Identifying and addressing student difficulties and misconceptions: examples from physics and from materials science and engineering

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Rebecca Rosenblatt, B.S.

Graduate Program in Physics

The Ohio State University

2012

Dissertation Committee:

Andrew F. Heckler, Advisor

Lei Bao

Bruce Patton

Nandini Trivedi

Abstract

Here I present my work identifying and addressing student difficulties with several materials science and physics topics. In the first part of this thesis, I present my work identifying student difficulties and misconceptions about the directional relationships between net force, velocity, and acceleration in one dimension. This is accomplished primarily through the discussion of the development, validation, and results of implementation of the FVA test, a research-oriented multiple-choice assessment instrument. In the second part of this thesis, I present my work identifying and addressing student difficulties in materials science through the design, implementation, and assessment of group work concept oriented tutorials. These tutorials were designed to mimic the tutorials developed by the University of Washington and the University of Maine and known to be effective in physics education. In addition, these tutorials include several teaching techniques found to be effective in physics education such as multiple representations, cognitive conflict, i.e. elicit-confront-resolve, and student dialog questions. While these tutorials still have a lot of room for improvement, the results suggest that these tutorials and recitation methods are effective in teaching students the difficult and important conceptual materials which they were designed to address. Furthermore, since the general design process used was not specific to that of materials science, there are wider implications that this process may be successful for a wide range of STEM courses. In this second part, I also discuss the development of a second multiple-choice assessment instrument which is designed to be more of an instructional tool, although it has been used for research as well to assess student conceptual understanding of the introductory materials science and engineering course.

To the student in each of us, may your curiosity never wane.

ACKNOWLEDGMENTS

I want to thank my family: Dad, Mom, Audrey, Ben, Ivana, Dave, Heather, and Jon. Without your love and support this would probably never have been written.

Special thanks also to my advisor Andrew Heckler whose advice was invaluable. I hope to be as good a mentor to future PER researchers.

Lastly I would like to thank my fellow PER graduate students and postdocs: Tomm Scaife, Ellie Sayre, Cameron Teichgraeber, and Brendon Mikula. I know these last 5 years would not have been nearly as informative and interesting without you.

VITA

2010	Master of Science in Physics from The Ohio State University
2008-2011	Center for Emergent Materials Research Fellow at The Ohio State University
2006-2008	Graduate Teaching Associate at The Ohio State University
2006	Bachelor of Science in Physics from University of Illinois Urbana/Champaign with high distinction in curriculum
2006	Research Assistant at the University of Illinois (Summer REU in Mathematics.)
2006	Lab instructor at the University of Illinois

Publications

1) Rosenblatt, R. and Heckler, A. F., "A Systematic Study of Student Understanding of the Relationships Between the Directions of Force, Velocity, and Acceleration in One Dimension." PRST-PER, 7 020112, 2011. (DOI: 10.1103/PhysRevSTPER.7.020112)

2) Rosenblatt, R., Heckler, A. F., Flores, K. "Tutorials for an Introductory Materials Engineering Course," to appear in 41st ASEE/IEEE Frontiers in Education Conference Proceeding, 2011.

3) Heckler A.F., & Rosenblatt, R., "Student difficulties with Basic Concepts in Introductory Materials Science Engineering," to appear in 41st ASEE/IEEE Frontiers in Education Conference Proceeding, 2011.

4) Rosenblatt, R., & Heckler, A.F., "Student understanding of the mechanical properties of metals in an introductory materials science engineering course," Proceedings of the Annual Conference of the American Society of Engineering Education, 2010.

5) Heckler, A.F., & Rosenblatt, R., “Student understanding of atomic bonds and their relation to mechanical properties of metals in an introductory materials science engineering course.” Proceedings of the Annual Conference of the American Society of Engineering Education, 2010.

6) R. Rosenblatt, E.C. Sayre, and A. F. Heckler. “Modeling students’ conceptual understanding of force, velocity, and acceleration.” In Proceedings of the 2009 PER Conference, AIP, Melville, NY, 2009, 245–248.

7) R. Rosenblatt, E. C. Sayre, and A. F. Heckler. “Toward a Comprehensive Picture of Student Understanding of Force, Velocity, and Acceleration.” In Proceedings of the 2008 PER Conference, AIP, Melville, NY, 2008, 183–186.

Fields of Study

Major Field: Physics Education Research

Studies in course materials and assessments design in the areas of Physics and Materials Science, assessment of student understanding and learning mechanisms, and understanding the role of diagrams in learning and problem solving: Andrew Heckler, PhD Thesis Advisor.

Table of Contents

II Identifying and Addressing Student Difficulties with Materials Science Concepts

Appendices

List of Figures

Figure **Page**

List of Tables

Table **Page**

Chapter 1

INTRODUCTION

1.1 Overview of the Thesis

I report here on two main areas of research and design. These two areas are "Identifying Student Difficulties and Misconceptions with a Specific Physics Concept", namely the directional relationships between net force, velocity, and acceleration in one dimension, and "Identifying and Addressing Student Difficulties with Materials Science Concepts."

1.1.1 Goals of the Thesis

The main goal of this thesis is to summarize and present the results of my work on identifying and addressing student difficulties with physics and materials science so that this important research and its instructional implications can be broadly available.

The goals of my work cover several topics.

Goal 1 was to develop a multiple-choice instrument to assess student understanding of the directional relationships between force, velocity, and acceleration in one dimension.

It is well known that students confuse the directional relationship between net force and velocity. They often indicate that a moving object must have a force in the direction of that motion, and they often indicate that when a net force is present the object must be moving in that direction. However, student confusion of the directional relationship between acceleration and velocity has not been very well studied. No study has been done considering a single student's understanding of all of the directional relationships between net force, velocity, and acceleration. The assessment instrument developed here is capable of answering questions like: "Does a student who understands that velocity and acceleration are not directionally related also understand that force and velocity are not directionally related?"

Goal 2 was to use this instrument to gain a better understanding of student knowledge about the relationships between net force, velocity, and acceleration and to gain a better understanding of how that knowledge changes with student level and instruction. This study gives us a much more global understanding of the occurrence, and therefore, the nature for the confusion among the three vectors. This understanding can then lead to improved instruction in this topic in which it is notoriously difficult to achieve student gains.

For example, we have found that many students have reponses for the relationships between velocity and net force and velocity and acceleration, that are neither the correct response (they are not related at all) or the common misconception response (they must be aligned with each other). Such knowledge can be invaluable to an instructor who is trying to teach the topic because he or she knows which examples will be the most informative for the most students. In addition, more complex patterns of student responses can be analyzed. For example, we can see if student responses are highly context dependant which might suggest students are relying on personal experience above formal reasoning. Or, we can see if students who respond that acceleration and net force uniquely imply each other are more likely, than other students, to use the same response-types for questions assessing the relationship between acceleration, and velocity and net force and velocity, which might suggest that students have and apply globally consistant models.

Goal 3 was to develop an effective set of group-work, conceptual worksheets for use in recitations in the introductory materials science course.

Materials science and engineering is a growing field of education research, but compared to physics education it is still in the early stages. While it has several factors in common with physics there are also many concerns specific to the disciplines of both engineering and materials science.

Group-work, concept-oriented worksheets, often called tutorials, have been developed for use in several areas of physics and been shown to be very effective teaching tools. There is a fairly well-known process that is used to develop these worksheets, and we believed that the process could also be used to develop effective materials science tutorials. We applied this process to the course. Identifying student difficulties was the most time consuming piece because so little research has yet to be done in this area. We will see that these tutorials achieved excellent results.

Goal 4 was to develop a multiple-choice assessment instrument to test the effectiveness of these conceptual worksheets, or ‘tutorials’, and which can generally be used in coordination with the materials concept inventory (MCI) to assess student understanding in materials science.

1.1.2 Thesis Content

Chapter 1 is the introduction. It provides an overview of the thesis and some theory and motivation for each area of research.

Chapter 2 is a brief discussion of test design, theory, and validation.

Chapter 3 is the force, velocity, and acceleration (FVA) test. It reviews the design and validation of the test, but more importantly the results from several sets of data that show interesting patterns in students' understanding of force, velocity, and acceleration, that in turn may have important instructional implications.

Chapter 4 is the start of the materials science part. It presents an overview of the project: literature reviews, motivation for the study, instructor goals for the course, implementation concerns, data collection methods, and results.

Chapter 5 continues the materials science discussion. It presents examples of student difficulties in each tutorial area: nature of atomic bonds, crystal structure, diffusion, the mechanical properties of metals, the effects of processing on properties, failure, phase diagrams, and TTT plots. A discussion about the tutorials themselves is included.

Chapter 6 provides a discussion of the learning and teaching methods used across several of the tutorials and their relations to overall course goals.

Chapter 7 is the last of the materials science chapters. It reports on the design and validation of the materials science concept evaluation (MSCE). A multiple-choice assessment instrument designed to test student knowledge of the tutorial materials.

Chapter 8 gives a brief conclusion and summing up of the reported research.

1.2 PER and Student Difficulties Research

We begin with a brief discussion of the history and importance of identifying student difficulites and misconceptions as part of PER and science education in general.

Physics education research (PER) has been a developing field of study for about 30 years. One of the main originating factors of PER was the discovery of significant student difficulties post instruction with simple conceptual topics that were previously assumed to have been understood by students. The development and use of the force concept inventory (FCI) was one of the most important assessments that made these difficulties widely publicized (Hestenes, Wells, and Swackhamer (1992)).

The FCI is from an expert perspective a straightforward and simple test of the concepts surrounding Newton's first, second and third laws. However, most students find these conceptual questions to be difficult and tricky. One of the most famous stories surrounding the FCI and student difficulties is the Harvard physics student who asked his instructor a few minutes after the test had started if he should answer the questions the way he was taught in class or the way he normally thought about the topic (Beichner (2009)). This discord between student's intuitions about physics and the formal scientific principles of physics can often make the learning of physics substantially more difficult for students.

The most common explanation for this is the constructivist theory of learning. This theory, first developed and promoted by Piaget, posets that people constantly update and build on their existing mental models via interaction with the world. From a learning perspective the most important part of this theory is that students do not simply add new and drop old information but rather build new knowelge into their existing knowledge structure (Taber (2006)). For this reason, having specific knowledge of a student's starting knowledge structure is very important in knowing how to help them build to the desired final target knowledge structure.

In addition, many physics curriculum, and instructors, support a theory of conceptual change that stipulates that students will not change or replace an existing theory without significant dissatisfaction with that theory (Posner, Strike, Hewson, and Gertzog (1982)). So, in order for a student to learn, he or she must be placed in a situation in which he or she will become dissatisfied with their old incorrect theory. Knowledge of specific student difficulties with a topic allows the instructors to adjust the curriculum appropriately. This is part of the well-known 'elicit-confront-resolve' curriculum design of the Washington tutorials (Heron (Shaffer); McDermott, Shaffer, and the PER Group at U. of Wash. (2002)) which have been shown to be highly effective in creating large student gains on the FCI (Hake (1998)).

Thus, it is clear that knowledge and awareness of specific student difficulties in a subject matter is of utmost importance to an education community. This importance is driven by both the need for awareness of instructional short comings and the need for instructional materials designed to address these shortcomings. In this thesis, I present a very specific, and as yet unaddressed, aspect of the already large body of knowledge about student difficulties with the directions of net force, velocity, and acceleration. In addition, I present a sizable range of student difficulties with materials science which adds to the body of knowledge in this developing field. I also present a new set of instructional tutorials designed from these difficulties in an 'elicit-confront-resolve' manner.

1.3 Motivations for Materials Science and Engineering (MSE) Research

Materials science and engineering as an educational discipline is relatively new. For example, the materials concept inventory (MCI) was published in 2002, about 10 years after the FCI. As previously stated, both materials science and engineering are distinct disciplines and therefore have their own educational concerns which are different from physics education. However, since a majority of both courses are made up of engineering students, the population is almost identical in terms of intelligence (or learning ability) and previous science and math background. In addition, the courses themselves are very similar in terms of the types of material the students are supposed to master - basic terms and definitions, graphical manipulations and interpretations, algebraic and functional mathematics, geometric visualization, and multivariable reasoning, to name a few. For these reasons, we believed that tutorials (i.e. conceptual group-work worksheets) could be designed with the process that was used to develop physics tutorials, and could be as effective as they have prooved to be.

There is a well-known process for how the physics tutorials were developed. This process is outlined and discussed in Chapter 4. I will not go into detail here. However, the first steps of this process are to identify the goals for a course and student's difficulties attaining those goals. Only once this is know can activities be created that would address these difficulties.

1.4 Discussion of the Similarities and Differences of the Two Parts of the Thesis

As previously stated, I report here on two main areas of research and design. These two areas are "Identifying Student Difficulties and Misconceptions with a Specific Physics Concept", namely the directional relationships between net force, velocity, and acceleration in one dimension, and "Identifying and Addressing Student Difficulties with Materials Science Concepts." These two areas are both very different and very similar. They are very different in the following three ways.

One, the first part of this thesis is a research driven study. Student difficulties with net force, velocity, and acceleration were already known to be an issue and instructional materials were already developed to improve instruction in this area. However, upon considering the existing research there was a gap in reported studies. Limited research had been done looking at the acceleration and velocity relationship, and no research had been done on a within student assessment of understanding of all relationships, see Figure 1.1. Such a study would answer questions like, "Does a student who believes a moving object must have a net force in the direction of its motion also believe that a moving object must be accelerating in the direction of its motion?" Filling this gap was the goal of the study.

Two, this first thesis part analyzes in detail a very small piece of a much larger picture. The topic, student understanding of the directional relationships, is very specific. It does not address two or three dimensions, it does not look at magnitude, and it does not consider speeding up, slowing down, or other situations where the change in velocity is a large part of the given information.

Lastly, the results are academic because the instructional implications of the study have not been tested. The data shows that students at an introductory level show gains in under-

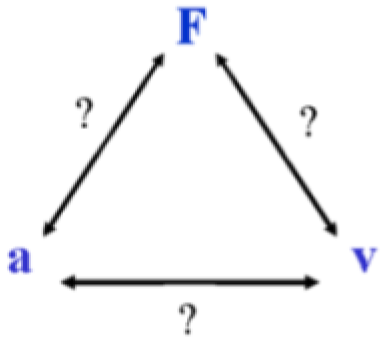


Figure 1.1: What do students think the directional relationships for net force, velocity, and acceleration are?

standing of the velocity and acceleration relationship before showing gains in understanding of the net force and velocity relationship. It also shows that almost no students who know the relationship between net force and velocity do not know the relationship between acceleration and velocity. However, it is unclear how, or if, a different instruction method would produce different student responses and such studies have not been done.

In comparison, the second part of this thesis is an instructional methods driven study. The scope of the study was very broad including the entire course. And, its purpose was to create immediately useful and effective instructional materials.

These two thesis parts are very similar in many ways as well. Both parts involve the creation and validation of instructional materials. They involve a study of student difficulties with particular areas of introductory course work. And, in each, results from a large set of mostly quantitative student data are reported, analyzed, and discussed. Thus, despite the differences in the two studies most of the research activities completed and the research approach are the same. For example, both included reading of the existing literature on the two areas of difficulty, initial exploratory data collection, definition of explicit research goals and method to achieve those goals, brain storming and group discussion of question wordings and topics, performing student interviews, data analysis with statistical software, and the creation of and analysis of a variety of validation tools. Because of these similarities, I include both of these topics together in this thesis under the wider heading of "Identifying and Addressing Student Difficulties and Misconceptions."

Chapter 2

Test Design, Theory, and Validation

Because test design is a large part of this thesis, here we provide a short discussion of test design as used in PER and this thesis, pros and cons of multiple-choice assessments, classical theory of test validation, and a preview of validation steps reported in this thesis in their respective parts.

2.1 Test Design as Used in PER and this Thesis

2.1.1 Generally Accepted Process for Multiple-Choice Test Development

The flow chart shown in Figure 2.1 is a well-known process for test development and was taken from Beichner's 1994 paper Beichner (1994). These steps are: 1. Identify a need for a new test; 2. Determine goals for the test; 3. Create the items; 4. Conduct preliminary testing for validity and reliability with revisions to the items as needed; 5. Perform validation and reliability checks on the final items; 6. Distribute and use the test. While this process is generally well known and commonly used, there are many levels of rigor and precision with which it is used, and there are many different ways that validity and reliability are established and reported (Lindell, Peak, and Foster (2007)). For example, the original FCI does not report any reliability or discrimination statistics Hestenes, Wells, and Swackhamer (1992), but these studies have been done and presented in other papers, (Wang and Bao (2010) and Rebello and Zollman (2004)) to name a few. In comparison the TUG-K describes in detail the process used to create the distractors, the way in which the content validity was established, students scores on an test designed to check equivalency reliability, content validity checks like higher scores with higher class level, point biserial coefficients, item difficulties, average item difficulty, and KR-20 (Beichner (1994)). This is not to say that the FCI was not rigorously developed. Simply that it was not rigorously presented. In the presentation of the design of the two tests in this thesis, I make every

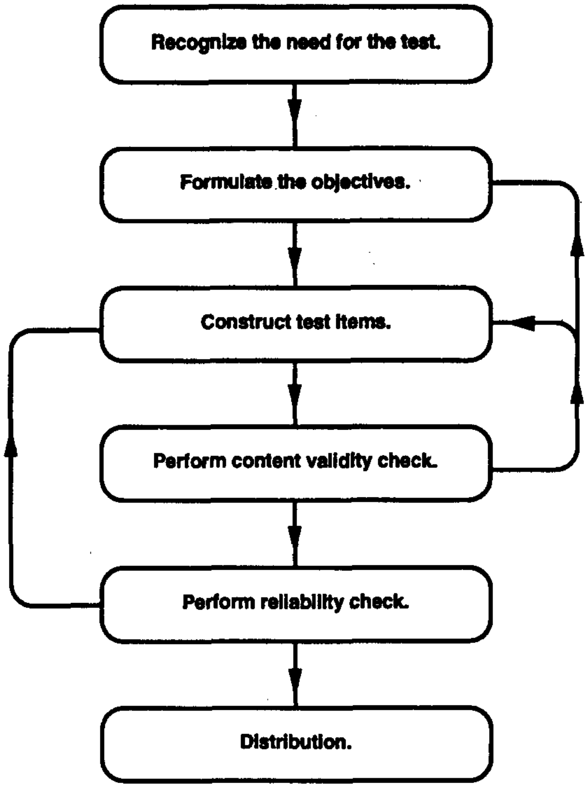


Figure 2.1: Well-known process for test development.

effort to follow the rigorous process exemplified by the TUG-K.

2.1.2 Differences in Classical Test Design Theory and PER Test Design

While PER does heavily use classical test design theory and validation, it differs in how it applies these principles because of specific goals for the tests. Two main areas of difference are in the use of distractors for test items and in the creation of broader tests that assess multiple factors.

Many science diagnostic tests use distractors as multiple-choice options - e.g. FCI, BEMA, FMCE, MCI (Ding, Chabay, Sherwood, and Beichner (2006); Hestenes, Wells, and Swackhamer (1992); Krause, Decker, Niska, and Alford (2002); Thornton and Sokoloff (1998)). Distractors are options that are representative of common incorrect student responses to the item, and they are often created from student responses to the item when given as an open ended question (Beichner (1994); Ding, Chabay, Sherwood, and Beichner (2006); Hestenes, Wells, and Swackhamer (1992); Krause, Decker, Niska, and Alford (2002)). While distractors have their issues, their usefulness as an informative teaching tool makes them invaluable as part of many educational assessment instruments.

Many tests designed as assessments in science do not actually test a single topic. Classical test validation proposes that a test should assess only one thing because the results reported from a test, the score, is than a measure of that one thing. When a test assesses more than one thing the score on the test no longer represents student ability in one thing (Messick (1995)). For example if half of the questions on a test are about force and the other about energy, a score of 50% could mean either that a student knows everything about one topic and nothing about another topic or that the student knows half of what they should in both, etc. Such ambiguity in data meaning and interpretation drives this classical stipulation. However, for educators it can often be very useful to know what a student knows about both topics at once. Thus, it is common for a single test to be given which assesses both and scores on groups of questions are then used to clarify the meaning of overall score. (This is what is done with the FVA test presented here.)

2.2 Why Multiple-Choice Tests are Used and Their Limitations

Multiple-choice tests are perhaps the most widely used test type. There are several well-known advantages to multiple-choice testing. The administration and grading of the tests is straightforward and much less time consuming than many other types of testing such as interviews or short-answer. The student responses are free of subjective grading or coding and can be easily converted to a spreadsheet for a myriad of statistical analysis techniques. Because the testing is not time consuming, a large data set can be collected which makes the results of reported data more likely to be generalizable to other students and institutions. And, when properly designed the tests can be used for complex diagnostic measures in a number of areas such as scientific reasoning (Lawson, A. E. (1978)), content knowledge and understanding in science areas including physics, chemistry, biology, and materials science (Hestenes, Wells, and Swackhamer (1992); Krause, Decker, Niska, and Alford (2002); Treagust, D. F. (1988)), and many other areas of reasoning, cognition, and knowledge.

The main disadvantage of multiple-choice tests is their limited ability to probe in depth into a student's thought process, and some research shows that students can provide correct answers without being able to justify their correct answer choice (Tamir (1990)). Also, it can be difficult to tell when students are guessing or not trying to do their best. However, if test questions are well developed and the testing environment is well controlled such drawbacks can be minimized.

(Note: the above discussion, while well accepted and used in many other articles and thesis, follows closely the points made in Engelhardt's *Introduction to classical test theory* Engelhardt (2009).)

2.3 Theory of Validation and Reliability

Test reliability and validity are necessary for stable and applicable inferences to be drawn from a test. Validity and reliability while commonly discussed separately are very closely related to each other; the distinction here is made for convenience, but in practice, a test which is not reliable is not valid and vice versa.

Test reliability is a measure of the error associated with the reported score on the test. There are three main types of reliability: internal consistency of the test questions; stability of testing responses; and equivalency. The first of these, internal consistency, concerns how related all the questions on the test are to each other. When we report and use testing data, only one number, student score, is used to distinguish that student's ability in area x. Thus, because the test should than be only measuring one thing, x, the internal consistency of the test should be high. The second of these, stability, addresses the likelihood that the same score would be received by a student if he or she could take the test twice without any learning or interference after the first administration of the test. In other words, how much range is there in the measurement of the theoretical trait being assessed or how closely does it measure a student to his or her actual trait. The last type of reliability is equivalency. This is how much agreement there is, or would be, between the test's measure of trait x and a separate set of *equivalent* items' measure of trait x (Engelhardt (2009)).

Test validity is usually defined as the extent to which a test measures what it is supposed to measure. As such, there is no single statistical measure of validity of a test. Contemporary thinking distinguishes three different types of validity for conveince. (Most validation experts promote a more global philosophy of validity and reliability that does not divide the validity into separate parts, e.g. Messick 1995.) Content validity measures a tests ability to assess the materials it is supposed to be testing. This type of validity is most often completed by consultation with experts in the testing domain. Construct validity measures a test's ability to assess a theoretical trait in the test taker. In our research, this is most often student understanding of a concept, but this can be a very broad category, such as intelligence, spatial reasoning, ethics, etc, depending on the type of test. Criterion validity measures a test's ability to assess transferable skills or abilities; in other words, to what extent does performance on the test relate to performance in other domains (Engelhardt (2009); Messick (1995)).

2.4 A Preview of Validation Steps Reported in This Thesis

For both the FVA and MSCE tests discussed here all three types of validity were measured and results confirming validity of the tests were seen. For the FVA test, we showed a group of three other PER researchers, and a few graduate students, several questions on the FVA test. We described the goals of the test, showed them some student response data, and explained the inferences we were drawing from the data. These researchers were asked to give feedback on the question wording and projects goals. From this, we received only one concern about content validity. This was that we needed to ensure that the wording was clearly referring to an instant in time for each item. This was put into action and all FVA test data reported in Part I has this edit. In this way the content validity of the FVA test was established.

For the MSCE test we used a more formal process of asking three course instructors, and materials scientists, to assess four aspects of the items: the importance of each items inclussion in the MSCE test (reported as a number 1 to 10 on a likert scale); the correct answer for each item; the level of concern for each items reported student response data (reported as a number 1 to 10 on a likert scale); and finally, any comments they had about question wording or content. This data was analyzed and items with low importance were dropped from the MSCE. In addition, items were reworded as needed to make them technically sound and clearer based on instructor comments and/or concerns.

The construct validity of the two tests was completed through a series of student think aloud interviews as the students completed the test questions. In addition, for the MSCE a set of "choose-an-answer and then explain-your-choice questions" were given because of the length of the test and the time consuming nature of think aloud interviews. In both cases such validation assessments were done before, during, and after item creation. With the before and during used to update and edit the items on the respective tests and the post being used as a validation measure.

The criterion validity of the two tests was completed through statistical comparisons of the response data with other measurements of conceptual understanding. In no case was a comparison made which suggested that the tests were invalid. This would have been the case if we expected to see a statistically significant positive or negative correlation and did not see one. Both tests were correlated with final exam scores. In addition, the FVA was correlated with the FCI which measures similar conceptual understanding. Both tests were correlated with increasing student levels. This was done in several different ways for the two tests. The MSCE was given to students and to teaching assistants for the course; this showed the expected results that the TAs would perform much better than their students. In addition, the MSCE test showed gains with increasing tutorial use which was expected since tutorials were shown to be effective teaching tools. The FVA test, on the other hand, was given to several levels of mechanics students and showed the expected trend in better score with higher student levels. It was also given pre and post course which showed within student gains.

The specific data and greater explanations of these different tests of validity are included in each of the later sections discussing the tests. Here we report only on some of the theory of test validity and give an accounting of the different methods we used to validate the test.

Again, in no case did any of these several measures for content, construct, and criterion validity show unexpected response or data trends which would have suggested that the tests were measuring things other than what the tests were designed to measure. Thus, we concluded that the two tests are, to the best of our knowledge, valid.

Volume I

Identifying Student Difficulties and Misconceptions with The Relationship Between the Directions of Net Force, Velocity, and Acceleration in One Dimension

Chapter 3

The Relationship Between the Directions of Net Force, Velocity, and Acceleration in One Dimension

3.1 Introduction

One of the earliest and most studied areas in physics education research is student understanding of force, velocity, and acceleration. For example, perhaps the most widely known and documented phenomenon in this field is the (incorrect) student belief that the net force on an object and its velocity must be in the same direction (Clement (1982); Halloun and Hestenes (1985); Hestenes, Wells, and Swackhamer (1992); Thornton and Sokoloff (1998); Viennot (1979)). It is also well documented that students often have difficulty distinguishing between the velocity and acceleration of an object (Reif and Allen (1992); Trowbridge and McDermott (1981)).

Nonetheless, even though this topic is relatively well studied, there remain many unanswered questions that are critical to both advancing our knowledge of student difficulties with force, velocity, and acceleration and applying this knowledge to improve student learning of these fundamental concepts. For example, empirically speaking, to what extent does the correct understanding of the relationship between, say, force and acceleration depend on the correct understanding of another relation, say, between force and velocity? Does the path to correct understanding of these relations empirically occur in steps? If so, what are the steps?

Furthermore, it is important to point out that when assessing student understanding of the relations between force, velocity, and acceleration, the questions posed typically involve *conditional relations*, though this has not been explicitly acknowledged or systematically studied in previous work. For example, in a landmark paper, Viennot (1979) posed questions of the form “given the velocity of an object, what is the (net) force on the object?”, which is a conditional relation of the form “given x, what is y?”. There were no questions in Viennot’s study probing the converse conditional relation “given a net force on an object, what is its velocity?”. Nor were there any questions regarding the relations between velocity and acceleration or acceleration and force (Viennot (1979)). Certainly, in other studies that followed Viennot’s paper, other conditional relationships were studied. However, as can be seen from Table 3.1, which summarizes the relationships studied in many of the existing research papers on students’ conceptual understanding of the directional relationships of force, velocity, and acceleration, there has been no systematic study of student understanding of all six possible paired conditional relations between the concepts of force, velocity, and acceleration. Furthermore, there has been an abundance of work on some of the six relations and little, if any, on others.

Table 3.1: Question types investigated in a sample of previous studies.
$\vec{x} \to \vec{y}$ notation indicates a question of the form: Given $\vec{x}$, what can be inferred about $\vec{y}$.

Number of Questions	Question Type	Citation of study
4	$\vec{v} \to \vec{F}$	Viennot (1979)
2	$(\Delta\vec{v}, dt) \to \vec{a}$	Trowbridge and McDermott (1981)
2	$\vec{v} \to \vec{F}$	Clement (1982)
2	$\vec{F} \to \vec{v}$	Clement (1982)
21	$\vec{v} \to \vec{a}$[a]	Reif and Allen (1992)
1	$\vec{a} \to \vec{v}$	Reif and Allen (1992)
7	$\vec{v} \to \vec{F}$	FCI test: Hestenes, Wells, Swackhammer (1992, 1995)
9	$\vec{F} \to \vec{v}, \Delta\vec{v}$[b]	FCI test: Hestenes, Wells, Swackhammer (1992, 1995)
4	$\vec{F} \to \Delta\vec{v}$	FCI test: Hestenes, Wells, Swackhammer (1992, 1995)
4	$\vec{v} \to \vec{F}$	Enderstein and Spargo (1996)
8	$\vec{v} \to \vec{F}$	Palmer (1997)
2	$\vec{F} \to \vec{v}$	Enderstein and Spargo (1998)
4	$\vec{v} \to \vec{F}$	Enderstein and Spargo (1998)
12	$\vec{v} \to \vec{F}$[c]	FMCE test: Thornton and Sokoloff (1998)
1	$\vec{a} \to \vec{F}$	FMCE test: Thornton and Sokoloff (1998)
3	$\vec{v} \to \vec{a}$	FMCE test: Thornton and Sokoloff (1998)

[a] In the problem set up many of the questions specify the speed in addition to the direction of the velocity.

[b] In the problem set up 5 of the questions specify a force, 2 gravity is implied, 2 zero force is implied.

[c] In the problem set up 6 of the questions specify the speed in addition to the direction of the velocity and 6 do not.

A systematic study of all possible pairs of conditional relations between force, velocity, and acceleration is important for two reasons. First, a within-student study of all possible pairs of relations will allow for a more holistic picture of student understanding of all relations and the possibility of determining whether understanding one relation may effect (or predict) the understanding of another relation. Second, it is not unreasonable to expect that for a given pair of variables, a conditional relation between the pair and its converse may not be answered similarly by the student. For example, the question "An object is accelerating in a certain direction, what can you infer about the object's velocity?" may be answered differently than the question "An object has a velocity in a certain direction, what can you infer about the object's acceleration?" Furthermore, if there is a causal relation between the variables (real or believed), such as between force and acceleration, then making inferences about the effect of a given cause may be different than making inferences about the cause of a given effect (Tversky and Kahneman (1980)).

Therefore, here we will investigate student understanding of all possible pairs of relations between force, velocity, and acceleration. To more precisely focus the investigation, we will only study student understanding of the relations between the *directions* of force, velocity, and acceleration in one dimension, and leave the investigation of multiple dimensions and the relations between the magnitudes of these variables for other studies.

While this investigation included a significant amount of student interviews and open-ended written answers, the bulk of the analysis is based on a multiple-choice test that we developed for this study. The multiple-choice test allows for, in principle, the identification of reliable patterns based on a large number of students. On the other hand, such a test can lack the subtlety and depth compared to a more qualitative study; nonetheless the validity and reliability of the results claimed here were corroborated by the interviews and written answers of students. Clearly an in-depth study using more qualitative data would also yield interesting results, but here we focus on some of the important, replicable patterns found via the carefully constructed instrument.

Finally, we have one further introductory comment before proceeding. In a relatively recent study, Alonso and Steedle Alonzo and Steedle (2009) have investigated middle-school student (12-14 year old) understanding of force and motion. They hypothesize increasingly expert-like levels of understanding of force and motion through which middle school students pass in a progression towards mastery of these concepts. Specifically, they construct a formal "learning progression" of force and motion for this population. The topic of learning progressions has recently generated significant interest in the science education community (e.g., see NRC (2007)) and is somewhat relevant to this study since we examine longitudinal and cross sectional data on student performance and we are interested in the steps and hierarchies in understanding the relation between the directions of force, velocity and acceleration. While the topic of learning progressions is not the focus of this part of the thesis, we will briefly comment on this topic and Alonso and Steedle's study in the final discussion section.

This chapter proceeds as follows. We first describe the development and careful construction of the short, multiple-choice assessment instrument and report on its validity and reliability. Next we present test results pre and post instruction, and results of students at different levels of physics knowledge. These results include an analysis of within and between-student answering patterns for all six conditional relations and how answering patterns change both over one course and from first to second year university physics students. Finally we summarize and discuss how the findings might be applied to the design of instruction aimed at improving student understanding of the relations between the directions of force, velocity, and acceleration.

3.2 Development of The Force, Velocity and Acceleration Direction Assessment Instrument

3.2.1 Description of the FVA Test

We constructed a 17-item multiple-choice test, called the FVA test, designed to assess student understanding of all six conditional relationships between the directions of force, velocity, and acceleration in one dimension. Each item in the test presents a simple scenario indicating the direction of one of the vectors for an object, say acceleration $\vec{a}$, and asks what this implies about the direction of one of the other vectors, say velocity $\vec{v}$. We label such a question as $\vec{a} \rightarrow \vec{v}$, which briefly means "given the acceleration, what can be inferred about the velocity?" (See Table 3.2 for an examples of an $\vec{a} \rightarrow \vec{v}$ and an $\vec{F} \rightarrow \vec{v}$ question.) Ten of the 17 items include two questions for each of the conditional relations $\vec{F} \rightarrow \vec{v}$, $\vec{v} \rightarrow \vec{F}$, $\vec{a} \rightarrow \vec{v}$, and $\vec{v} \rightarrow \vec{a}$ and one question each for $\vec{a} \rightarrow \vec{F}$ and $\vec{F} \rightarrow \vec{a}$. These 10 items directly probe the 6 conditional relations between force, velocity, and acceleration, which are of particular interest here.

The specific results of the remaining 7 items also provide additional interesting information. However, except for being part of the reported total score results and item statistics of the FVA test, a detailed analysis of response patterns from these 7 items are not reported here as they are not the focus of this thesis. Nonetheless it is worth mentioning that these seven items were included in the FVA test for several reasons. First, they provide variety in answering, so that the correct answer is not always "a, b, and c are possible" (see Table 3.2 or 3.3), which is the case for the eight $\vec{v} \rightarrow \vec{F}$, $\vec{F} \rightarrow \vec{v}$, $\vec{a} \rightarrow \vec{v}$, $\vec{v} \rightarrow \vec{a}$ items. We have found that if the answer choice is always (or often) the same, then students start thinking more about the "tricks" of the item format rather than the content of the question. Second, these items probe different aspects of understanding the directional relations of force, velocity, and acceleration and as such are part of a more valid and reliable FVA test. For example,

Table 3.2: Explanation and examples of the $\vec{x} \rightarrow \vec{y}$ notation.

A $\vec{x} \rightarrow \vec{y}$ question is designed to probe a student's understanding of how a given vector's direction, $\vec{x}$, is related to another vector's direction, $\vec{y}$. For example, an $\vec{a} \rightarrow \vec{v}$ question provides a simple scenario indicating the direction of the acceleration on an object and asks the student what this implies about the direction of the object's velocity. Two specific examples from the developed test are provided below.

Example of $\vec{a} \rightarrow \vec{v}$ question: "A car is on a hill and the direction of its acceleration is uphill. Which statement best describes the motion of the car at that time?"	Example of $\vec{F} \rightarrow \vec{v}$ question: "At a particular instant of time, there are several forces acting on an object in both the positive and negative direction, but the forces in the negative direction (to the left) are greater. Which statement best describes the motion of the object at this instant?"
a) it is moving uphill	a)it is moving to the right
b) it is moving downhill	b) it is moving to the left
c) it is not moving	c) it is not moving
d) both a and b are possible	d) both a and b are possible
e) both a and c are possible	e) both b and c are possible
f) a, b, and c are possible	f) a, b, and c are possible

two of the seven items provide situations in which an object is explicitly at rest (item 12) or has zero net forces acting on it (item 3). Furthermore, three of the items (2, 7, and 8) provide (or ask for) information of both the velocity and the change in speed. Finally, two of the items (2 and 15) are very familiar and easy, and help to establish a baseline of student understanding. Detailed analysis of response patterns on these items is a topic of further study. The complete instrument is reproduced in Appendix A.

Item construction occurred over a period of several years, beginning with open-ended pencil and paper questions and over 40 individual student interviews in a think-aloud format. This was followed up by over 60 individual debriefings for students completing the final versions of the FVA test. The test development involved two major iterations, as explained in more detail in the section "The construct validity of the question and answer choice format." This process revealed that there were several possible student response choices for the questions posed, and it was important to include all of these possibilities as response choices in the multiple-choice format.

Specifically, for one dimensional motion there are three ways in which any one the three relevant vectors (net force, velocity, or acceleration) can be related to any other: the vectors can be aligned, (symbolized as $\uparrow\uparrow$), opposite to one another, ($\uparrow\downarrow$), or one of them can be zero, ($\uparrow 0$). Using acceleration and velocity to illustrate the seven possible combinations of these relationships yields: $\vec{a} \uparrow\uparrow \vec{v}$ ($\vec{v}$ must be parallel to $\vec{a}$); $\vec{a} \uparrow\downarrow \vec{v}$ ($\vec{v}$ must be antiparallel to $\vec{a}$); $\vec{a} \uparrow 0\ \vec{v}$ ($\vec{v}$ is zero to a non-zero $\vec{a}$); $\vec{a}(\uparrow\uparrow, \uparrow 0\ )\vec{v}$ ($\vec{v}$ can be zero or parallel to $\vec{a}$);

Table 3.3: Available student response choices for each FVA item.

Almost all possible choices for the relationship between two vectors are available responses for a student to choose from when answering a question. The other possible choices were almost never chosen, thus excluded. Consider the $\vec{a} \rightarrow \vec{v}$ question as an example: "A car is on a hill and the direction of its acceleration is uphill. Which statement best describes the motion of the car at this time?"

Response Choices	Symbolic Representation of Choices	Description of Most Common Choices
a) it is moving uphill	$\vec{a} \uparrow\uparrow \vec{v}$	Common "Misconception"
b) it is moving downhill	$\vec{a} \uparrow\downarrow \vec{v}$	
c) it is not moving	$\vec{a} \uparrow 0\ \vec{v}$	
d) both a and b are possible	$\vec{a}(\uparrow\uparrow, \uparrow\downarrow)\vec{v}$	($\vec{v}$) "Cannot-be-Zero" *partially correct*
e) both a and c are possible	$\vec{a}(\uparrow\uparrow, \uparrow 0\ )\vec{v}$	($\vec{v}$) "Cannot-be-Opposite" (to $\vec{a}$) *partially correct*
f) a, b, and c are possible	$\vec{a}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{v}$	Correct

$\vec{a}(\uparrow\uparrow, \uparrow\downarrow)\vec{v}$ ($\vec{v}$ can be parallel or antiparallel to $\vec{a}$); $\vec{a}(\uparrow\downarrow, \uparrow 0)\vec{v}$ ($\vec{v}$ can be antiparallel to $\vec{a}$ or zero); $\vec{a}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{v}$ ($\vec{v}$ can be parallel or antiparallel to $\vec{a}$ or zero).

Of these seven possible combinations, we found that students rarely if ever considered the physically unnatural possibility of "can only be opposite or zero", thus usually only six response choices were provided. Table 3.3 provides an example of an item and the six possible response choice "models".

3.2.2 The Content Validity of the Questions.

In order to establish the content validity of the FVA test, we showed a group of three other PER researchers, and a few graduate students, several questions on the FVA test. We described the goals of the test, showed them some student response data, and explained the inferences we were drawing from the data. These researchers were asked to give feedback on the question wording and projects goals. From this, we received only one concern about content validity. This was that we needed to ensure that the wording was clearly referring to an instant in time for each item. This was put into action and all FVA test data reported in Part I has this edit. In this way the content validity of the FVA test was established.

3.2.3 The Construct Validity of the Answer-Choice Format.

We report here on the construct validity of the items, including the question and answer choice format, which was supported through several stages of interview and testing-based modifications.

The test went through two major changes to the question style during development before it became the final FVA test. The first change was to go from single response questions, with four answer options, to a more dynamic response choice where students circled always, sometimes, or never for a series of options for the net force, velocity, or acceleration's direction. (See Table 3.4 for examples of each question style.) Test 1, the original test design, was simple and students' answers could be practically categorized into two categories - either the answer was correct or the answer was consistent with a common misconception such as velocity is in the direction of net force, but there were two main drawbacks to the test as designed.

First, was the tests limited ability to capture a student's individual model. On Test 1 only four of the seven possible responses were available answers. Any students holding the other three models were grouped into one of the four available responses making the data less meaningful. For example, a student who believed that velocity could be either against or in the direction of acceleration but not zero, i.e. response $\vec{a}(\uparrow\uparrow, \uparrow\downarrow)\vec{v}$ which is sometimes referred to as Cannot-be-Zero, would probably have said that there was not enough information to answer the example question when it was posed in the Test 1 format. Therefore, Test 1 provided no way to distinguish these students from the students who believed that velocity could be either in the direction of the acceleration or zero but not opposite, i.e. response $\vec{a}(\uparrow\uparrow, \uparrow 0\ )\vec{v}$ which is some time referred to as Cannot-be-Opposite, or from students who were correct and thought velocity could be in either direction or zero, i.e. response $\vec{a}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{v}$. (See Table 3.4.)

The second main drawback was that the correct answer was often "not enough information." This is a unique kind of answer choice, and students may or may not have chosen it for reasons that were unrelated to the physics content of the problem. Informal interviews gave anecdotal evidence that for some small number of students this may have occurred.

The second main change in test style went from the dynamic options of always, sometimes, or never back to a multiple-choice test which this time had seven of the eight options for the vector relations. There were a few benefits that the dynamic options offered over the multiple-choice options. All combinations of answers were available for the students to choose from, which reduced the chance of grouping students with different models together. Also, it allowed for more accurate categorization of students into distinct conceptual models. For example, only students who said sometimes for all options were labeled correct, and only students who responded that given a force the velocity was always in the same direction and never opposite to the force and never zero were labeled as holding the common misconception, $\vec{F} \uparrow\uparrow \vec{v}$.

However, interviews revealed that while students often would answer in a (formally) logically consistent manner within a given question (e.g. "if it is always moving right, then it is never moving left or at rest"), on other questions the same student would answer in

Table 3.4: Examples of answer choices for three different test forms which test different question formats.

Examples of answer choices for the different tests for an $\vec{a} \rightarrow \vec{v}$ question. The prompt is "A car is on a hill and the direction of its acceleration is uphill. Which statement best describes the motion of the car?"

Test 1	Response Models Available for Test 1
a) It is moving uphill	$\vec{a} \uparrow\uparrow \vec{v}$ "Misconception"
b) It is moving downhill	$\vec{a} \uparrow\downarrow \vec{v}$
c) It is not moving	$\vec{a} \uparrow 0 \vec{v}$
d) Not enough information	$\vec{a}(\uparrow\uparrow, \uparrow\downarrow)\vec{v}$ "Cannot-be-Zero"
	$\vec{a}(\uparrow\uparrow, \uparrow 0\ )\vec{v}$ "Cannot-be-Opposite"
	$\vec{a}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{v}$ "Correct"

Test 2 (Always, Sometimes, Never)	Responses Models Available for Test 2
(A) (S) (N): The car is moving uphill	ANN: $\vec{a} \uparrow\uparrow \vec{v}$ "Misconception"
(A) (S) (N): The car is moving downhill	NAN: $\vec{a} \uparrow\downarrow \vec{v}$
(A) (S) (N): The car is not moving	NNA: $\vec{a} \uparrow 0 \vec{v}$
	SSN: $\vec{a}(\uparrow\uparrow, \uparrow\downarrow)\vec{v}$ "Cannot-be-Zero"
	SNS: $\vec{a}(\uparrow\uparrow, \uparrow 0\ )\vec{v}$ "Cannot-be-Opposite"
	SSS: $\vec{a}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{v}$ "Correct".
	And Other Undesirable Models: ASN, SAN, AAN...

Test 3	Response Models Available for Test 3
a) it is moving uphill	$\vec{a} \uparrow\uparrow \vec{v}$ "Misconception"
b) it is moving downhill	$\vec{a} \uparrow\downarrow \vec{v}$
c) it is not moving	$\vec{a} \uparrow 0 \vec{v}$
d) both a and b are possible	$\vec{a}(\uparrow\uparrow, \uparrow\downarrow)\vec{v}$ "Cannot-be-Zero"
e) both a and c are possible	$\vec{a}(\uparrow\uparrow, \uparrow 0\ )\vec{v}$ "Cannot-be-Opposite"
f) a, b, and c are possible	$\vec{a}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{v}$ "Correct"

logically inconsistent ways (e.g. "it is *always* moving right, but *sometimes* it could be going left"). On average, questions were answered illogically 9% of the time. In all, only 29 out of the 72 students answered all Test 2 (Always, Sometimes, Never) questions logically. While the lack of self consistency is interesting and could provide some insight into student answering and reasoning habits, we were not interested in studying it as part of the test. Thus, the final test, Test 3, which we now call the FVA test, was created. This test included most possible (and logically consistent) answers, resulting in a multiple-choice format with up to 7 mutually exclusive choices. (Again see Table 3.4 for an example.) (We did not usually include the response of can be zero or antiparallel, $\uparrow\downarrow, \uparrow 0$, because student did not use this response.)

Table 3.5: Comparison of student responses for equivalent questions in three different test forms. Each student took only one test form. Correct answers are designated with a *. Students were in the calculus mechanics course.

$\vec{F} \to \vec{a}$ A group of workers is pushing on a car on a level driveway and there is a net force on the car toward the street. What can you say about the acceleration of the car?

Test Type	N	$\vec{F} \uparrow\uparrow \vec{a}$	$\vec{F}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{a}$	$\vec{F}(\uparrow\uparrow, \uparrow\downarrow)\vec{a}$	$\vec{F}(\uparrow\uparrow, \uparrow 0)\vec{a}$	Other
Test 1	78	89%*	9%	NA	NA	2%
Test 2	40	35%*	13%	15%	20%	17%
Test 3	119	56%*	13%	4%	20%	7%

$\vec{F} \to \vec{v}$ A soccer player pushes on a soccer ball with her foot. What can you say about the motion of the ball?

Test Type	N	$\vec{F}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{v}$	$\vec{F} \uparrow\uparrow \vec{v}$	$\vec{F}(\uparrow\uparrow, \uparrow\downarrow)\vec{v}$	$\vec{F}(\uparrow\uparrow, \uparrow 0)\vec{v}$	Other
Test 1	191	38%*	58%	NA	NA	4%
Test 2	32	22%*	25%	31%	6%	16%
Test 3	119	11%*	61%	12%	14%	2%

$\vec{a} \to \vec{v}$ A car is on a hill, and the direction of the acceleration is uphill. What can you say about the motion of the car?

Test Type	N	$\vec{a}(\uparrow\uparrow, \uparrow\downarrow, \uparrow 0)\vec{v}$	$\vec{a} \uparrow\uparrow \vec{v}$	$\vec{a}(\uparrow\uparrow, \uparrow\downarrow)\vec{v}$	$\vec{a}(\uparrow\uparrow, \uparrow 0)\vec{v}$	Other
Test 1	82	37%*	59%	NA	NA	4%
Test 2	40	35%*	28%	20%	8%	9%
Test 3	119	23%*	48%	17%	12%	0%

Note on Notation. All questions provide information about one of three quantities (force, velocity, or acceleration) and ask about another. $\vec{A} \to \vec{B}$ means that information is given about $\vec{A}$ and the student is asked about $\vec{B}$. $\vec{A} \uparrow\uparrow \vec{B}$, $\vec{A} \uparrow\downarrow \vec{B}$, $\vec{A} \uparrow 0\ \vec{B}$ mean respectively that given information about $\vec{A}$ the student responds that $\vec{B}$ can be in the same direction as $\vec{A}$, opposite $\vec{A}$, or zero given $\vec{A}$ is nonzero. See the text for a more complete description.

3.2.4 Analysis of Differences in Student Responses due to Answer Choice Format.

Qualitative reasons for the development of the three test styles are discussed above. Here we compare the quantitative results of the three test styles. Table 3.5 presents the response percentages for three questions in each of the test formats. For these examples, the question posed was the same, and the only difference was in the answer choices. There are several points to consider from these results. First, it appears as though students taking Test 1 chose the correct answer more often than those taking Tests 2 or 3 ($p < 0.01$). The differences in the answering patterns between Tests 2 and 3 are smaller, though a chi-squared test for independence produces $p < 0.01$ for the $\vec{F} \to \vec{v}$ question. Overall, the scores for students taking Test 3 tended to be the lowest of the three tests. Second, for Tests 2 and 3, anywhere from 20% to 35% of the students answered either $(\uparrow\uparrow, \uparrow\downarrow)$, i.e. the two vectors could be parallel or antiparallel but not zero aka 'Cannot-be-Zero', or $(\uparrow\uparrow, \uparrow 0)$, i.e. the vectors could be parallel or zero but not antiparallel aka 'Cannot-be-Opposite' , choices which were not available on Test 1.

The most dramatic and perhaps the most interesting difference between the tests appears in the responses to the $\vec{F} \rightarrow \vec{a}$ question. Table 3.5 shows that almost 90% of the students chose the correct response in Test 1 and less than 10% chose "not enough information." Thus it would seem that students fully understood that net force and acceleration are necessarily parallel, in accordance with Newton's Second Law. However, when given other options, such as in Test 2 or Test 3, 30% fewer students chose the correct answer. Many students chose answers indicating that the net force on an object and its acceleration are not necessarily parallel, even after instruction on Newton's Second Law.

The $\vec{a} \rightarrow \vec{v}$ question shows a similar, though less dramatic, difference. On Test 1 almost 60% of the students incorrectly responded that acceleration and velocity were necessarily parallel, but on Tests 2 and 3 only between 30% and 50%, responded that they must be parallel.

A possible explanation for the difference between Test 1 and Tests 2 and 3 is that the presence of extra answer choices helped remind students of other options for the acceleration/velocity relationship while it incorrectly made them unsure of the force/acceleration relationship. We have few interviews with students taking Test 1, so we can only speculate about the confidence they have in their answers. In interviews with students taking Test 2, students occasionally chose "sometimes" even though they did not have a good reason in mind. This may account for this test's greater shift of the response from aligned, $\uparrow\uparrow$, into the other models.

In addition, there is a parallel effect in the $\vec{F} \rightarrow \vec{v}$ question in Table 3.5. There was a large shift in the percentage of students answering $\vec{F} \uparrow\uparrow \vec{v}$, aka 'Misconception', on Test 1 into $(\uparrow\uparrow, \uparrow\downarrow)$, aka 'Cannot-be-Zero', and $(\uparrow\uparrow, \uparrow 0)$, aka 'Cannot-be-Opposite', on Test 2. However, there was almost no change in $\vec{F} \uparrow\uparrow \vec{v}$ answering in Test 3.

This difference is believed to be an artifact of a new wording used in place of "net force" on Test 3. The new wording seemed to change the pattern of responses from ones consistent with net force and velocity not necessarily being parallel to ones where net force and velocity are necessarily parallel. However, as discussed below in 'Changes to Question Wording,' Test 3's response percentages are believed to be more representative of students' understanding of force and velocity.

3.2.5 Discussion of Changes to Question Wording

In addition to the development of the answer choice format, there was development of the text of the posed questions. The most significant change in question format involved the term "net force." Interviews revealed that students often misinterpret the term "net force." For example, if a question reads "*The net force is to the right. Which way is the velocity?*" a student might respond by saying that the velocity could be left or right as long as there is another force, in addition to the net force, pushing parallel to the velocity. Consequently, we revised test questions to reduce the use of "net force," replacing it with longer phrases such as, "there may be several forces but the forces to the left are greater." Interviews showed that students did not have difficulties understanding this question format, and there was no evidence that they interpreted the question differently than what was intended.

3.2.6 Effect of Story Context on Responses

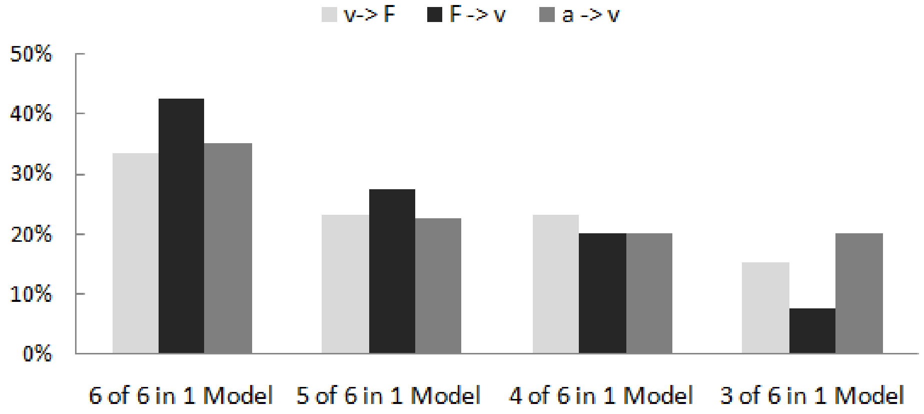
One significant threat to the validity of a particular item is its potential sensitivity to construct-irrelevant changes to the item. Thus far, we have only addressed issues of potential sensitivity to the item structure and format. Here we would like to address the issue of potential sensitivity to the story context of the item. For example, a force and motion question about a playground ball might be regarded differently by the student compared to an analogous question (from the perspective of the expert) about satellites in space. In order to limit the test to a reasonable length, the FVA test has at most two different story contexts for each question category, $\vec{F} \to \vec{v}$, $\vec{a} \to \vec{v}$, $\vec{v} \to \vec{F}$, etc. Therefore if the effects of story context are significant, this could severely limit the generalizability of any conclusions based on student response patterns in the FVA test. We constructed a series of tests and analyzed the results in two ways to investigate the possibility that our results were simply an artifact of story context.

A multiple context, fixed question type, construct validation quiz

We constructed and administered three separate "multiple context" tests to assess consistency of responses across a variety of story contexts for each of the three major question categories with which students have the most difficulty. Specifically, each multiple context test consisted of ten questions, six of which were all either $\vec{a} \to \vec{v}$, $\vec{F} \to \vec{v}$, or $\vec{v} \to \vec{F}$, and, for variety, four of which were $\vec{a} \to \vec{F}$ or $\vec{F} \to \vec{a}$ questions. Students were randomly assigned to complete one of the three multiple context tests, with 40 students in the standard mechanics course per test. We analyzed the results in two ways.

First, we analyzed the data to determine whether there were consistent within-student response patterns for a given question category. We found that on average across the three tests (see Figure 3.1), 37% of students consistently (within-student) chose the same answer

Figure 3.1: Percentage of students in the Standard Mechanics Course who responded using only one model – either Correct, Cannot-be-Zero, Cannot-be-Opposite, or “Misconception” – for 3, 4, 5 or 6 out of the 6 questions of the same question type – $\vec{v} \rightarrow \vec{F}$, $\vec{F} \rightarrow \vec{v}$, or $\vec{a} \rightarrow \vec{v}$ – on the “multiple context” tests. Note that a majority of students used only one model for 5 or 6 out of 6 questions.



choice for all six of the questions, and 61% of students answered at least five of the six questions with the same answer choice (within-student).

It is also worth noting that each of the major answer choice “models” corresponding to Correct, “Misconception”, Cannot-be-Zero, and Cannot-be-Opposite were consistently answered on all or 5 out of six questions by at least some students. This suggests that these four answer choices were not just random distractors that were occasionally attractive to the student for certain question contexts; rather, they were consistently chosen as that student’s model for that question type.

In contrast, on the regular FVA test, only 3% of students consistently (within-student) answered all six of the $\vec{a} \rightarrow \vec{v}$, $\vec{F} \rightarrow \vec{v}$, and $\vec{v} \rightarrow \vec{F}$ questions on the FVA test with the same answer choice and only 24% use the same answer choice on five of these six questions. Overall, these results suggests that for a given question type, for a variety of story contexts, within-student responses follow a specific model, such as Cannot-be-Opposite, but students do not necessarily use this model for other question types.

Second, we compared the answering patterns in the multiple context tests with the answering patterns in the FVA test to determine whether the response patterns for corresponding question types agree. Table 3.6 reports the average percentages and standard errors across the 6 different questions for each test and compares it to the averages from the two questions used for each question type on the FVA test. Both inspection of Table 3.6 and a chi-squared test for independence reveals that there are no significant differences between answer patterns on the multiple context tests that focus on one question category, and the

Table 3.6: Comparison of average response choice percentages for the FVA test and the "multiple context" tests (which include six different story contexts) for each Question Type. Reported are averages ± standard error between the questions.

Question Type	Correct	Cannot-be-Zero	Cannot-be-Opposite	"Misconception"
$\vec{v} \rightarrow \vec{F}$	22±2.4%	7±1.7%	7±1.1%	60±2.9%
FVA test	15±2.7%	3±1.6%	11±2.9%	63±4.4%
$\vec{F} \rightarrow \vec{v}$	21±3.5%	12±1.4%	7±1.9%	58±4.9%
FVA test	19±3.1%	13±3.1%	6±1.8%	56±4.6%
$\vec{a} \rightarrow \vec{v}$	38±1.9%	24±2.0%	4±1.1%	34±4.7%
FVA test	30±3.4%	22±3.9%	8±2.4%	35±4.4%

answer patterns on corresponding questions in the FVA test ($\chi^2(3) = 3.47, p = 0.325$ for $\vec{v} \rightarrow \vec{F}$ questions, $\chi^2(3) = 0.11, p = 0.991$ for $\vec{F} \rightarrow \vec{v}$ questions, and $\chi^2(3) = 1.05, p = 0.789$ for $\vec{a} \rightarrow \vec{v}$ questions).

In sum the results of both kinds of analysis of the focused tests reveal that the *averaged* FVA test responses for each question type are relatively insensitive to story context and in that sense the results are fairly generalizable.

A fixed context, multiple question type, construct validation quiz

It could be argued that a test which covers only one category, $\vec{a} \rightarrow \vec{v}$, $\vec{v} \rightarrow \vec{F}$, or $\vec{F} \rightarrow \vec{v}$, may elicit different models from students than a test that covers many different categories. So a second set of three tests to control for context were developed. These tests held the question context fixed as the FVA test context for that question but changed the question type. For example, on each test the first question was about a boat on a lake, the context of #1 on the FVA test, but different versions had slightly different wordings so that on one test the questions might be $\vec{v} \rightarrow \vec{F}$ but on another it might be $\vec{a} \rightarrow \vec{v}$. (1. At exactly 2:31PM, a boat is moving to the north on a lake. Which statement best describes the forces on the boat at this time? or 1. At exactly 2:31 PM, a boat is accelerating to the north on a lake. Which statement best describes the motion of the boat at this time?.) This type of test does not allow for an in depth look at the consistency of student responses across different contexts since each student receives at most three questions for each question type, but it provides a way to control for context while giving students a test which is almost identical to the FVA test becuase it has a mix of question types together and has the same contexts seen on the FVA test.

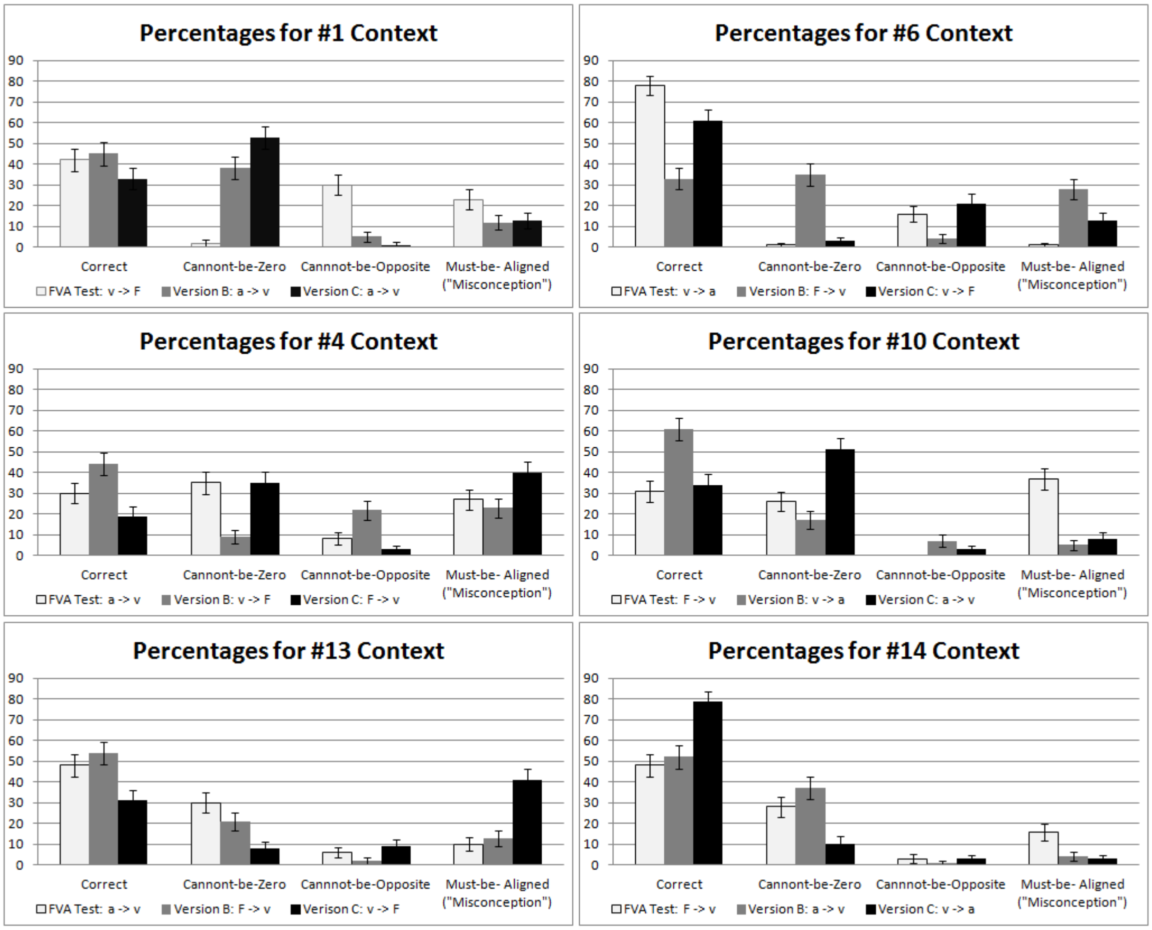
These tests were given to approximately 80 students per test version. The 240 students were randomly assigned to each test version and all took the test during the same week

Table 3.7: Comparing the FVA test averages for each Question Type, 2 questions, to each Mixed Question Test. (Honors Calculus Mechanics students, N = 240)

Question Type	Number of Questions	Correct	Cannot-be-Zero	Cannot-be-Opposite	"Misconception"
$\vec{v} \to \vec{F}$	5	48±6.6%	7±1.1%	16±3.0%	$25 \pm 5.6\%$
FVA test	2	38±5.6%	4±2.8%	26±6.4%	29±7.7%
$\vec{F} \to \vec{v}$	4	$31 \pm 8.3\%$	$34 \pm 4.6\%$	$3 \pm 0.5\%$	$28 \pm 5.5\%$
FVA test	2	$40 \pm 8.5\%$	$27 \pm 1.4\%$	2±1.5%	27±10.5%
$\vec{a} \to \vec{v}$	4	$38 \pm 6.6\%$	$46 \pm 4.9\%$	$2 \pm 1.1\%$	$11 \pm 2.6\%$
FVA test	2	$39 \pm 9.0\%$	$33 \pm 2.5\%$	$7 \pm 1.0\%$	$19 \pm 8.5\%$
$\vec{v} \to \vec{a}$	2	$70 \pm 9.0\%$	$14 \pm 3.5\%$	$5 \pm 2.0\%$	$4 \pm 1.0\%$
FVA test	2	$75 \pm 3.5\%$	$3 \pm 2.0\%$	$18 \pm 2.0\%$	$2 \pm 0.5\%$
$\vec{F} \to \vec{a}$	2	$85 \pm 3.5\%$	$5 \pm 3.5\%$	$6 \pm 0.5\%$	$4 \pm 2.0\%$
FVA test	1	$88 \pm N/A\%$	$4 \pm N/A\%$	$1 \pm N/A\%$	$5 \pm N/A\%$
$\vec{a} \to \vec{F}$	2	$82 \pm 3.0\%$	$8 \pm 1.5\%$	$3 \pm 0\%$	$3 \pm 2.5\%$
FVA test	1	$88 \pm N/A\%$	$5 \pm N/A\%$	$4 \pm N/A\%$	$1 \pm N/A\%$

near the end of the honors calculus mechanics course. Table 3.7 summarizes the effects of question context and type. Considering this data, the average responses on the FVA test are not as close to the total averages for the different tests as before. However, they are not highly dissimilar either, 10% being the largest difference. Also, when the response percentages for different question types are considered, it is clear that the effects of the question type on student responses are larger than the effects of context. So, even when questions of the same type are not averaged, and especially when they are, it is usually easy to tell what type a question belonged to by looking at the average student responses. $\vec{F} \to \vec{a}$ and $\vec{a} \to \vec{F}$ questions are distinguishable from the other question types by their large correct response percentages. $\vec{v} \to \vec{a}$ also are distinguishable by their correctness and in general are differentiable from the $\vec{F} \to \vec{a}$ and $\vec{a} \to \vec{F}$ by their lower correctness and greater use of the responses Cannot-be-Zero, $(\uparrow\uparrow, \uparrow\downarrow)$ and Cannot-be-Opposite, $(\uparrow\downarrow, \uparrow 0)$. $\vec{v} \to \vec{F}$ is then distinguishable from $\vec{F} \to \vec{v}$ and $\vec{a} \to \vec{v}$ by their higher responses in Cannot-be-Opposite where $\vec{F} \to \vec{v}$ and $\vec{a} \to \vec{v}$ have much higher Cannot-be-Zero responses. Finally, the $\vec{F} \to \vec{v}$ and $\vec{a} \to \vec{v}$ questions, which are the hardest to distinguish, can on average be separated by the $\vec{a} \to \vec{v}$ questions' lower misconceptions and higher Cannot-be-Opposite response. (This is only an average distinction and is not as clear a signal as the other differences.) Thus, while question context does play a role and is important to consider when looking for smaller differences in student responses, it seems fair to use the two FVA questions for each type of question. (See Figure 3.2 for a visual representation of the differences in responses for different question types but fixed contexts.)

Figure 3.2: This shows how the response patterns change with a fixed context but changing question type. The question number is that on the FVA test so the context can be seen by looking in Appendix A for that question number.



3.2.7 Student Responses During Think-Aloud Interviews

In order to assess how students were interacting with the FVA test, 25 formal think-aloud interviews were done. (A “think allowed interview” is an interview where students are prompted to talk through their thought process as they go through the test.) There was minimal dialog between the interviewer and the interviewee while he or she was initially working through the problems. After students had worked through the whole test, the interviewer then went back through and asked students for more detail in their answers on a few selected questions to test students more in their confidence and reasoning. In this case the interviewer would specifically question students. Asking, for example, “You said a. and c. were possible but what is wrong with answer choice b?” There are three main points to note from the interviews. First, in almost all questions, with almost all students, the answer which a student circled was consistent with their apparent thought process going through the question. Second, students often used a ruling out process when picking a response so that if they chose a partially correct response it was because they ruled out a, b, or c as a possibility. Thus, the terminology ‘Cannot-be-Zero’ and ‘Cannot-be-Opposite’ for certain student responses does seem to accurately represent a student’s thoughts or model because the student chose this response by deciding that Zero or Opposite was not an acceptable option. Third, the reasoning process used was often local to that question only and changed between questions. For example, a student might be correct on #13 a $\vec{a} \rightarrow \vec{v}$ question with acceleration to the left and then move on to #14 a $\vec{F} \rightarrow \vec{v}$ question with force to the left and choose a partially correct model missing the turnaround point they had just considered on #13. When the interviewer took students back through the questions, students would sometimes conceded the possibility of other responses or realize their inconsistency, but again this would often be a local concession only for that question context or category. A few examples of students comments are given in Table 3.8 with their responses.

Table 3.8: An ad hoc selection of student comments during think-aloud interviews.

In reference to #4, a $\vec{a} \rightarrow \vec{v}$ question, “[Went to pick d] No if it were going down the hill it would be accelerating that way. I know it is not c [zero velocity option] because then it would not be accelerating so that means [crossed out e and f and circles a].”
In reference to #6, a $\vec{v} \rightarrow \vec{a}$ question, “It could be speeding up or slowing down so a and b are ok. [long pause] I guess it could be constant too. [circles f].”
In reference to #13, a $\vec{a} \rightarrow \vec{v}$ question, “The forces would be equal in order for c, [not moving] so [picks a].”
In reference to #16, a $\vec{v} \rightarrow \vec{F}$ question, “The force of friction would have to be overcome for it to be moving that way [picks a].”

3.2.8 Assessment of the Students’ Confidence in their Responses on the FVA Test

Also, there is some data to suggest that students were fairly confident in their responses. Instead of given the questions as multiple-choice, four of the FVA questions were given to students with these instructions: “For each item on the questionnaire, you are given 100 ‘confidence points’ that you will allocate among the answer options. Read each answer option and decide how likely it is to be the correct answer. Then, spread out the 100 confidence points among the answer options in order to best reflect your confidence in each option.” The students were then given two examples and told to ask the proctor if they did not understand. The students all seemed to be relatively comfortable with this novel response system, and only one student out of 111 students had points which did not add to 100. When you look at students responses, students used all 100 points on one answer 62% of the time. 27% of students always used all 100 points on one answer, but 23% use it on three of the four questions and 24% on two of the four. So, it follows that students show high confidence in their answers to these fva questions, i.e. putting 100 in one answer, and showing high confidence is not just hesitation to use multiple responses or eagerness to simplify the problem (because many students using 100 on one question did not use 100 on another). The average confidence of the correct answer for each question was: $\vec{v} \to \vec{F}$, 24.6%; the $\vec{a} \to \vec{F}$, 70%; the $\vec{a} \to \vec{v}$, 44%; $\vec{F} \to \vec{v}$, 42%.

3.3 A Brief Summary of the Validation Steps Taken in the Previous Section

Because the previous section is long and contains many important points, a brief summary of the validation methods is provided.

1. The questions format, e.g. given the direction of $\vec{v}$ what is $\vec{F}$, was created from several stages of prior interview and testing data. 2. The response choices available to the questions went through three different stages before settling on the current six or seven option multiple-choice items, which we believe, from student interviews and analysis of a large set of response data, is the best of the three. 3. The actual content of the FVA items' wordings were validated by a few PER faculty and graduate students. 4. The susceptibility of students' responses to the superficial question context was then validated from a student perspective. This was done by fixing the question type and giving multiple contexts, and by fixing the contexts and giving multiple question types. The results of these tests show that while students do base their responses on superficial contexts, there is a much larger importance placed on question type. 5. Several think-aloud interviews were done to insure that students were choosing an answer that agreed with their thought process and knowledge. 6. A subset of the FVA test questions were given to students to test their confidence in the answer they picked and to see if there was a large indication that correct or incorrect students were substantially more confident. The results were that students were in general very confident in their answers whether right or wrong.

On face value the FVA test looks simple and perhaps a little contrived. However, the various validation steps taken above ensure that students' answers to the test items are replicable, reliable, purposeful, and relatively unaffected by factors other than the question type and student level.

Table 3.9: Outline of test administration and description of courses and populations

Course Name	Quarter /Year	Number of Students	Type of Testing Done	Description of Course and Population (All Mechanics w/ Calc.)
Standard	2005-2008	> 500	Test Development	Students' activities include show work, free response, interviews, and multiple-choice data.
Standard	Wi/2008 Wi/2009	117 111	Post-FVA Post-FVA FCI	Traditional lecture and recitation format. Majority are first or second year science or engineering students.
Honors Engineers	Au/2008 Au/2009	86 230	Post-FVA Pre-FVA Post-FVA	Lecture and recitation format with some PER reforms included, e.g. clickers in lecture and group work problem solving in recitation. All first year honors engineering students.
Honors Physics Majors	Au/2009	49	Pre-FVA Post-FVA	Traditional lecture. Majority are first year honors physics majors.
Second Year Physics Majors	Wi/2009	65	Post-FVA	Traditional lecture. This is the second quarter of the second year physics majors sequence. Majority are second physics majors, with some engineering majors as well.

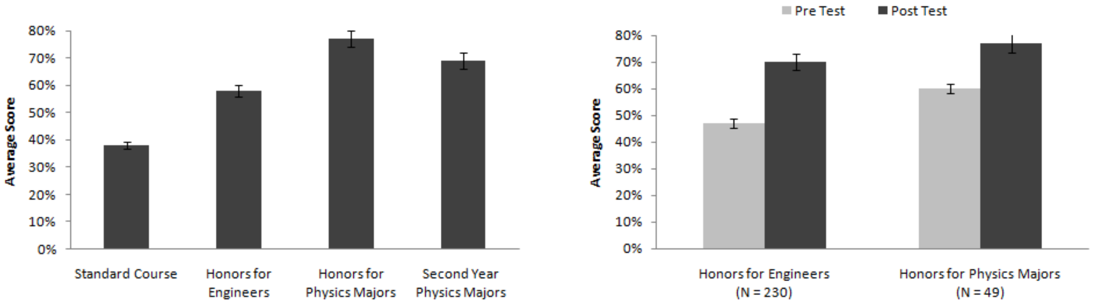
3.4 Other Measures of Reliability and Validity of the FVA Test

We reported above on the content and construct validity of the items - including the answer choice format, the question wording, and the story context of the questions - which was supported through several stages of interview and testing-based modifications. We report here on other measures of validity and reliability, including correlations of the test with other measures such as course level, course grade, and the FCI (all measures of student knowledge) as well as various reliability measures of the instrument.

3.4.1 Increases in FVA Score with Increasing Course Level and Instruction

We administered the FVA test post instruction to four different class levels - standard calculus mechanics, honors calculus mechanics for engineering majors, honors calculus mechanics for first-year physics majors, and mechanics for second year physics majors. See Table 3.9

Figure 3.3: Summary of FVA scores. The graph on the left shows a trend of increasing post test score with increasing class level for four different calculus mechanics courses ranging from first-year introductory to second-year physics majors. The graph on the right shows pre and post test scores for the two different honors introductory mechanics courses.



for a description of each course and the enrolled students. Figure 3.3 reveals that the average score on the FVA test tended to increase with course level such that the average score of the second year physics majors was 0.9 standard deviations above that of the standard mechanics course. One exception was the higher score for the first honors physics majors compared to the second year majors course. This difference may be due to slightly different population, since the second year course enrolls some students who are not physics majors and not honors students. Note also that the increase in average score with course level was not an artifact of an increase in score on a small number of questions; rather the increase was spread among all the question types. Similarly, for all questions, the percentage of "misconception" responses decreased as class level increases. Thus, the average student in the higher level class did better on the FVA test, by both decreasing his or her misconception responses and increasing his or her correct responses, than the average student in the lower level course.

In addition to measuring post test score differences between different course levels, we also administered pre and post tests to measure any changes in scores within a given course. For two courses, Figure 3.3 reveals a within-student pre-post test increase in correct responses. In particular, a paired t-test reveals significant gains from pre to post for the honors engineering introductory calculus mechanics class ($t(229) = 16.50$, $p < 0.001$, effect size $d = 0.95$) and for the honors physics majors introductory calculus mechanics class ($t(48) = 7.13$, $p < 0.001$, effect size $d = 0.71$). Furthermore, it is interesting to note that we pre and post tested students in the standard calculus mechanics course with very similar, earlier version of the FVA test and found no significant difference between pre and post test averages. This suggests that for the standard calculus course there may be little gain or evolution in the concepts the FVA test assesses. This lack of significant gain in the

Table 3.10: Summary of FVA test statistics

Course	N	Post-test FVA avg. ± sd	KR 20	Correlation: Grade-Score	Correlation: Grade-Misc. Score
Standard Mechanics	228	$37.6 \pm 17.7\%$	0.721	0.311	-0.286
Honors Engineers	86	$57.9 \pm 21.4\%$	0.801	0.428	-0.508
Second Year Majors	65	$69.2 \pm 22.8\%$	0.849	...	...
Honors Engineers[1]	230	$70 \pm 24.0\%$	0.872	0.464	-0.436

traditional course is consistent with previous research in student conceptual understanding of force and motion (Halloun and Hestenes (1985), Thornton and Sokoloff (1998), Clement (1982), Viennot (1979), and Hestenes, Wells, and Swackhamer (1992)).

3.4.2 Psychometric Properties and Correlations with Course Grade, Course Level and FCI.

Table 3.10 reports overall tests statistics for several class levels and Table 3.11 reports individual item statistics for several class levels as well. The FVA test has reasonably high statistical reliability scores (KR $20 = 0.7 - 0.85$), indicating that correct responses to all of the FVA items are somewhat correlated. Furthermore, there were moderate (0.3-0.4) correlations between FVA score and final course grade. Likewise, the FVA "misconception" responses were negatively correlated with grade in the class and were on average about -0.4. These correlations tended to be a larger for the higher level classes. This data is consistent with the FVA test assessing a portion of the skills necessary to do well in the class. The correlations between FVA score with course level and the gains in pre-post scores suggest that the FVA-final grade correlation is not simply caused by something more general such intelligence but rather by gained knowledge of force, velocity, and acceleration.

Furthermore, we administered the Force Concept Inventory (FCI) to the winter 2009 calculus mechanics class in order to compare the FVA test to a standard benchmark and further assess the validity of the FVA. The FCI is a multiple-choice concept inventory developed to assess understanding of basic concepts in force and motion. It has been widely used and generally accepted as a standard and reasonably reliable assessment, and has also been used to evaluate instructional interventions at the high school and university level (Hake (1998); Halloun and Hestenes (1985)). We found a relatively strong correlation between FVA score and FCI score ($r = 0.569$), while the correlation of FCI with final grade was 0.387, about the same as the FVA test- final grade correlation.

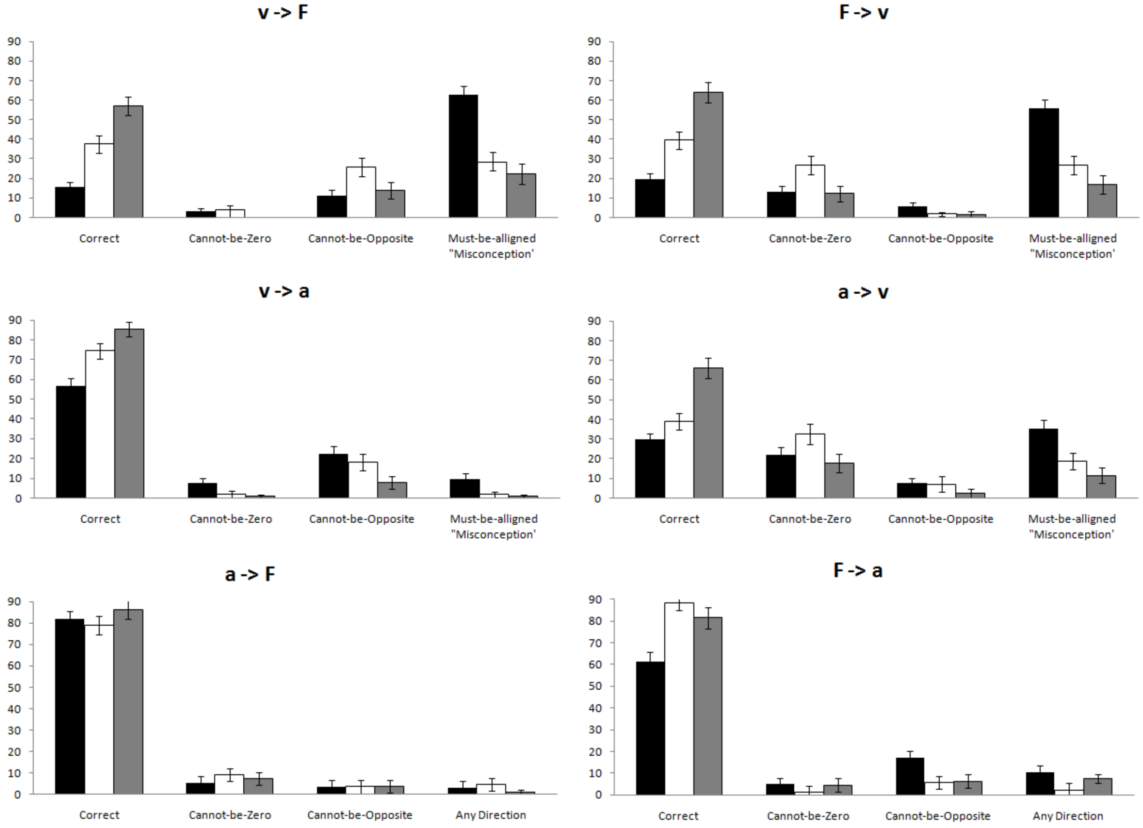
In sum, the positive correlations of FVA score with other measures (or expectations) of force and motion conceptual understanding such as course level, pre/post, grade, and FCI score help to support the criterion validity of FVA test.

Table 3.11: Summary of individual FVA test item statistics for three course levels. Reported are the response percentages for each available response on all 17 items. Correct responses choices are in bold. P-Bi stands for point biserial correlation coefficient.(Note that question 9 required students to "circle all that apply". Thus response percentages reflect the percentage of students circling that response. The correct response was circling a, b and d.)

		Standard Introductory Mechanics Average Total Score FVA Test $= 37.6 \pm 17.7\%$								Honors Mechanics for Engineers Average Total Score FVA Test $= 57.9 \pm 21.4\%$								Second Year Mechanics for Physics Majors Average Total Score FVA Test $= 69.2 \pm 22.8\%$							
Item	Ques. type	P-Bi	a	b	c	d	e	f	g	P-Bi	a	b	c	d	e	f	g	P-Bi	a	b	c	d	e	f	g
1	$\vec{v} \to \vec{F}$	0.377	56	4	4	4	12	**21**	...	0.427	23	1	1	2	30	**42**	...	0.486	22	3	3	0	12	**60**	...
2	$\vec{v}, \Delta\vec{v} \to \vec{a}$	0.151	**88**	2	0	7	3	0	...	-0.033	**93**	1	0	2	1	2	...	0.075	**92**	0	0	8	0	0	...
3	$\vec{F} = 0 \to \vec{v}$	0.317	1	1	82	1	**15**	...	...	0.628	0	0	47	2	**51**	...	...	0.475	0	0	29	8	**63**	...	...
4	$\vec{a} \to \vec{v}$	0.299	41	2	0	23	10	**24**	...	0.291	27	0	0	35	8	**30**	...	0.523	17	0	0	19	5	**60**	...
5	$\vec{f} \to \vec{a}$	0.162	**61**	3	3	5	17	11	...	0.196	**88**	1	1	1	6	2	...	0.562	**82**	0	0	5	6	7	...
6	$\vec{v} \to \vec{a}$	0.410	11	1	1	6	19	**62**	...	0.546	1	1	2	1	16	**78**	...	0.421	2	0	2	0	8	**90**	...
7	$\vec{v}, \Delta\vec{v} \to \vec{F}$	0.312	42	**39**	4	6	4	2	4	0.284	16	**78**	0	1	0	2	2	0.529	9	**72**	3	0	3	3	9
8	$\vec{a} \to \vec{v}, \Delta\vec{v}$	0.427	19	2	1	23	**43**	12	...	0.544	20	0	1	9	**61**	9	...	0.507	5	0	0	6	**69**	20	...
9	$\vec{v} \to \vec{F}$-list	0.279	**95**	**93**	57	**93**	0	...	...	0.275	**100**	**100**	12	**99**	0	...	...	0.522	**100**	**100**	16	**100**	0	...	...
10	$\vec{F} \to \vec{v}$	0.402	62	2	0	11	10	**14**	...	0.562	37	4	1	26	0	**32**	...	0.634	23	2	2	17	2	**55**	...
11	$\vec{a} \to \vec{F}$	0.163	**84**	4	0	5	4	3	...	0.238	**79**	1	1	9	4	5	...	0.427	**86**	0	2	5	5	3	...
12	$\vec{v} = 0 \to \vec{F}$	0.437	2	**30**	60	1	1	5	1	0.539	1	**60**	30	0	0	4	4	0.516	3	**68**	19	0	0	8	3
13	$\vec{a} \to \vec{v}$	0.405	15	20	0	21	2	5	**36**	0.527	11	2	2	30	6	1	**48**	0.673	6	0	2	17	3	0	**72**
14	$\vec{F} \to \vec{v}$	0.632	4	52	1	16	2	**25**	...	0.625	4	16	0	28	4	**48**	...	0.704	0	11	3	8	2	**76**	...
15	$\vec{a} \to \Delta\vec{v}$	0.199	8	1	**91**	...	...	...	...	0.169	5	1	**92**	...	...	...	...	0.337	2	0	**98**	...	...	...	...
16	$\vec{v} \to \vec{F}$	0.387	73	4	0	2	11	**10**	...	0.526	34	4	1	6	21	**34**	...	0.427	24	0	3	2	16	**56**	...
17	$\vec{v} \to \vec{a}$	0.380	8	2	4	9	23	**54**	...	0.532	2	0	1	5	20	**71**	...	0.585	0	0	6	2	8	**84**	...

3.5 Analysis of FVA Test Results

Figure 3.4: Mean student response percentages for all six conditional relation question types and for three course levels. Black: standard calculus mechanics, N = 228. White: honors for engineers calculus mechanics, N = 86. Grey: second year physics majors, N = 65. Error bars are ± 1 standard error.



The previous sections have focused on the development and validation of the FVA test. The rest of this part focuses on analyzing pre and post FVA test data from students enrolled in different levels of physics courses. This analysis will allow for an investigation of possible structure and hierarchy of student understanding of the relations between the directions of force, velocity, and acceleration as well as an investigation of evolution of this understanding.

The FVA test was administered to students either during an extra session (counted as part of the total homework grade with full credit for participation) in which students came to our lab to complete the test, or as an in-class activity completed during the regular laboratory or lecture for the course. In both situations, students had plenty of time to finish the quiz and appeared to take the activity seriously.

3.5.1 General Response Patterns for Different Course Levels

Figure 3.4 presents average students response patterns for all six question types for three class levels. There are four important observations about the response patterns presented in Figure 3.4, as described below.

3.5.2 Evidence of intermediate levels of understanding

There was a small but significant fraction of students (20%-30%) who displayed intermediate levels of understanding of relations between the directions force, velocity, and acceleration as suggested by their choice of partially-correct responses. By “partially correct”, we mean that the response included some physically valid possibilities not considered in the common misconception response. For example, for a $\vec{v} \rightarrow \vec{F}$ question, the common misconception response assumes that the inferred force must be non-zero and aligned with the velocity. In contrast, the somewhat common response choice that includes the possibilities that the net force is aligned *or* is zero (i.e. the “Cannot-be-Opposite” model) is more accurate than the common misconception response, and could be considered an intermediate, partially correct response. As seen in Figure 3.4, intermediated levels of understanding occurred in all of the conditional relations between force, velocity, and acceleration.

Interviews with students further revealed that those choosing partially correct answers were often confident about their answers, for example allowing for the possibility that a moving object can have a net force aligned with the motion or a zero net force, but certain that the net force cannot be opposite the motion.

3.5.3 Asymmetry in response patterns between $\vec{x} \rightarrow \vec{y}$ and $\vec{y} \rightarrow \vec{x}$

Figure 3.4 also reveals two significant asymmetries in response patterns between a given conditional relation $\vec{x} \rightarrow \vec{y}$ and its converse $\vec{y} \rightarrow \vec{x}$. First, there were often asymmetries in scores, depending on the course level and the question types. For example, while there were only small differences between the $\vec{v} \rightarrow \vec{F}$ scores and $\vec{F} \rightarrow \vec{v}$ scores (effect sizes less than 0.12 standard deviations), there were significant differences between the $\vec{a} \rightarrow \vec{v}$ and $\vec{v} \rightarrow \vec{a}$ scores for the standard calculus-based physics course [2] (31% versus 57% correct, paired t-test, $t(110) = 5.78, p < 0.001$, effect size $d = 0.54$.) and for the honors physics majors course (39% versus 74% correct, paired t-test, $t(85) = 7.49, p < 0.001$, effect size $d = 0.74$.) Clearly most students correctly understand that a moving object can have an acceleration

---

[2] Wi2008 students received a older version of the fva test which did not have $\vec{v} \rightarrow \vec{a}$ questions. Thus, they were excluded from this analysis.

in any direction or zero acceleration, but many students also believe incorrectly that an accelerating object must be moving in the direction of its acceleration.

Another perhaps more surprising asymmetry in scores occurs for the $\vec{F} \rightarrow \vec{a}$ versus $\vec{a} \rightarrow \vec{F}$ questions. While there were no significant differences in responses for the first and second year physics majors courses (perhaps because they were answering at ceiling), there was a difference in responding to these two questions categories for the students in the standard calculus-based course. Specifically, the average score for the $\vec{F} \rightarrow \vec{a}$ question was 21% lower compared to the 82% score for the $\vec{a} \rightarrow \vec{F}$ question, which is a significant difference (paired t-test, $t(227) = 5.50, p < 0.001$, effect size d = 0.36.). Interestingly, a similar asymmetry in scores occurs in pretest results for the first year physics majors course (58% correct for $\vec{F} \rightarrow \vec{a}$ and 84% correct for $\vec{a} \rightarrow \vec{F}$, paired t-test, $t(227) = 6.94, p < 0.001$, effect size d = 0.46.), but not for the posttest. This asymmetry in responding might be considered somewhat surprising since the relation $\vec{F} = \mathrm{m}\vec{a}$ is a central relation in these physics courses (and readily recited by all students), but the results of the FVA test demonstrate that students in lower level courses often did not consider the conditional relationships between (net) force and acceleration to be symmetric.

A second significant kind of response asymmetry occurred in the *kinds* of intermediate, partially correct responses chosen by students. For example, for $\vec{v} \rightarrow \vec{F}$, the partially correct response chosen tended to be Cannot-be-Opposite, while the $\vec{F} \rightarrow \vec{v}$, the partially correct response chosen tended to be Cannot-be-Zero. Therefore, it appears as though if it is given that an object is moving, students more readily accepted that it may not have a zero net force acting on it rather than accepting that it could have an opposing net force acting on it. On the other hand, if it is given that an object has a net force acting on it, students more readily accepted that it can move opposite the net force, rather than accepting that it is not moving at all. For $\vec{v} \rightarrow \vec{a}$ versus $\vec{a} \rightarrow \vec{v}$ questions, there are significant differences in all response choices, with the $\vec{v} \rightarrow \vec{a}$ questions tending to be answered correctly significantly more often. Similar to the $\vec{v} \rightarrow \vec{F}$ versus $\vec{F} \rightarrow \vec{v}$ questions, students tended to choose the Cannot-be-Opposite partially correct response for $\vec{v} \rightarrow \vec{a}$ and Cannot-be-Zero for $\vec{a} \rightarrow \vec{v}$ questions.

3.5.4 Other differences in scores between question types

In addition to differences in scores between a given conditional relation and its converse, there were also significant differences between other combinations of relations. For the standard calculus-based physics course the question types can be ranked as $\vec{v} \rightarrow \vec{F}$, $\vec{F} \rightarrow \vec{v}$, $\vec{a} \rightarrow \vec{v}$, $\vec{v} \rightarrow \vec{a}$, $\vec{F} \rightarrow \vec{a}$, $\vec{a} \rightarrow \vec{F}$, in order of increasing average score. The scores varied the most for the standard calculus-based physics course, but there were similar but reduced differences for the higher level courses. We will investigate the possible hierarchy of understanding of these relations in more detail in the section *"Investigating Possible*

Hierarchies".

3.5.5 Difference in course levels: evidence of evolution of understanding

While there were qualitative similarities between the patterns of the different course levels for each question type, there appears to be an "evolution" of the patterns from lower to higher course level. However, while percentage of correct responses increased as the class level increased, the change in misconception score between two classes was not always equal in magnitude to the change in correct score. This appears to have been caused by a significant fraction of students choosing the partially correct Cannot-be-Zero and Cannot-be-Opposite responses, depending on the course level. For example, comparing the standard mechanics course to the honors for engineers mechanics course, the decrease in misconception responses was greater than the increase in correct responses, and the difference was comprised of students choosing one of the partially-correct responses. Furthermore, when the difference between the honors and second year course is considered, it is apparent that the increase in correct responses is greater than the decrease in misconception responses, and the balance is comprised of a decrease in the partially-correct responses. These differences in response patterns between course levels suggest that a significant number of students evolved from an initial high level of misconceptions to the correct answer by passing through a partially-correct response "state", which indicates more knowledge than the common misconception but lacks the completeness of the correct response.

However, there is a danger in interpreting this data as evidence of evolution of understanding since it is cross-sectional rather than longitudinal, and sometimes represents different kinds of students (e.g. physics majors vs. engineering majors). Nonetheless, this data is consistent with the interesting possibility that students evolve though a partially-correct "state" on the path to fully understanding the relations between the directions of force, velocity and acceleration. We will investigate longitudinal data in the next section.

3.5.6 Pre & Post Responses: evidence of progression through intermediate levels

Pre and post FVA test data (i.e. longitudinal data), was analyzed in order to more closely investigate the evolution of student understanding of the relations between force, velocity, and acceleration. We were especially interested in determining whether the progression of student understanding involved passing through an intermediate, partially correct level of understanding. We administered the FVA test both pre and post instruction to 230 students in an honors calculus mechanics class for engineers and 49 students in a first year honors calculus mechanics for physics majors class. As mentioned in the reliability and validity section and presented in Figure 1, there were significant gains in the average scores for both

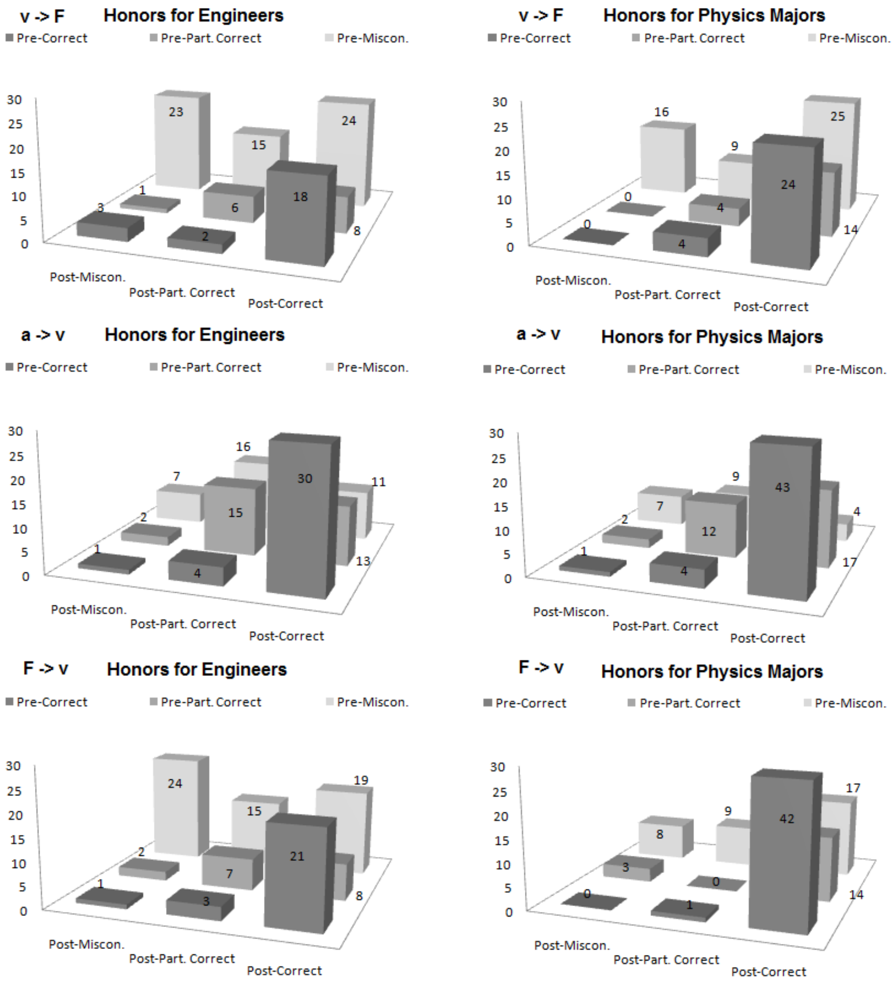


Figure 3.5: Within-student, pre versus post test response choice percentages for the 3 lowest scoring question types. Responses were categorized as either “Correct”, “Part. Correct” (partially correct response), or “Miscon.” (misconception-like response). The data presented is from the Honors Calculus Mechanics Course for Engineers, N = 230, and the First Year Honors Physics Majors Calculus Mechanics Course, N = 49. Note that roughly half of the students did not change their answer from pre to post test - these students are represented in the three diagonal columns from back left to front right of each graph. Also, note that a little less than half of students improved by moving into or out of a partially correct response, or directly from the misconception response to the fully correct response, - these students are represented in the three columns behind and to the right of the diagonal columns - and roughly 10% of students answered less correctly from pre to post - represented in the three columns to the front and left of the diagonal. (A few examples from the Honors for Engineers $\vec{v} \to \vec{F}$ plot. 23% of students responded with a Misconception on the pretest and a Misconception on the posttest. 15% of students responded with a misconception on the pretest and a Part. Correct, partially correct response, on the posttest. 8% of students responded with a Part. Correct, partially correct response, on the pretest and a Correct response on the posttest.)

classes. Perhaps more interesting, we examined within-student pre versus post test shifts in response choices for each item in the FVA test. Figure 3.5 presents a cross tabulation of within-student pre and post test responses on a select set of FVA test items for the two courses. There are two important observations about the data represented in these figures.

First, it is helpful to describe the general patterns of shifts in student answering. Considering the honors mechanics for engineers course for the $\vec{v} \rightarrow \vec{F}$, $\vec{F} \rightarrow \vec{v}$, and $\vec{a} \rightarrow \vec{v}$ questions on average 51% of students did not change their answers, 43% answered "more correctly" on the post versus the pre by either changing from the misconception response to either a partially correct response or to the correct response or changing from a partially correct response to the correct response, and conversely 6% answered less correctly. The results for the first-year physics majors course were somewhat similar, where 52% of students did not change their answers, 38% answered more correctly, and conversely 5% answered less correctly from pre to post test.

Second and more importantly, on average approximately 15% of students moved from the misconception response to a partially correct response and approximately 10% of students moved from a partially correct response to the correct response. These averages are representative of $\vec{v} \rightarrow \vec{F}$, $\vec{F} \rightarrow \vec{v}$, and $\vec{a} \rightarrow \vec{v}$ questions for both courses. This is to be compared with approximately 20% of students who moved directly from the misconception response to the correct response.

These results provide strong evidence that for many students the progression of student understanding involves passing through an intermediate, partially correct, level of understanding. Specifically, over half of the students who changed their answer changed either to or from an intermediate, partially correct, response.

3.6 Investigating Possible Hierarchies in Student Responses

In this section we are interested in investigating the question: "Does correctly answering a given conditional relation necessarily imply that another specific conditional relation was also answered correctly?" For example, does correctly answering $\vec{v} \rightarrow \vec{F}$ questions necessarily imply that $\vec{v} \rightarrow \vec{a}$ questions were also answered correctly? Note that while one might make reasonable physical arguments to answer this question from an expert point of view, we are first interested in this as strictly an *empirical* question.

If such patterns in answering do exist, then one can proceed to make inferences as to the causes of these patterns. There are some standard analysis practices, such as Guttman scaling or scalogram methods, for determining the hierarchical-like structures of items for a unidimensional instrument. In fact, a full item response theory analysis of the FVA test can be used to find such hierarchies. However, here we are interested in the hierarchical relation between a number of different dimensions, such as understanding $\vec{v} \rightarrow \vec{F}$ or $\vec{v} \rightarrow \vec{a}$ probed

Generic Contingency Table			
		Question y	
		Incorrect	Correct
Question x	Incorrect	a	b
	Correct	c	d

Hypothetical Example Data			
		$\vec{v} \to \vec{a}$	
		Incorrect	Correct
$\vec{v} \to \vec{F}$	Incorrect	50	25
	Correct	0**	25

**There are no counterexamples to the statement: "correctly answering $\vec{v} \to \vec{F}$ questions implies correctly answering $\vec{v} \to \vec{a}$ questions."

Table 3.12: A simple method for ruling out or finding supporting evidence for possible hierarchical structure in answering. Consider the generic contingency table above on the left. Here, for example, a = number of students answering both x and y incorrectly. If $c = 0$, $d \gg 1$ and $a \gg 1$, then this is consistent with the statement "answering x correctly implies answering y correctly" and the logical equivalent: "answering y incorrectly implies answering x incorrectly". One could reasonably also use the conditions, $d \gg c$ and $a \gg c$, since c will not be zero in practice due to random guessing etc. Furthermore, if these conditions on c are violated, then these statements are *disproved*, since c represents the number of counterexamples to these statements (and b represents the number of counterexamples to the converse of these statements). The table above on the right presents a hypothetical example for $\vec{v} \to \vec{F}$ versus $\vec{v} \to \vec{a}$ questions. In this case one can claim this hypothetical data is consistent with the statement "correctly answering $\vec{v} \to \vec{F}$ questions implies correctly answering $\vec{v} \to \vec{a}$ questions" and "incorrectly answering $\vec{v} \to \vec{a}$ questions implies incorrectly answering $\vec{v} \to \vec{F}$ questions." On the other hand, the relatively high counts in the "b" cell *disproves* the converse statement "correctly answering $\vec{v} \to \vec{a}$ questions implies correctly answering $\vec{v} \to \vec{F}$ questions."

by different items within the FVA test. Therefore, we will examine cross tabular results between a pairs of question types within the FVA test. A full Guttman scaling and/or item response theory analysis could also be informative from a more global perspective and is worth further study, but here we will focus on hierarchies within the six conditional relations of interest.

Table 3.12 provides an example of a simple method to rule out or provide supporting evidence for the existence of hierarchies in response patterns for pairs of question types. In this hypothetical example, *all* of students that answered $\vec{v} \to \vec{F}$ questions correctly also answered $\vec{v} \to \vec{a}$ correctly, but only *half* (25 out of 50) of the students that answered the $\vec{v} \to \vec{a}$ questions correctly answered the $\vec{v} \to \vec{F}$ correctly. This hypothetical data is consistent with the statement "correctly answering $\vec{v} \to \vec{F}$ questions necessarily implies correctly answering $\vec{v} \to \vec{a}$ questions" (the data is also consistent with the logically equivalent statement "incorrectly answering $\vec{v} \to \vec{a}$ questions implies incorrectly answering $\vec{v} \to \vec{F}$ questions"). Furthermore, this hypothetical data provides evidence to *disprove* the converse statement "correctly answering $\vec{v} \to \vec{a}$ questions implies correctly answering $\vec{v} \to \vec{F}$ questions", since 25 out of 50 students are counterexamples to this statement. Therefore, one can analyze pairs of question types in this manner to either provide evidence *disproving* or supporting (but not *proving*) the existence of a particular hierarchy in answering, namely that answering question type x correctly requires answering question type y correctly (but not the converse).

We analyzed within-student response patterns for all 15 possible pairs of question types on the FVA test (see Table 3.13 to 3.27) using the simple method shown in Table 3.12, and found that there were no cases in which there were *zero* counterexamples for the statement "answering relation x correctly requires answering relation y correctly". However, there were a number of cases in which there were a relatively small number of counterexamples.

These few counterexamples may be due to uninteresting causes such as random guessing. Thus, when there are only a relatively small number of counterexamples (rather than zero), this can still suggest the existence of a hierarchy in the answering pattern.

Using the constraint of a small number of counterexamples rather than zero counterexamples to indicate a hierarchy in answering, inspection of Tables 3.13 to 3.27 reveals a trend: for most pairs of relations if the average score of a question type x was significantly less than the score for question type y, then it is also the case that correctly answering question type x implied correctly answering question type y, but *not* the converse. More specifically, if a student correctly answered the question types with the lowest average scores (the "difficult" question types), namely $\vec{v} \rightarrow \vec{F}$ or $\vec{F} \rightarrow \vec{v}$, then most of the time this student also to correctly answered question types with high average scores (i.e. "easy" question type), namely $\vec{v} \rightarrow \vec{a}$, $\vec{F} \rightarrow \vec{a}$, or $\vec{a} \rightarrow \vec{F}$.

For example, when comparing student responses to both $\vec{F} \rightarrow \vec{v}$ and $\vec{v} \rightarrow \vec{a}$ questions for the standard calculus-based physics course, we found that the scores are 20% and 56% for $\vec{F} \rightarrow \vec{v}$ and $\vec{v} \rightarrow \vec{a}$ questions respectively. As shown in Table 3.19, when comparing within-student responses, over 90% of students answering $\vec{F} \rightarrow \vec{v}$ questions correctly answered $\vec{v} \rightarrow \vec{a}$ questions correctly, $\frac{32}{32+3} \approx 91\%$, and over 90% of students answering $\vec{v} \rightarrow \vec{a}$ questions incorrectly answered $\vec{F} \rightarrow \vec{v}$ questions incorrectly, $\frac{28}{28+3} \approx 90\%$. For example, only about 40% of students answering $\vec{v} \rightarrow \vec{a}$ questions correctly answered $\vec{F} \rightarrow \vec{v}$ questions correctly, $\frac{32}{32+44} \approx 42\%$. These results are consistent with (but do not prove) the statement "correctly answering $\vec{F} \rightarrow \vec{v}$ questions necessarily implies correctly answering $\vec{v} \rightarrow \vec{a}$ questions", and the logical equivalent "incorrectly answering $\vec{v} \rightarrow \vec{a}$ questions implies incorrectly answering $\vec{F} \rightarrow \vec{v}$ questions". Furthermore, this contingency table disproves the converse statement: "correctly answering $\vec{v} \rightarrow \vec{a}$ questions necessarily implies correctly answering $\vec{F} \rightarrow \vec{v}$ questions."

Each table also includes the ϕ coefficient, which is a measure of correlation between the scores of each question type in the table. For example, in Table 3.19 discussed above, $\phi = 0.314$, denoting a medium-level correlation between the scores on $\vec{F} \rightarrow \vec{v}$ and $\vec{v} \rightarrow \vec{a}$ questions.

Within-student cross tabulations of scores between various question types. Cells represent numbers of students. Data reported are from the standard calculus mechanics course. A cell count is represented in bold face for tables which roughly satisfy conditions (discussed in Table 3.12) which are consistent with significant hierarchies between the indicated question types. Some question types had two questions posed; in this case the label “Correct” indicates that at least one question of that type was answered correctly, and “Incorrect” indicates that zero questions of that type were answered correctly. Note: Phi is the mean squared contingency coefficient between the question types, equivalent to the correlation coefficient for a 2×2 table.

$\phi = 0.324$	$\vec{F} \rightarrow \vec{v}$	
$\vec{v} \rightarrow \vec{F}$	Incorrect	Correct
Incorrect	58	17
Correct	13	18

Table 3.13: comparison one

$\phi = -0.026$	$\vec{a} \rightarrow \vec{v}$	
$\vec{v} \rightarrow \vec{F}$	Incorrect	Correct
Incorrect	39	38
Correct	17	13

Table 3.14: comparison two

$\phi = 0.140$	$\vec{v} \rightarrow \vec{a}$	
$\vec{v} \rightarrow \vec{F}$	Incorrect	Correct
Incorrect	25	50
Correct	**6**	25

Table 3.15: comparison three

$\phi = 0.138$	$\vec{a} \rightarrow \vec{F}$	
$\vec{v} \rightarrow \vec{F}$	Incorrect	Correct
Incorrect	16	59
Correct	**3**	28

Table 3.16: comparison four

$\phi = 0.131$	$\vec{F} \rightarrow \vec{a}$	
$\vec{v} \rightarrow \vec{F}$	Incorrect	Correct
Incorrect	27	48
Correct	**7**	24

Table 3.17: comparison five

$\phi = 0.451$	$\vec{a} \rightarrow \vec{v}$	
$\vec{F} \rightarrow \vec{v}$	Incorrect	Correct
Incorrect	49	23
Correct	**7**	28

Table 3.18: comparison six

$\phi = 0.314$	$\vec{v} \rightarrow \vec{a}$	
$\vec{F} \rightarrow \vec{v}$	Incorrect	Correct
Incorrect	28	44
Correct	**3**	32

Table 3.19: comparison seven

$\phi = 0.130$	$\vec{a} \rightarrow \vec{F}$	
$\vec{F} \rightarrow \vec{v}$	Incorrect	Correct
Incorrect	16	57
Correct	**4**	31

Table 3.20: comparison eight

$\phi = -0.066$	$\vec{F} \rightarrow \vec{a}$	
$\vec{F} \rightarrow \vec{v}$	Incorrect	Correct
Incorrect	22	50
Correct	13	22

Table 3.21: comparison nine

$\phi = 0.238$	$\vec{v} \rightarrow \vec{a}$	
$\vec{a} \rightarrow \vec{v}$	Incorrect	Correct
Incorrect	22	34
Correct	**9**	42

Table 3.22: comparison ten

$\phi = -0.022$	$\vec{a} \to \vec{F}$	
$\vec{a} \to \vec{v}$	Incorrect	Correct
Incorrect	10	46
Correct	10	41

Table 3.23: comparison eleven

$\phi = -0.013$	$\vec{F} \to \vec{a}$	
$\vec{a} \to \vec{v}$	Incorrect	Correct
Incorrect	18	38
Correct	**5**	34

Table 3.24: comparison twelve

$\phi = 0.064$	$\vec{a} \to \vec{F}$	
$\vec{v} \to \vec{a}$	Incorrect	Correct
Incorrect	7	24
Correct	13	63

Table 3.25: comparison thriteen

$\phi = 0.0384$	$\vec{F} \to \vec{a}$	
$\vec{v} \to \vec{a}$	Incorrect	Correct
Incorrect	11	20
Correct	24	52

Table 3.26: comparison fourteen

$\phi = 0.126$	$\vec{a} \to \vec{F}$	
$\vec{F} \to \vec{a}$	Incorrect	Correct
Incorrect	9	27
Correct	11	61

Table 3.27: comparison fifteen

A Summary of Hierarchy Trends suggested by Tables 1.11-1.25.		
Correct Responses on this Question Type.	**Implies** →	Correct Responses on this Question Type.
$\vec{v} \to \vec{F}$		$\vec{v} \to \vec{a}$
$\vec{v} \to \vec{F}$		$\vec{a} \to \vec{F}$
$\vec{v} \to \vec{F}$		$\vec{F} \to \vec{a}$
$\vec{F} \to \vec{v}$		$\vec{a} \to \vec{v}$
$\vec{F} \to \vec{v}$		$\vec{v} \to \vec{a}$
$\vec{F} \to \vec{v}$		$\vec{a} \to \vec{F}$
$\vec{a} \to \vec{v}$		$\vec{v} \to \vec{a}$
$\vec{a} \to \vec{v}$		$\vec{F} \to \vec{a}$

Table 3.28: summary

Finally, we use the method described in Table 3.12 on the data patterns in Tables 3.13-3.27 to present a summary of potential hierarchies in Table 3.28. Note that these hierarchies are only suggested by trends in the data tables. Nonetheless, these tables do provide strong evidence that statement converse to those in Table 3.28 are not true. For example, there is strong evidence, (via a significant number of counterexamples) that the statement "correctly answering $\vec{v} \to \vec{a}$ questions implies correctly answering $\vec{v} \to \vec{F}$" is not true.

3.6.1 Comments on Hierarchies in Responses

There are three points we would like to address concerning the determination of hierarchies of understanding. First, from the perspective of traditional design constraints on item statistics of a valid and reliable instrument (i.e., items have high internal consistency), it is not altogether unexpected that an answering hierarchy is aligned with increasing relative item score. Specifically, the constraint of choosing only items with a relatively high discrimination index implies that for a given student, if items with low averages are answered correctly then items with high averages also are answered correctly. Nonetheless, this does not diminish the significance of the finding that answering patterns of some pairs of questions types, such as $\vec{F} \to \vec{v}$ versus $\vec{v} \to \vec{a}$ are *not independent* and have a hierarchy (i.e., x implies y, but y does not imply x).

Second, it is worth pointing out that this analysis of hierarchies of question types can be viewed from the perspective of diagnostic assessment. Specifically, if students answer the difficult $\vec{v} \to \vec{F}$ and $\vec{F} \to \vec{v}$ questions correctly, then they are very likely to answer all other questions on the FVA test correctly. Therefore, to the extent that the FVA test measures understanding of the relations between force velocity and acceleration, one could view the $\vec{v} \to \vec{F}$ and $\vec{F} \to \vec{v}$ questions as the most *diagnostic* for determining understanding, at least for the level of students tested in this study.

Finally, while we have found evidence of hierarchies in student responses, our claims about hierarchies of student *understanding* of force, velocity, and acceleration are more qualified. It is important to keep in mind that "evidence of understanding" and any inferences of hierarchies that follow from such evidence depends on a careful characterization of how 'understanding" is operationally defined. For example, when judging whether a student adequately understands the relation between force and velocity, it could be considered reasonable to require that a student correctly and explicitly distinguishes the differences between velocity and acceleration *as part of* their explanation of (or answers to question about) the relation between force and velocity. However, one might not expect the converse, namely one might not require a student to explicitly distinguish the relation between force and velocity in order to demonstrate an understanding of the relation between velocity and acceleration. In this case, the evidence for a hierarchy of understanding $\vec{v} \to \vec{a}$ before $\vec{F} \to \vec{v}$ is strongly determined by the nature of the definition of "understanding".

In this research, we have instead asked questions about a specific relation, say $\vec{v} \to \vec{F}$, *without* any explicit reference to other relations, such as $\vec{v} \to \vec{a}$. To the extent that specific items in the FVA test measure understanding of each relation individually, without reference to other relations, observed hierarchical answering patterns suggest hierarchies in student understanding of relations between force, velocity, and acceleration. This finding is not simply an inevitable result of the operational definition of understanding of the relations, but appears to suggest that students at least implicitly connect different relations.

3.6.2 Hierarchies and Evolution of Responses

If there are hierarchies in responses to questions about the relations between force, velocity, and acceleration, then, in the course of learning, the evolution of responses should follow paths consistent with these hierarchies. Generally speaking, if correctly answering x necessarily implies correctly answering y, then gains in scores on y should precede gains in scores on x. For example, in the last section we provided evidence that correctly answering $\vec{F} \to \vec{v}$ questions implied answering $\vec{v} \to \vec{a}$ questions, but not the converse. Therefore, for students initially performing poorly on both $\vec{v} \to \vec{a}$ and $\vec{F} \to \vec{v}$ questions, we would expect that within-student gains in answering $\vec{F} \to \vec{v}$ questions correctly would not occur without gains in answering $\vec{v} \to \vec{a}$ (though one might expect to see the converse).

The pre-post FVA test data described in the section "Pre & Post Responses" allows for such a comparison of within-student gains in correct answering of the various question types. We analyzed all 15 possible pairs of question types (Tables 3.29 - 3.43) using the simple hierarchy method shown in Table 3.12, and found that there were no cases in which there were *zero* counterexamples for the statement "a gain scores for relation x necessarily implies a gain scores for relation y". However, for a few pairs of relations we did find cases in which there were relatively few counterexamples to such a statement, and these cases were exactly the ones that one would expect from the evidence of hierarchies of understanding described in the previous section.

Specifically, as shown in Table 3.35, over 84% of students who improved their score on $\vec{F} \to \vec{v}$ questions also improved their score on $\vec{v} \to \vec{a}$ questions,$\frac{38}{38+7} \approx 84\%$, and over 80% of students who *did not* improve their score $\vec{v} \to \vec{a}$ questions also did not improve their score on $\vec{F} \to \vec{v}$ questions, $\frac{29}{29+7} \approx 80\%$. Note that these patterns are not found for the converse.

For example, less than 45% of students who improved their score on $\vec{v} \to \vec{a}$ questions also improved their score on $\vec{F} \to \vec{v}$ questions, $\frac{38}{38+47} \approx 45\%$. As mentioned, with a relatively small number of counterexamples notwithstanding, these findings are consistent with the evidence of hierarchy of understanding the relations $\vec{F} \to \vec{v}$ and $\vec{v} \to \vec{a}$.

In Table 3.44, we compiled the data even further to demonstrate then general pattern that a gain in either a $\vec{v} \to \vec{F}$ or a $\vec{F} \to \vec{v}$ question necessarily implies a gain in either a $\vec{v} \to \vec{a}$ or a $\vec{a} \to \vec{v}$ question. That is over 92% of students who improved their score on either $\vec{F} \to \vec{v}$ or $\vec{v} \to \vec{F}$ questions also improved their score on either $\vec{v} \to \vec{a}$ or $\vec{a} \to \vec{v}$ questions, and over 77% of students who *did not* improve their score $\vec{v} \to \vec{a}$ or $\vec{a} \to \vec{v}$ questions also did not improve their score on $\vec{F} \to \vec{v}$ or $\vec{v} \to \vec{F}$ questions. This would support the finding discussed earlier that correctly answering questions about the relations between acceleration and velocity tends to be required in order to correctly answering questions about the relations between force and velocity. We did not find significant hierarchies for gains involving in $\vec{F} \to \vec{a}$ or $\vec{a} \to \vec{F}$ questions, most likely because the score to these questions where already near ceiling, leaving little room for gain.

Within-student cross tabulations of *gains* in scores between various question types. Cells represent numbers of students. Data reported are from the honors for engineers course. A cell count is represented in bold face for tables which roughly satisfy conditions (discussed in Table 8) which are consistent with significant hierarchies between the indicated question types. Cases in which the score was 2 out of 2 both on pre and post tests on a specific question types were removed. This helps to remove the less interesting "ceiling cases" that would register as "no gain". Some question types had two questions posed; in this case the label "Gain" indicates an increase of at least one correct response for that question type, and "No Gain" indicates either no increase in correct responses or a loss in correct responses for that question type. Note: ϕ is the mean squared contingency coefficient between the question types, equivalent to the correlation coefficient for a 2×2 table.

$\phi = 0.352$	$\vec{F} \rightarrow \vec{v}$	
$\vec{v} \rightarrow \vec{F}$	No Gain	Gain
No Gain	66	20
Gain	43	67

Table 3.29: gains comparison one

$\phi = 0.228$	$\vec{a} \rightarrow \vec{v}$	
$\vec{v} \rightarrow \vec{F}$	No Gain	Gain
No Gain	54	27
Gain	44	58

Table 3.30: gains comparison two

$\phi = 0.205$	$\vec{v} \rightarrow \vec{a}$	
$\vec{v} \rightarrow \vec{F}$	No Gain	Gain
No Gain	25	40
Gain	**11**	46

Table 3.31: gains comparison three

Phi = 0.341	$\vec{F} \rightarrow \vec{a}$	
$\vec{v} \rightarrow \vec{F}$	No Gain	Gain
No Gain	28	21
Gain	14	45

Table 3.32: gains comparison four

$\phi = 0.117$	$\vec{a} \rightarrow \vec{F}$	
$\vec{v} \rightarrow \vec{F}$	No Gain	Gain
No Gain	10	11
Gain	12	16

Table 3.33: gains comparison five

$\phi = 0.389$	$\vec{a} \rightarrow \vec{v}$	
$\vec{F} \rightarrow \vec{v}$	No Gain	Gain
No Gain	76	32
Gain	**20**	53

Table 3.34: gains comparison six

$\phi = 0.232$	$\vec{v} \rightarrow \vec{a}$	
$\vec{F} \rightarrow \vec{v}$	No Gain	Gain
No Gain	29	47
Gain	**7**	38

Table 3.35: gains comparison seven

Phi = 0.386	$\vec{F} \rightarrow \vec{a}$	
$\vec{F} \rightarrow \vec{v}$	No Gain	Gain
No Gain	33	26
Gain	**9**	41

Table 3.36: gains comparison eight

In sum, the contingency tables of the pre-post gains in student scores are in agreement with the hierarchies deduced from the contingency tables of within-student answering at a single time. This provides yet more evidence that specific hierarchies in student understanding of the relations between force, velocity, and acceleration and these hierarchies affect the

Continuation of within-student cross tabulations of *gains* in scores between various question types. Cells represent numbers of students. Data reported are from the honors for engineers course. For more information see the previous page.

$\phi = 0.133$	$\vec{a} \to \vec{F}$	
$\vec{F} \to \vec{v}$	No Gain	Gain
No Gain	13	17
Gain	10	12

Table 3.37: gains comparison nine

Phi = 0.091	$\vec{v} \to \vec{a}$	
$\vec{a} \to \vec{v}$	No Gain	Gain
No Gain	22	46
Gain	12	38

Table 3.38: gains comparison ten

$\phi = 0.221$	$\vec{F} \to \vec{a}$	
$\vec{a} \to \vec{v}$	No Gain	Gain
No Gain	26	24
Gain	15	32

Table 3.39: gains comparison eleven

$\phi = 0.108$	$\vec{a} \to \vec{F}$	
$\vec{a} \to \vec{v}$	No Gain	Gain
No Gain	13	12
Gain	8	11

Table 3.40: gains comparison twelve

$\phi = 0.239$	$\vec{F} \to \vec{a}$	
$\vec{v} \to \vec{a}$	No Gain	Gain
No Gain	14	11
Gain	20	28

Table 3.41: gains comparison thirteen

$\phi = 0.252$	$\vec{a} \to \vec{F}$	
$\vec{v} \to \vec{a}$	No Gain	Gain
No Gain	4	4
Gain	11	15

Table 3.42: gains comparison fourteen

$\phi = 0.148$	$\vec{a} \to \vec{F}$	
$\vec{F} \to \vec{a}$	No Gain	Gain
No Gain	7	8
Gain	7	12

Table 3.43: gains comparison fifteen

Phi = 0.313	$\vec{v} \to \vec{a}$ or $\vec{a} \to \vec{v}$	
$\vec{v} \to \vec{F}$ or $\vec{F} \to \vec{v}$	No Gain	Gain
No Gain	17	35
Gain	**5**	58

Table 3.44: summary

evolution of student understanding.

3.7 Comment on Learning Progressions

As mentioned in the introduction, Alonso and Steedle have hypothesized successive levels of understanding of force and motion Alonzo and Steedle (2009). There are several major differences between our study and theirs. For example, they studied middle school students, they studied understanding of magnitude (including change in magnitude) of quantities of force and motion, and only to a lesser extent did they also study understanding of relative direction of force and motion. Furthermore, they did not systematically study student understanding of the concept of acceleration (including direction) and its relations to velocity. Instead they focus on “motion” and occasionally make explicit references to “acceleration”. Nonetheless, an examination of their hypothesized levels of understanding reveal that in their model, students tend to understand issues concerning the relations between the direction of force and motion before they come to understand issues about the relations between the magnitude (and changes in magnitude) of force and motion. This result is certainly worth confirming in further focused empirical studies.

In contrast, our study is focused solely on the understanding of the relations between the directions of force velocity and acceleration, and as such our results do not contradict or confirm their results. Instead, our results may add more detail and depth to a portion of a larger learning progression framework for force and motion that may also include Alonso and Steedle’s work. While the term “learning progression” has not been uniquely defined (for example, see the discussion in Alonzo and Steedle’s work Alonzo and Steedle (2009)), the general idea is one of successive stages of student understanding of a concept or topic, starting from incomplete or incorrect knowledge and ending with some defined level of mastery, usually described by a particular science education standard. We have not set out a priori to construct a learning progression, rather we found that a consistent progression emerged out of our longitudinal and cross sectional data. This is somewhat in contrast to typical work on learning progressions (including that of Alonzo and Steedle), which rather than being primarily empirically driven, were typically constructed by an expert as some logical progression (from an expert point of view) toward mastery, with only some input on empirical data on how students are thinking or how they might progress toward mastery. As Alonzo and Steedle’s article states, “the learning progression represents a hypothesis about student thinking, rather than a description” (Alonzo and Steedle (2009)).

Indeed our approach is more empirical. We carefully designed questions to probe student understanding of logically and scientifically relevant dimensions (from an expert’s perspective), namely the six conditional relations. Nonetheless, understanding these relations could also be seen as sub-goals of understanding force and motion in general, and while an expert might logically order how these sub-goals would best be learned, this does not preclude the order in which students actually learn them which is an empirical question investigated here.

In sum, some of the results reported here could be used to link with recent efforts to identify learning progressions of force and motion. Specifically, our results could be used to construct a more formalized, empirically based learning progression of student understanding of the directions of force, velocity and acceleration, and this could be useful for instruction. Others implications for instruction are discussed in the next section.

3.8 Implications for Instruction

We will focus on some of the most important implications of the three major findings summarized in a previous section. First, instruction may be more effective if it accounts for the existence of intermediate states of understanding, especially since these intermediate states vary, depending on the specific conditional relation. For example, if an instructor focuses on the point "an object moving at constant velocity must have zero net force acting on it," this may help some students move from the misconception level into the somewhat common "Cannot-be-Opposite" intermediate state for that conditional relation, but it is unlikely to help the significant population of students who were already in that intermediate state to advance to fully correct understanding. Instead, instructors should be aware of the importance of focusing on the point that "an object which is moving may have a net force opposite to its direction of motion." Furthermore, if instructors are not careful in their assessment, they may incorrectly infer that students in the intermediate, partially correct, state have a complete understanding.

Another implication for instruction stems from asymmetry in responses to conditional relations between force, velocity, and acceleration. This implies that students may consider conditional examples differently during instruction. That is, a student who sees an example demonstrating that an object with a given instantaneous velocity can have any value of net force acting on it may perceive this differently than an example in which an object with a given net force can have any value of velocity. Furthermore, these two different examples may address different intermediate levels of understanding, as mentioned earlier. Therefore, attention must be given to both kinds of examples in order to fully address student difficulties with understanding these relations.

Finally, evidence for potential hierarchies in understanding the relations between the directions of force, velocity, and acceleration naturally has important implications for the order of instructional units and priorities for their mastery. For example, the results of this study imply that instructors must *first* ensure that students understand the relation between the direction of velocity and acceleration as well as force and acceleration in order to ensure that the students understand the relation between velocity and force, which is the source of common, compelling misconceptions. While from an expert point of view this order seems quite reasonable and perhaps expected, it is important to keep in mind that this study implies that teaching in the reverse order will not be as effective. Namely, teaching students *first* about the common misconceptions involving the relations between velocity and force may not be effective in preparing them to learn about the relations between velocity and acceleration or force and acceleration. While this and other implications of the order of instruction following from evidence of hierarchies of understanding is an interesting result, clearly more carefully controlled intervention studies are needed in order to better establish their validity.

In sum, we have found that the carefully designed FVA test has provided more comprehensive insight into student understanding of the relations between the directions of force, velocity and acceleration in one dimension. Clearly the levels of understanding these concepts has a rich and interesting structure, and the results of this study can help to inform careful decisions about the order and priorities of instruction as well as the identification and use of critical types of example questions to improve student understanding of this fundamental topic.

3.9 Summary

We have developed a 17 item multiple-choice test, the “FVA test”, designed to probe students understanding of the relationships between the directions of net force, velocity, and acceleration of (or on) a particle in one dimension. While previous research has examined one or two of these relationships at a time, the goal here was to holistically examine answering patterns for all six possible pairs of conditional relationships in order to obtain a more coherent picture of student understanding of these relations among the concepts of force, velocity, and acceleration. The development of the instrument included multiple stages of revisions with feedback via interviews and testing with standard and honors calculus-based introductory university students as well as second year physics majors. The test has been shown to have significant statistical reliability as well as validity for the population tested, as shown for example by significant correlations of FVA test score with course grade, level of the student, and the Force Concept Inventory score. The overall test scores indicate that traditional calculus-based physics students performed poorly on the test, with an average score of about 40%, and even second year physics majors find these questions somewhat challenging, with an average score of 70%. Furthermore, detailed patterns in student responses to the FVA test were analyzed, and several interesting findings were reported, as summarized below.

First, we consistently found evidence of an intermediate, partially correct level of understanding of the relations between force, velocity, and acceleration held by up to 30% of the students pre or post instruction. This is in addition to finding that a significant number of students answer consistent with the well-known and common student “misconception” that the vector quantities should always point in the same direction. Specifically we found two intermediate models. The first model is the belief that two vectors, such as force and velocity, need not be aligned, but they may also be pointed in opposite directions, but one cannot be zero (“Cannot-be-Zero” model). The second model is the belief that the two vectors need not be aligned, though one of them could be zero but not pointed in the opposite direction as the other (“Cannot-be-Opposite” model). Furthermore, we found that about half of the students who improved their understanding of the relations between the directions of force, velocity, and acceleration did so by evolving through these partially correct "states". Roughly speaking, from pre to post test in the honors physics sections, we found that about half of the students did not change their responses, about one-quarter changed from the misconception answered to the correct answer, and about one-quarter either changed from misconception answer to the partially correct answer or from the partially correct answer to the correct answer.

Second, we found an asymmetry in student responses to conditional relations. That is, students often treated questions that probe the concept *motion implies acceleration* differently than the concept *acceleration implies motion.* Likewise they often treated questions about *motion implies force* differently than *force implies motion*, and perhaps surprisingly, they often treated questions about *Force implies acceleration* differently than *acceleration implies force.* The differences are reflected in the response frequencies of each answer choice. For example for $\vec{v} \rightarrow \vec{a}$ versus $\vec{a} \rightarrow \vec{v}$ questions, there were differences in the number of correct and misconception responses as well as in the kinds of partially correct responses (Cannot-be-Opposite versus Cannot-be-Zero). For the $\vec{v} \rightarrow \vec{F}$ versus $\vec{F} \rightarrow \vec{v}$ questions, there were no differences in the correct and misconception response frequencies, but there were differences in the partially correct Cannot-be-Opposite versus Cannot-be-Zero responses.

Third, we found evidence of specific hierarchies in correct responses to different question types. The evidence included both within-student scores at one point in time and within-student gains in scores from pre to post. For example, we found evidence that if students correctly answered $\vec{F} \rightarrow \vec{v}$ questions, then they were very likely to also correctly answer $\vec{v} \rightarrow \vec{a}$ questions, but not the converse. Further, if $\vec{v} \rightarrow \vec{a}$ questions were answered incorrectly, then it was very likely that $\vec{F} \rightarrow \vec{v}$ questions were also answered incorrectly (but not the converse). Likewise we found that for a given student, gains in $\vec{F} \rightarrow \vec{v}$ scores most likely occurred in the presence of gains in $\vec{v} \rightarrow \vec{a}$ scores, but the not converse. These findings are indeed interesting and suggest that it may be required to understand the relationship between the direction of velocity and acceleration in order to understand the relationship between the direction of force and velocity. However, to more firmly establish a possible causal link between understanding these relations, one must first be careful to explicitly define what is meant by *understand*, and, second, a controlled intervention (for example manipulating the amount of velocity-acceleration instruction) is needed.

Volume II

Identifying and Addressing Student Difficulties with Materials Science Concepts

Chapter 4

A Tutorial Design Process Applied to an Introductory Materials Engineering Course

4.1 Introduction

This part of the thesis presents my work in identifying and addressing student difficulties with materials science concepts. This project was to develop effective concept oriented group-work worksheets, often called tutorials, for use in the introductory materials science course. Because the developers on the project were PER researchers, and not materials scientists, we took the well-known and effective tutorial design process that has worked to create several effective physics tutorials, and we mapped this process onto the materials science course. The different steps and how we implemented them are discussed in detail in the following chapters, but the main process is to find the most important student difficulties with the course material, aim the tutorial activities at these difficulties in an 'elicit-confront-resolve' process of conceptual change, and test to make sure that the intervention was effective. In the following chapters we present this process, the positive results of tutorial use on student conceptual understanding, the data collected on student difficulties, the tutorials created, and a test designed to evaluate tutorial effectiveness.

This chapter, Chapter 4, provides an over view of the project as a whole: it introduces tutorials as used in physics education; the tutorial design process for physics; the motivation for believing this process could be mapped effectively to materials science; how the process was applied to the introductory materials science course; how the tutorials were implemented and the course concerns for their implementation, such as TA training and course exams reflecting recitation material; data collection methods; and, finally, results of the tutorials on student conceptual understanding.

Chapter 5 reports on the large, and still incomplete, set of student difficulties with the various materials science concepts. In addition, it presents a brief discussion of the tutorial material and reasons for using that specific material.

Chapter 6 discusses over all course goals and the pedagogies used across several of the tutorials, such as multiple representations and elicit-confront-resolve.

Chapter 7 concludes this part of the thesis with a presentation of the materials science conceptual evaluation (MSCE) that was developed to test student understanding. It reviews the item design process, the steps taken to validate the items, and several sets of data from students taking versions of the MSCE.

4.2 A Physics Tutorial Design Process for a Materials Science Course

In the area of physics education, an iterative process of design and implementation of instructional units called "tutorials" has been shown to be very effective in improving student understanding of a wide variety of physics topics. The process of curricular and instructional development which we will refer to as the tutorial design process was first developed and implemented over 20 years ago by the University of Washington's Physics Education Research group, led by Lillian McDermott (McDermott, Shaffer, and the PER Group at U. of Wash. (2002); McDermott (1991); Shaffer and McDermott (1992)). It has been applied to numerous topics in Physics including electric circuits (Shaffer and McDermott (1992)), force and motion (e.g., McDermott, Shaffer, and Somers (2004); Shaffer and McDermott (2005); Heron, Loverude, Shaffer, and McDermott (2003)), intermediate mechanics, (Ambrose (2004)), geometric optics, (Wosilait, Heron, Shaffer, and McDermott (1998)), interference and diffraction of light, (Wosilait, Heron, Shaffer, and McDermott (1998)), pressure, (Loverude, Heron, and Kautz (2010)), sound (Wittmann, Redish, and Steinberg (2003)), quantum mechanics (Singh (2008); Wittmann, Morgan, and Bao (2005); Bao and Redish (2002)), and special relativity (Scherr, Shaffer, and Vokos (2002)). These studies have also demonstrated that the implementation in the classroom has resulted in significant, consistent gains in conceptual understanding, reasoning and simple problem solving. Based on much of this research, several groups have published tutorial workbooks for introductory physics (McDermott, Shaffer, and the PER Group at U. of Wash. (2002); Wittmann, Steinberg, and Redish, (2005)). A large and comprehensive study has demonstrated that application of the tutorial method in introductory mechanics (using the workbook of McDermott et al., 2002) results in consistent gains for instructors with a variety of backgrounds and experience (Pollock and Finkelstein (2008)). Finkelstein and Pollock (2005) have also identified and investigated key conditions of implementation.

4.3 Development of Tutorial Materials and Methods

The development of tutorial materials and methods is an iterative and often parallel rather than serial process consisting of four tasks, outlined in Table 4.1. The outline of the development of tutorial materials and methods is simple and straightforward, however, several of the tasks often require significant time and effort, and there is significant overlap among tasks. The outline describes key features of the tutorial design process and is based on the work mentioned above and summarized in articles such as McDermott (1991) and McDermott (2001).

The first task is to identify a target content topic and instructional goals. Prevalent and widely accepted goals for standard introductory courses can be identified, for example, via interviews of course instructors and analysis of the syllabus and textbooks. It should be noted that tutorials usually focus on conceptual understanding and scientific reasoning rather than, for example quantitative problem solving or memorization.

The second task is to identify and characterize specific student difficulties with the target topic. This task is considered a fundamental component of the tutorial design process, and typically involves student interviews and testing over a period of a year or more. This task is critical for providing specific and useful information for the iterative development of materials, methods and assessments. Furthermore, the careful identification of specific student difficulties often results in a modification of instructional goals.

The third task is to construct specific material and methods aimed at addressing student difficulties. The typical strategy is to present questions that will reliably elicit specific student difficulties, engage the students in confronting the difficulty, and providing activities and exercises that help them to resolve the difficulty. It is common to use multiple representations including graphs, diagrams, equations, and an emphasis on clear, logical explanations. The exercises often stem from the interview questions posed to students in the previous task to discover student difficulties.

The fourth task is field testing and assessment of effectiveness. Field testing typically requires several iterations of material and methods revisions over several semesters. Assessment of the effectiveness typically emerges as part of the process of instructor/teaching assistant (TA) feedback and construction of key questions and problems that diagnose student understanding. A distinct advantage of this process is that the materials are empirically tested in classrooms so they are known to function in a real classroom environment.

4.4 Implementation of Tutorials

The development and the implementation of tutorials are closely related, since the development includes several iterations of implementation. However, there are a number of important issues and factors in implementation that are separate from development of materials

	Task	Resources
1	Identification of target content topic and instructional goals	Instructor interviews, syllabus, and textbook.
2	Identification and characterization of student difficulties with target topic.	Student interviews, instructor interviews, test performance, existing Education Research.
3	Construction of interactive small-group tasks aimed at the instructional goals and addressing specific student difficulties.	Findings from Task 2, findings from education research, existing materials from instructors, textbooks.
4	Field-testing and assessment of effectiveness.	Test and homework performance, in-class interaction, feedback from TA's and course instructors.

Table 4.1: Outline of the iterative development of tutorial materials and methods.

and methods. Finkelstein and Pollock (2005) discuss critical features of implementation at several levels, including at the level of the task, the classroom environment, the course structure and the support at the departmental and university level. Here we will briefly describe some of the key conditions of implementation, as outlined in Table 4.2. The extent to which each condition alone is important has not been rigorously investigated. However, the conditions described here are common to most tutorial implementations that have demonstrated success in student learning. The first condition is the use of well-developed tutorial materials, using the tutorial design process described above.

The second and third conditions describe the format of the tutorial class activity. Students are placed in groups of 3 or 4, and they actively participate in the tutorial activity. The TA's facilitate student participation and elicit more complete and correct student explanations by asking the students well-posed questions that respond to the immediate student difficulties. The role of the TA is not to provide explanations, but to have students provide explanations.

The fourth condition is TA preparation. There are two components to this. First, the TA's must learn the skills of facilitating student explanations and handling the group-work environment. Second, the TAs must be fully prepared for the specific issues of each week's tutorial. Particularly, the TAs must be aware of common, specific student difficulties that will arise that week and be prepared with key questions that can help students to progress.

Finally the fifth condition is to support the importance of the tutorial material by placing questions on exams that are explicitly based on tutorial materials.

1	Well-developed tutorial tasks which challenge students, elicit difficulties, and provide pathways for resolution.
2	Class is structured in group-work, active participation format.
3	TA's continually emphasize and facilitate complete explanations from students. TA's do not simply give explanations and answers.
4	TA's are prepared weekly for specific student difficulties that go with each tutorial and general skills in group facilitation.
5	Tutorial material is explicitly tested on course exams.

Table 4.2: Critical conditions for tutorial implementation.

4.5 A Necessary Digression Describing Participants and Data Collection Methods

Data was collected in this study in order to (a) identify and characterize specific student difficulties and (b) implement tutorials and assess student learning from the tutorials. In this section we describe details of the participants and how the data was collected.

The participants in this study were enrolled in the introductory materials science course for engineers, which is a required core course for many of the engineering major programs at Ohio State University, a large public research university. The students ranged from 1st to 5th-year engineering students. About 10-15% of the students intended on becoming materials science engineering majors, and about 35% of the students were mechanical engineering majors, the most common major in the course. Data was collected over a period of 7 quarters, for a total of approximately 1000 participants. The data was collected in five ways. First, we conducted individual or group interviews on over 200 students. These interviews consisted of asking a wide range of open ended and multiple-choice questions, such as those presented in this paper. Several dozen interviews were videotaped, and the rest were recorded via interview notes. The interviews were used to first explore areas of difficulty, then to focus on specific difficulties identified in the initial interviews and free response tests. Most interviews were conducted individually, but some were given in groups of 3 or 4. The second method of data collection was via free response and multiple-choice tests. In addition to the standard homework, students were given a "flexible homework" assignment with credit for participation as part of the course grade. The flexible homework assignment consisted of participation in a one-hour session where students completed some combination of testing and interviewing. Throughout the quarter, students were randomly selected to participate in the flexible homework. Typically, about 95% of students participated in the flexible homework. The tests items were in either multiple-choice, free-response, or a multiple-choice-with-explanation format. Students completed the material at their own

Quarter	Number of Students	Description of Experiments done
Sp 2008	45	Exploratory interviews and testing data collected via student volunteers.
Au 2008	45	
Wi 2009	150	Pilot Testing of Tutorial activities with Mechanical Properties during special 48 minute homework sessions held in our lab, i.e. “flex homework”.
Au 2009	175	Tutorials used in weekly recitations. Questions pertaining to recitation were given on exams. Short-answer and multiple-choice data on student difficulties collected in 48 minute homework sessions held in our lab, i.e. “flex homework”.
Sp 2010	175	Short-answer and multiple-choice data on student difficulties collected in 48 minute homework sessions held in our lab, i.e. “flex homework”.
Au 2010	120	Reworked tutorials used in weekly recitations. Questions pertaining to recitation were given on exams. Multiple-choice data on student difficulties collected in 48 minute homework sessions held in our lab, i.e. “flex homework”.
Wi 2011	220	
Sp 2011	150	No tutorials used in recitations. Questions pertaining to usual recitation material given on final exam. Multiple-choice data on student difficulties collected in 48 minute homework sessions held in our lab, i.e. “flex homework”.

Table 4.3: Summary of different data collection methods used and numbers of participants.

pace at individual stations in a quiet room. Afterwards we would informally ask students whether they had any questions and/or to explain their answers. We observed during these sessions that students made a good faith effort to answer the questions to the best of their ability.

The third method for collecting data was again via flexible homework. However, the sessions were conducted in a more regular classroom environment during the 7, 8 or 9th week of the quarter only. Students were given 30 to 35 minutes to complete a set of about 30 multiple-choice items. (Students were aware that their score on the “quiz” would not affect their course grade, but again we observed the majority of students making a good faith effort to answer the questions.) Students were then asked to spend approximately 10 minutes discussing the correct answers to the items in small groups with TA assistance which allowed for some informal interview data on the validity of the multiple-choice items.

The fourth method for collecting data was via observations of small group work and collected tutorial responses in recitations. The authors participated in some of the recitations which were conducted in small group format. This method was used to further verify and/or clarify student difficulties found via interviews and tests, to evaluate student understanding and interaction with the tutorial questions, and to assess student ability to correctly complete the tutorial questions. Finally, the fifth method for collecting data was via the official exams administered as part of the course. The exams were in multiple-choice format, and some of the items (about 10-20%) were designed by us in collaboration with the instructor. This method helped to assure that student answering was not simply an artifact of the testing context, i.e., whether performance would dramatically improve for high-stakes testing contexts.

Most tests and interviews were at least one week after the relevant instruction, however some were administered before relevant instruction. The data reported here is all post instruction. Most of the difficulties reported here were first found in interviews. We subsequently devised questions to demonstrate the relative frequency of these difficulties in the student population. Thus incorrect answers to the questions should not be viewed as uninteresting artifacts of the particular questions, but rather indicative of student difficulties with understanding the materials science concepts underlying the questions, or possibly, as in the case with questions in graph or diagram format, some of the difficulty arises from the format itself.

4.6 Initial Exploration of Course Goals

The course is a standard course in introductory materials science and engineering at the college level. The course is required by many engineering majors at Ohio State University. The topics covered are well-represented by the textbook of the course, which is the common textbook by Callister (2007). At the beginning of the project, we interviewed 5 faculty who had recently taught the course and asked them what they saw as the goals of instruction of the course. Perhaps as to be expected, there were common themes to the responses. First, specific topics were mentioned, such material properties, diffusion, defects, fatigue etc. - topics that were usually chapters in the text. Second, the instructors also described more general goals, and we found five general goals commonly stated. The goals are listed in Table 4.4.

The overwhelming message from the instructors was that basic conceptual understanding of basic materials science ideas was the most important goal. This goal lends itself well to the tutorial design process, which is typically aimed at conceptual understanding and basic reasoning. Therefore the content topics chosen for the tutorials were closely aligned with lecture topics and chapters in the textbook and we focused on conceptual understanding, keeping in mind the general goals in Table 4.4.

4.7 Recitation Format

The course consisted of three lecture and one recitation class per week. Recitations were 48 minutes in duration, and recitation attendance was voluntary. The number of the students per class varied but was on average 20 to 25 students. A senior experienced instructor, usually a faculty member, and two teaching assistants, typically graduate or undergraduate students, were present for each recitation. Students worked in groups of 3 or 4 to complete the tutorials and instructors circulated to answer questions, ask questions of the groups, and in general facilitate the activities. About 37% percent of students typically attended recitations, or every student attends on average 3.4 recitations out of nine total. This attendance rate is similar to the attendance rate of the traditional recitations used before the tutorials, which typically consisted of mini-lectures and discussion of solutions to homework problems. Therefore the tutorials themselves did not appear to significantly affect attendance. Data from the most recent quarter of tutorial implementation had an attendance of 62% or about every student going to 5 recitations out of the eight available that quarter. This suggests that the tutorials might have a small, positive effect on attendance. However, other factors such as instructor encouragement, student population, etc. might have contributed to and/or be the cause for this increase.

4.8 The Importance of TAs to Facilitate the Tutorials

Instructor preparation for the tutorials is a critical component of the implementation. While the tutorials consist of carefully constructed questions, they are designed to be complemented by questions and dialog between the students and the teaching assistants (TAs). Every week the TAs have an hour long training session in which they discuss the correct answers to the tutorials, difficulties they can expect students to have, how to assess what difficulties the students are having, and how to guide students to overcome their difficulties

Table 4.4: General goals of the introductory materials science course identified by instructors.

1. Understanding of basic definitions, e.g., yield strength.
2. Basic understanding of major concepts, e.g., fatigue, diffusion.
3. Basic understanding of the relation between structure, processing, and properties.
4. Conceptual intuition for basic properties of common materials and material classes.
5. Basic conceptual understanding of material selection, e.g. necessity of tradeoffs.

6.* Ability to interpret and use basic materials science graphs and diagrams.

Note. The sixth goal was identified in the course of this study, as a result of discovering pervasive student difficulties with graphs and diagrams.

by asking thoughtful and responsive questions instead of simply telling the student the correct answer. As a result of TA training, the TAs are prepared for potential issue and have prepared ways to help students. Effective dialog between the students and TAs is critical for the tutorials to be successful and preparation of methods to overcome known common difficulties is an important part of effective dialog.

4.9 TA Training and Learning

We have discussed the need for effective teaching assistants to facilitate the tutorials and the training needed to prepare them. Here we will discuss the results of such training on TA knowledge of materials science, on their ideas about teaching, and on their feeling of preparedness to teach each week.

How much did you learn about MSE?	Lots	5	4	3	2	1	Nothing
	TA respones	25%	37%	25%	13%	0%	
How much did you learn about teaching?	Lots	5	4	3	2	1	Nothing
	TA respones	31%	44%	25%	0%	0%	
How do you feel about MSE content now?	I can ace the final	5	4	3	2	1	I know little and would not do well on a final
	TA respones	25%	69%	6%	0%	0%	
How did you feel in terms of being prepared each week to teach	I always felt well prepared	5	4	3	2	1	I usually felt unprepared
	TA respones	19%	63%	19%	0%	0%	

Table 4.5: TAs survey responses of preparedness and teaching knowledge (N = 16).

To test TA knowledge we gave various versions of the MSCE test the first day of TA training and the last day of TA training. Such results are reported in detail in Chapter 7 where the MSCE test is discussed. The main effect seen is twofold. First, the TAs pretested at approximately 15% better than the average student post tested (around 70% correct). Second they gained about 20% after a quarter of teaching so that their posttest scores were around 90%.

In addition, TAs were given a survey to fill out with a combination of likert scale 1 to

5 ratings and short-answer responses. We post tested 16 TAs this way. A selection of their responses for certain questions is shown in Table 4.5. For the most part the TAs responses were positive but without a lot of variation: mostly indicating a 4 out of 5 for things like learning the mse materials, learning of teaching, and feeling prepared to teach.

It is also interesting to consider TAs responses to the item “Please give a quick explanation of your idea of the role of a TA.” This was given to 10 TAs in a pretest. Of these 2 did not respond, 2 indicated holding homework help, grading, or recitation sessions and did not elaborate, 4 indicated that there role was answering questions. Only 2 of the 10 TAs, i.e. 20%, indicated a more sophisticated teaching ideology by mentioning asking questions of the students and/or guiding students to correct concepts. On a posttest, of 16 TAs 4 did not respond, 1 was off topic, 1 was ambiguous, 2 mentioned explaining things, and 8, i.e. 50%, showed a more sophisticated teaching ideology saying their role was: “a guide”, “...encourage students to dig deeper”, “ask questions...” These are small numbers, and this was one open ended survey question. However, the results are encouraging that the training session may have positively affected TAs ideas about teaching.

4.10 Iterative Construction of the Tutorial Materials

In the following chapters, chapters 5 and 6, the important student difficulties identified in this study and the design and pedagogy of the tutorials used to address these difficulties is discussed in detail. Here we describe the basic construction and evaluation methods and give an example of the developmental stages of the diffusion tutorial.

The writing of tutorial materials began after the initial identification of course goals and student difficulties with specific topics. In collaboration with the instructor, high-priority student difficulties were chosen for the tutorials among the many student difficulties discovered in the initial interviews and testing. An initial version of the tutorials was drafted based on interview and test questions used in the initial difficulty-identification stage. One quarter (i.e. course term) of small mock recitation sections (with students attending as part of a homework assignment) was used to field-test the first version of a handful of the tutorial activities. Feedback from the written material produced by the students and comments from the instructors implementing the sections was used to improve on the materials. After this, the tutorials were implemented in the regular recitation sections for three quarters. For each quarter, the tutorials were redesigned based on in-class observations, assessments of submitted group responses to the activities, instructor feedback, and assessments of the tutorial’s effectiveness based on exam scores.

4.11 Results of Implementation

Nine tutorial worksheets were implemented in weekly recitations (as outline earlier) in three separate lecture sections in three separate quarters. Two of the sections (Autumn 2010 and 2011) were taught by an experienced materials Science Engineering faculty member who taught the course for a number of years, and one of the sections was taught by a faculty member who taught the course for the first time. All lectures were mainly traditional format, with the exception that electronic voting machine questions were presented and discussed for a portion of most of the lectures.

Regarding assessment, as mentioned earlier, a Materials Concept Inventory is available (Krause, Kelly, Triplett, Eller, and Baker (2010)), but we found that this instrument was not adequate for the purposes of our study because, for example, the instrument has several items assessing topics that were not covered in the course.

Therefore, we evaluated the effectiveness of the tutorials by analyzing final exam scores. The final exams were constructed in multiple-choice format by the instructors and were similar in nature in all three sections, yet each exam was different. The final exams were on average 32 questions long with approximately 18 questions related to the tutorials given in recitations and 14 questions unrelated to recitation tutorials. The metrics of the exams were within a reasonable range, for example the Cronbach's alpha measures were in the range of 0.7-0.75 and average scores in the 70-75% range.

While the use of final exam scores does not allow for a comparison between lecture sections, for example between a section using tutorials vs. a section that employs traditional recitation lecture and homework reviews, one still may make informative comparisons of exam performance within lecture sections. Specifically, as noted earlier, students attended on average only about 37% of the recitations. This variability in attendance allowed for a comparison of final exam scores of students who attended the voluntary recitations to scores of students who did not attend recitation. Figure 4.1 reveals a clear linear, increasing relation between the number of recitations attended and the final exam score. Students gained about 0.10 ± 0.02 standard deviations on the final exam score for every recitation they attended. (See Figure 4.1 which shows similar trends in all three quarters.)

However, this increase performance with increased tutorial participation is necessary but not sufficient evidence for the added benefit of tutorials. Because the recitations were voluntary, it might be argued that the gain in score could be due to the fact that better students may have more often attended recitation. While this is a plausible contributing factor to the score-attendance relation measured, we argue that there is evidence of additional learning due to recitation participation. In particular, we made a within student comparison of performance on exam questions that were related to the recitation material vs. exam questions that were not related to recitation material. If one interprets the score

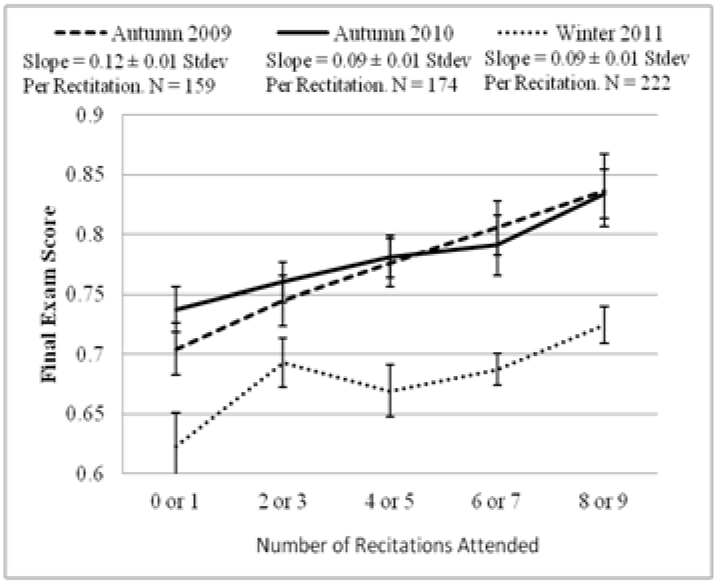


Figure 4.1: Final exam score vs. recitation attendance for three separate quarters. The slope for each quarter is also provided. The recitations are binned by two to reduce fluctuations from topic difficulty, tutorial strength, and the number of final exam questions on each topic. (Note, Winter 2011 had only 8 recitations)

on the non-recitation related questions as a measure of the mastery level of the student, then one can use this information to factor out the quality of the student and determine if there is any residual effect of recitation attendance alone.

In order to formally analyze this possibility, we performed a 2 x 2 repeated measures analysis of variance, with high/low attendance (greater or less than average attendance) as the between student factor, and question type (recitation related vs. not recitation related question) as the repeated (within student) factor. Figure 19 presents the data separated into these factors. Perhaps unsurprisingly, there was a main effect of attendance, with high attenders outscoring low attenders by 0.36 standard deviations on all questions on average (F(1, 553) = 20.06, $p < 0.001$). There was no main effect of question type, thus students performed on average equally well on recitation vs. non-recitation related questions, 74% vs. 73% (F(1, 553) = 3.76, p = 0.053). Most interestingly, as seen in Figure 4.2 there was a significant interaction between attendance and question type (F(1, 553) = 19.55, $p < 0.001$), with a clear value added on recitation related questions for those attending recitation. Thus, even accounting for the fact that slightly "better" students attended the recitations there was a valued-added effect of the recitations that improved relative student performance on recitation related questions by 0.35 standard deviations. This effect size is fairly impressive considering that difference in attendance between high and low attenders is about 4 recitations, which is less than 3.5 hours of tutorial participation.

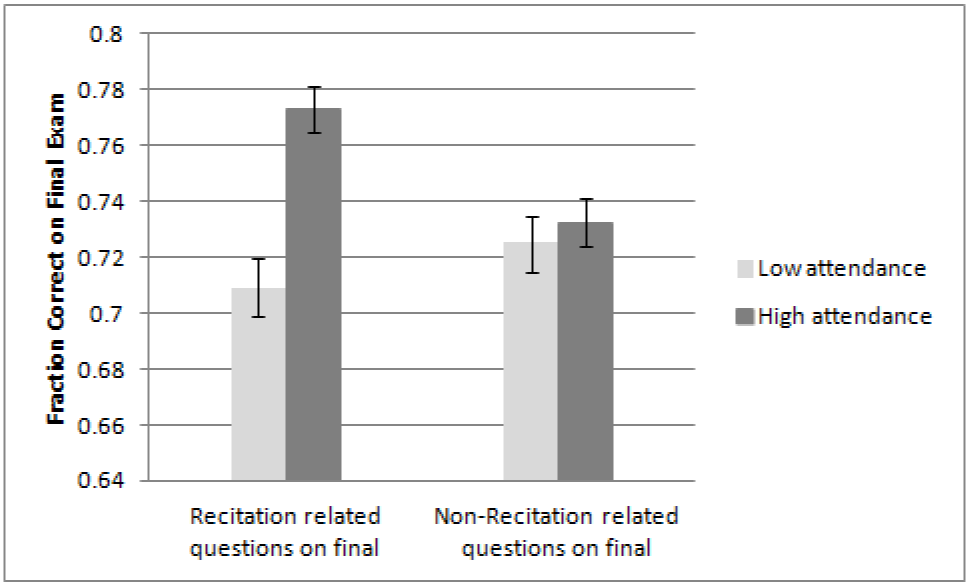


Figure 4.2: Interaction of recitation attendance and recitation related vs. non-recitation related questions (N=553). Low attendance is less than average attendance 3, recitations or less, and high attendance is greater than average attendance, 4 or more.

4.12 Summary

What is somewhat remarkable is that there exists a fairly well-defined process for developing physics curricula that significantly improves student learning of important and difficult concepts in physics, and this study demonstrates that this same process can also work for an introductory materials science engineering course. The implication is that this process may be applied to all STEM topics. While the process is fairly straightforward, it nonetheless requires significant and lengthy effort typically on the order of several years. This process includes the careful identification of specific student difficulties and the iterative process of developing interactive group activities facilitated by suitably-prepared TA's.

More specifically, we have reported here on the successful application of this tutorial design process to an introductory materials science engineering course. We have identified a number of prevalent student difficulties with fundamental concepts in materials science, and designed tutorials as group-work activities aimed at addressing these difficulties. The tutorials follow the a large extent many of the chapters of a common book used in the course (Callister (2007)), and include the topics of atomic bonding, crystal structure, diffusion, mechanical properties, plastic deformation, cold-working, creep, fatigue, failure, phase diagrams, TTT plots, ceramics, polymers and composites. The tutorials not only address specific topics but also more general instructional goals such as the interpretation of diagrams and graphs, and the relation between structure, processing, and properties. The tutorials guide student through a series of questions which are designed to elicit known student difficulties and encourage students to explicitly confront these difficulties via group discussions of posed questions and/or dialog with a TA. In addition, the tutorials address larger course goals - including understanding basic materials science terms, visualizing the microstructure of materials, and expertise with graphs and diagrams - through incorporating these skills in each of the tutorials. Analysis of final exam performance suggests that participation in the tutorial based recitations significantly improves student understanding. In all, these results suggest that these methods and materials are effective in teaching students the difficult and important conceptual materials which they were designed to address.

Chapter 5

Examples of Several Specific Student Difficulties Identified and the Tutorial Materials Designed to Address Them

5.1 Relevant previous work in identifying student difficulties in material science engineering

Several previous studies have identified and described a number of student difficulties with concepts in introductory materials science (Krause, Decker, and Griffin (2003); Krause, Tasooji, and Griffin (2004); Kitto (2008); Krause, Kelly, Tasooji, Corkins, Baker, and Purzer (2010)). In addition, a "Materials Concept Inventory" has been developed to assess a very basic conceptual understanding for some of the topics in materials science (Krause, Decker, and Griffin (2003); Krause, Kelly, Tasooji, Corkins, Baker, and Purzer (2010)). Krause and collaborators (Krause, Kelly, Triplett, Eller, and Baker (2010)) have also categorized some materials science "misconceptions" and investigated how different methods of instruction affect these categories of misconceptions. They found that interactive concept sketching activities were the most effective at raising scores on the Materials Concept Inventory, followed by interactive concept card sorting activities, interactive discussions and passive lecture (in order of decreasing effectiveness).

In this study, we have used some of the these previous findings on student difficulties as a starting point to characterize and investigate in more detail identified difficulties as well as investigate a wide range of additional student difficulties with basic materials science concepts. Therefore some of the information from previous studies about student difficulties has in a sense become a part of the tutorial design process.

5.2 Atomic Bonding

5.2.1 Student Difficulties with Atomic Bonding

Here we describe a number of deep and fundamental student difficulties in understanding atomic bonding and its relation to macroscopic properties of metals. These misunderstandings may significantly contribute to barriers in understanding materials at an atomic level, and how the atomic level physics effects macroscopic properties. We will discuss four major difficulties with basic concepts. 1. Students often believe that bonds themselves can be permanently weakened or stretched. 2. Students often confuse density, strength of atomic bonds, melting temperature, and yield strength. 3. Students often also confuse bond strength with force rather than energy. 4. Students have considerable conceptual and graphical difficulties with Lenard-Jones bonding potential energy graphs. All of these results were obtained after traditional instruction that explicitly covered these topics.

We found that students often believe that atomic bonds can be modified, such as becoming permanently stretched, much like the phenomenon of a permanently stretched spring. Evidence of the stretched-bond model is shown in Figure 5.1, and was verified in numerous interviews. Students were asked post instruction to compare the atomic separation in a metal before and after plastic deformation, and 71% of them indicated that the bonds would be stretched after plastic deformation. Note that this idea of stretched bonds is similar to the results found previously regarding a question in the Materials Concept Inventory, in which many students answered (pre-instruction) that when a wire is drawn through a tapered hole, the bonds have been compressed (Krause, Decker, and Griffin (2003)).

This incorrect model of atomic bonding and plastic deformation reveals that students do not understand the process of plastic deformation at a microscopic level, and this may in turn contribute to difficulties in understanding how yield strength is determined by the propagation of dislocations, rather than the permanent stretching of atomic bonds. In addition to the idea of stretched bonds, many students also believe that bonds can be weakened. Consistent with previous findings (Krause, Tasooji, and Griffin (2004)), we also found that about 40% of interviewed students believed that when a metal is heated, atomic

A metal is permanently elongated by a load, then the load is removed. Which of the following is true?

(29%)	**Atoms will be rearranged compared to before the elongation.**
(11%)	The atomic bonds will be stretched compared to before the elongation.
(60%)	Both a) and b) will occur.

Figure 5.1: Example question demonstrating student difficulty with the concept of plastic deformation. Student response percentages in parentheses, and correct answer in bold. (N = 64. SE = 6%.)

Material A has a greater (average) atomic separation than Material B. Which of the following must also be true given this information? (You may choose more than one.)

(72%) Material B has a greater mass density.
(75%) Material B has a great atomic bond strength.
(44%) Material B has a greater yield strength.
(40%) Material B has a greater melting temperature.

Figure 5.2: Example question demonstrating student difficulty with the concepts of bonding, atomic seperation, and strength. Student response percentages in parentheses. None of these options is correct. Question regarding atomic separation and material properties. (N = 67, SE = 6%.)

bonds become weakened and they believed this explains, for example, why a heated metal expands.

Second, we found that an overwhelming majority of students assumed that high mass density necessarily implied small atomic separation. In this case, students ignored the fact that mass density depends on both atomic separation and atomic mass. When pressed in interviews, most students quickly recognized that atomic mass is a factor. However the neglect of atomic mass when considering mass density was quite pervasive, as shown in Figure 5.2, response a.

The assumption that mass density necessarily determines atomic separation and vice versa may seem like a minor and innocuous oversight. Students may have simply interpreted (implicitly or explicitly) that "density" means "number density" rather than the more commonly assumed "mass density" (even if "mass" is explicitly stated), and the focus on number density might be expected, since the lessons on crystal structure focus on numbers of atoms, for example when calculating the atomic packing factor, rather that the mass of the atoms. However, Figure 5.3 provides evidence that this assumption is a symptom of a much deeper misunderstanding of microscopic and macroscopic properties which may lead to common errors in multi-step reasoning involving density. Responses to the question in Figure 5.2 and interview responses reveal that many students use a train of incorrect steps to argue that density predicts strength and melting temperature. First, most students assume that relatively high mass density implies relatively small average atomic separation. Second, most students also believe that relatively small average atomic separation necessarily implies relatively large atomic bond strength. Finally, most students believe (correctly) that high atomic bond strength necessarily implies high melting temperature and (incorrectly) that high atomic bond strength necessarily implies high yield strength. Therefore, the idea that high density implies high melting temperature and high yield strength is compelling to students because there is a natural and plausible (yet incorrect) mechanism: stronger atomic bonding due to smaller atomic separation. This is the common student reasoning

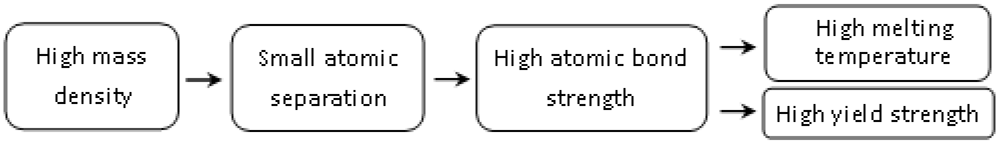


Figure 5.3: Common incorrect line of reasoning about the relation between density, melting temperature, and yield strength. Note that all but one of the steps is incorrect.

for equating the properties of mass density with the macroscopic properties of melting temperature and yield strength. (See Figure 5.3 for a summary of this reasoning pattern.)

The third basic student difficulty with understanding atomic bonding is that students often confuse the concepts of force and energy when referring to the strength of atomic bonds and use the two terms interchangeably in their explanations (see also Heckler and Rosenblatt, 2010). Atomic bonds are often described by instructors as being either "strong" or "weak". Unfortunately this can be misleading or confusing to the students because sometimes the word "strong" refers to the force of the bond and sometimes it refers to the bond energy. Like many misconceptions, the use of a common word can lead to difficulties in understanding the proper scientific concept. In everyday usage, "strength" usually refers to force, whereas normally when an expert speaks of a strong atomic bond, it is meant in terms of a large binding energy. In general we observed that it was common for students to use the terms force or energy when discussing the origins of macroscopic properties such as elasticity, strength and melting temperature, with little regard for the scientific accuracy of their own usage of the words. The failure to distinguish between energy and force in atomic bonds may be contributing to student difficulty in understanding how the properties of atomic bonds are related to macroscopic properties such as strength and elasticity.

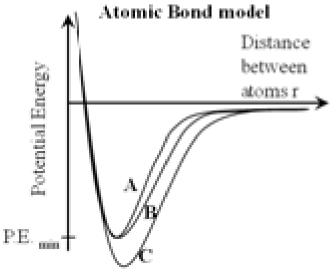
Finally, we found that students have considerable difficulties using a (Lennard-Jones) bonding potential energy graph to answer common questions about melting temperature, Young's modulus, and atomic separations. Comparing student's responses for the two questions shown in Figure 5.4; it is evident, from the almost identical responses for the two separate questions, that students do not understand what parts of the graph are related to the requested properties of the bond. In addition, students' responses to the even simpler question in Figure 5.5 illustrates that many students do not have even a basic understanding of the meaning of the graph.

5.2.2 The Atomic Bonding Tutorial

Student difficulties with understanding atomic bonding are addressed in a number of the tutorials to some extent, since the concepts are so basic and pervasive. Nonetheless, we developed a specific tutorial, which is presented in Appendix C, which focuses on the Lennard-

Shown below are the Potential Energy curves for three different materials. Circle the answer with the correct rankings of the three materials' melting temperatures.

(13%) $A > B > C$
(33%) $\mathbf{C > B = A}$
(11%) $C < B = A$
(42%) $C > B > A$



Shown below are the Potential Energy curves for three different materials. Circle the answer with the correct rankings of the three materials' stiffness (i.e. their Young's moduli, E).

(18%) $\mathbf{A > B > C}$
(30%) $C > B = A$
(11%) $C < B = A$
(41%) $C > B > A$

Figure 5.4: Example questions demonstrating student difficulties interpreting graphs of the Lenard-Jones potential. Student response percentages in parentheses, and correct answer in bold. (N = 61, SE = 6%.) Note: To save space the graph which is common to the two questions is shown only once and placed between the two questions. This was not the case when students saw the questions.

If the total energy of the system is EN, what is the smallest possible separation between the two atoms?

(3%) The smallest separation is zero.
(26%) The smallest separation is very small but must be greater than zero.
(54%) The smallest separation is r1.
(16%) The smallest separation is ro.

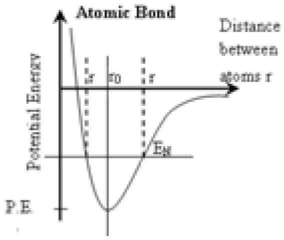


Figure 5.5: Example question demonstrating student difficulties interpreting a graph of the Lenard-Jones potential. Student response percentages in parentheses, and correct answer in bold. (N = 61, SE = 6%)

Jones potential energy of two metal atoms as a function of separation. First the tutorial provides a brief explanation of two common models for atomic bonding - the Lennard-Jones model and the spring model. Then students are prompted to compare the two models. This includes describing and comparing the motion of the atoms in the two cases, and how the average separation between the atoms changes when the average energy increases. We found that even though these concepts were mentioned in the textbook and in lecture, most students were not at all familiar with the Lennard-Jones potential or how it is related to the behavior of the atoms. Thus such simple questions were useful to familiarize students with the two models and their graphs. There are also a number of questions prompting students to use the potentials to address macroscopic properties of materials such as its melting temperature, Young's modulus, coefficient of thermal expansion, and the energy necessary to break bonds as a function of temperature.

We chose this tutorial activity for several reasons. First, a basic (and we stress basic) conceptual understanding of the major features of the potential is fundamental to understanding the nature of atomic bonds, and this understanding can be used throughout the course to help students think about the microscopic structure of a material. (This is discussed further in Chapter 6.) Second, students difficulties using the Lennard-Jones potential to answer simple questions, which was already, pre-tutorial, a part of the expected atomic bonding curricula for the course, would have suggested some instruction with the Lennard-Jones potential anyway. Third, the graph, is a convenient visual representation that facilitates understanding of the independence of the fundamental dimensions of bond energy, average separation, and curvature that can be conceptually linked to the independent macroscopic properties of melting temperature, number density, and Young's modulus. This can help students to more clearly distinguish between, and understand the independence of, atomic separation, bond energy, yield strength (related in part to depth of well), and elasticity (related to curvature of well). Finally, comparisons of the two models can assist in illustrating for the students when a spring model is not a good model to use for bonding, for example that it does not explain thermal expansion, which can be built upon later when students are learning plastic deformation.

5.3 Crystals: Atomic Packing Factor, Atomic Weight, and Defects

5.3.1 Student Difficulties Understanding Crystal Structure, Atomic Packing Factor, Atomic Weight, and Defects

Student difficulties with the crystal structure of metals are widespread going all the way from the simple, "A metal has a crystal structure," to the complex, how to visualize a dislocation in three dimensions. The three main difficulties that were found to be pervasive were difficulties with the concept of atomic packing factor, difficulties with the atomic weight and atomic percentage formulas and concepts, and difficulties with defects.

Student difficulties with atomic packing factor are illustrated in Figure 5.6 and 5.7. Many students could not recall post instruction the definition of atomic packing factor. When questioned in more detail it was clear that the knowledge that they had about atomic packing factor was factoidal and disconnected. For example, students might recall apf was 0.68 or that it was unit less but still have said that the apf was the volume, or the number of atoms, in the unit cell. In addition, most students did not have any conceptual idea of how apf was derived.

In a related topic, many students struggled with the formulas for atomic weight and atomic percent relations. When asked to reason qualitatively about atomic weight many

What is the definition of the atomic packing factor?

(25%) It is the number of atoms in a unit cell.
(16%) It is the volume of a unit cell that is occupied by atoms.
(20%) It is how tightly packed atoms are.
(38%) It is the fraction of a unit cell that is occupied by atoms.

Figure 5.6: Student difficulties with the definition of APF. Student response percentages in parentheses, correct answer in bold. (N = 155, SE = 3%)

What is the constraining relation between the side length of the unit cell, a, and the radius of an atom, r, for a body centered cubic crystal?

(43%) $\mathbf{4 * r = \sqrt{a^2 + a^2 + a^2}}$.
(29%) $4 * r = \sqrt{a^2 + \sqrt{2}a^2}$.
(5%) $4 * r = \sqrt{\sqrt{2}a^2 + 2a^2}$.
(21%) $4 * r = \sqrt{a^2 + a^2}$.

Figure 5.7: Student difficulties with the relationship between atom size and unit cell size. Student response percentages in parentheses, correct answer in bold. (N = 155, SE = 3%)

students made logical mistakes that lead to sign errors. When allowed to use calculators more students were successful but some still made formula mistakes which lead to a similar sign error.

Because of the visual nature of defects, no multiple-choice questions were created to test student understanding. However, a small number of students were given interview questions on this topic. Students were asked to draw what a metal looks like on an atomic level and then describe their picture. Students were also asked to sketch grains and dislocations. Common student mistakes were: drawing atoms randomly not in a crystal structure; drawing atoms in a crystal structure and then not being able to explain why the atoms were arranged regularly; and, drawing gains with lines representing atomic chains and not knowing why the lines were there.

5.3.2 The Properties of Crystals and Defects Tutorial

To address these difficulties with crystal structures the tutorial starts with an atomic packing factor activity which has students step through the derivation of the atomic packing factor of a bbc crystal by deriving the geometric relationship between the side length of a units cell and the size of the atoms creating the cell. After this derivation the tutorial then asks students a conceptual question that students commonly miss which highlights the use of apf as a unitless factor not dependant of atomic size. The goal of this activity is to ground

A material is made from 80% wt Mg and 20% wt of Cu. Given that Cu has a greater atomic weight, in grams per mole, than Mg, what is the atomic percent of Cu in the material?

(43%) less than 20%
(5%) equal to 20%
(47%) greater than 20%
(5%) Not enough information

Figure 5.8: Student difficulties with qualitative reasoning about atomic weight and atomic percent. Student response percentages in parentheses, correct answer in bold.
(N = 155, SE = 3%)

the apf definition and therefore aid in recall and conceptualization. Then the tutorial has students describe a set of defects in crystal structure and then represent each defect with a picture. These activities utilize both multiple representations and atomic structure sketching both of which are valuable teaching tools and are discussed in more detail in Chapter 6. The main goal of these activities is student understanding of the respective defects. However, these activities more generally help enforce the concepts of atomic level microstructure in metals. For example, that a metal has a crystal structure.

5.4 Diffusion

5.4.1 Student Difficulties with Diffusion

We have found that from the perspective of introductory materials science, virtually every aspect of diffusion is difficult for students to understand. It has been recognized that the process of diffusion is difficult to understand by a number of researchers (Streveler et al., 2008; Chi, 2005). Chi, for example has discussed this difficulty in terms of the perspective that diffusion is an emergent process, so the cause of diffusion is often misunderstood by students leading to confusion. In our studies, we also found these issues. However, we also found other significant post-instruction student difficulties with diffusion that are important to materials science, including a lack of understanding of the terminology, mathematical equations, and graphs characterizing diffusion and a lack of a consistent microscopic model and physical intuition of the process of diffusion in solids.

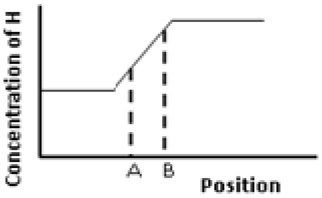
For example, students have difficulty with the concepts underlying Fick's first law, which states that the diffusion flux is proportional to the (negative of the) concentration gradient. Specifically, students often confuse higher concentration with higher diffusion flux rather than higher concentration gradient, and they do not have a physical understanding of why a gradient is necessary to obtain a non-zero net diffusion flux. Furthermore, they have difficulty understanding the meaning of concentration versus position graphs and making

The figure to the right shows the concentration of Aluminum as a function of position. How does the diffusion flux of Aluminum at point A compare to that at point B?

(8%) $A > B$

(29%) $A < B$

(63%) A = B



In the figure to the left, in which direction is there a net diffusion of Hydrogen at point A?

(51%) To the right (+ x direction)

(36%) To the left (- x direction)

(7%) The concentration profile is in steady state, so the net diffusion is zero.

(7%) There is no direction to the diffusion.

Figure 5.9: Example questions demonstrating student difficulty with diffusion and concentration vs. position graphs. Student response percentages in parentheses, and correct answer in bold. (N = 62, SE = 6%.)

meaningful inferences about diffusion from such graphs. Figure 5.9 provides an example of these difficulties, with about 30% of students considering the height rather than slope, and more importantly, only 36% of students being able to determine the direction of the net diffusion flux from the graph. This indicates that either students do not have any understanding of the relation between diffusion and concentration gradient or that they have a significant lack of understanding of the meaning of the position versus concentration graphs. Further interviews indicates that some students do have some physical understanding of the relation between concentration gradient and diffusion flux, but they are unable to connect this understanding with the graphic representation of position versus concentration.

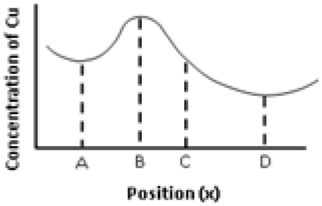
Another example of student difficulty with diffusion is with Fick's second law, which relates the rate of change in concentration with the second derivative of concentration with respect to position. For Fick's second law, the scope of students' difficulties becomes more pronounced. For example, as shown in Figure 5.10, only about 25% of students recognized that the concentration would increase with time at a local minimum in concentration, and only about 15% recognized that concentration would change the fastest at points of "high curvature" in concentration. Instead, when asked about the graph in Figure 5.10, most students chose according to the slope, presumably since we are asking how concentration is changing with time, and rates of change are often associated with slope. Unlike with questions regarding Fick's first law, students did not do better on these questions when pressed in interviews. This suggests a more serious lack of understanding of the graphs and the physical processes underlying Fick's second law.

For a final example, we consider student responses to a question that probes student understanding of the nature of diffusion in metals on an atomic level, shown in Figure 5.11. Only 36% of students correctly responded that a copper atom will diffuse through a sample

The figure shows the concentration of Copper as a function of position in a sample at 2000 C. At point A, how is the concentration of Copper changing with time?

(24%) **It is increasing with time.**
(4%) It is decreasing with time.
(45%) It remains constant with time.
(27%) It could be increasing or decreasing with time.

The figure below shows the concentration of Copper as a function of position. At which point is the concentration of Copper changing the fastest with time?



(1%) A
(16%) **B**
(67%) C
(16%) D

Figure 5.10: Example questions demonstrating student difficulty diffusion and concentration as a function of time. Student response percentages in parentheses, and correct answer in bold. (N = 376, SE= 3%.)

Consider a single particular Copper atom in a sample that is 100% Copper at 3000C. Which of the following is correct?

(17%) The Copper atom will not diffuse through the sample because the atoms are fixed in a crystal lattice.
(43%) The Copper atom will not diffuse through the sample because the sample is 100% Cu and therefore movement of atoms is not energetically preferred.
(4%) The Copper atom will diffuse through the sample because Copper atoms are not in a crystal lattice.
(36%) **The Copper atom will diffuse through the sample because of thermal energy.**

Figure 5.11: Example question demonstrating student difficulty the concept of atomic diffusion in a metal. Student response percentages in parentheses, and correct answer in bold. (N = 376, SE = 3%.)

of copper. From these responses and from interviews, it was clear that students do not have strong consistent models for diffusion in metals on an atomic level.

5.4.2 The Diffusion Tutorial

The tutorial starts by providing students with a series of conceptual questions about diffusion, such as questions concerning diffusion flux, concentration, and steady state diffusion. These questions highlight not only basic definitions and units of critical terms, but also how and why these terms are interrelated. Students then work through a series of questions related to Fick's equations, physical descriptions of diffusion, and the drawing of concentration vs. position graphs for the special case of steady state diffusion. These activities are designed to help students connect the concepts of the mathematical equations of Fick's laws with that of graphical concentration information. In addition, students work through an activity designed to help them connect the macroscopic properties of concentration and diffusion flux with what is happening at the atomic level through drawing of a very simplified atoms-in-a-bin picture of concentration and connecting this with a concentration vs. position graph and a diffusion flux graph (see Appendix C).

As a final activity, the tutorial guides students through a non-steady state concentration graph very much like the one in Figure 5.10. Students are asked several questions like, "Where is the diffusion flux greatest, where is the concentration greatest, ..." These questions are designed to help students evaluate their understanding and catch students who fall back on naive responses to give them a last chance to correct their reasoning.

5.5 Mechanical Properties

5.5.1 Student Difficulties with Mechanical Properties

We have identified two general and inter-related kinds of student difficulties with mechanical properties (see also Heckler & Rosenblatt, 2011; Kitto, 2008; Krause et al., 2010; Rosenblatt & Heckler, 2010). First, we found that even after instruction students often equated mechanical properties concepts and terms, used them interchangeably, or at very least thought the properties were necessarily correlated or anti-correlated. Perhaps the most prevalent and fundamentally important confusion was between the concepts of material strength and elasticity. Even if some of the students did understand the difference in the definitions, they often believed that the two properties must be correlated. That is, they believed a stiff material must also be strong and vice versa. This is highlighted by the questions in Figure 5.12 and 5.13. The first question concerns a conceptual definition of modulus of elasticity. Only one-third of the students answered correctly, with most students confusing the concept of yield strength with elasticity. This question requires a careful reading of the answer choices and is somewhat subtle, yet interviews revealed that student understood the options and chose purposefully.

The next question (Figure 5.13) is somewhat more straightforward, yet only 13% of the students answered correctly. Approximately 40% of students believed that the material with a higher yield strength will also have a higher tensile strength, which is not unreasonable, and over 40% (the majority) of students answered that the material with a higher yield strength, also has a higher tensile strength and higher modulus of elasticity.

It was also somewhat common for students to believe that there is a strict anti-correlation between yield strength and ductility, namely that a highly ductile material has low strength, as shown in Figure 5.14.

In addition to difficulties with confusion of mechanical properties, we found that students had difficulty with understanding and applying the basic concepts and quantities necessary

What is the Young's modulus of elasticity or 'stiffness' of a material?

(33%)	**A measure of a material's resistance to elastic strain when under stress.**
(19%)	A measure of a material's ability to return to its original shape after a load is applied.
(11%)	A measure of a material's ability to stretch or deform without breaking.
(37%)	A measure of a material's ability to withstand an applied stress without permanently deforming.

Figure 5.12: Example question demonstrating student difficulty with the concept of elasticity. Student response percentages in parentheses, and correct answer in bold. (N = 62, SE = 6%)

Two pieces of metal, A and B, are the same size and shape but Metal A has a greater yield strength than Metal B. Which of the following statements is true?

(13%)	**Metal A will permanently deform at a greater stress than Metal B**
(2%)	Metal A will have a greater tensile strength than Metal B
(2%)	Metal A will have a greater young's modulus of elasticity than Metal B
(40%)	Both a & b
(44%)	a, b, & c are all true

Figure 5.13: Example question demonstrating student confusion of the concepts of yield strength, tensile strength and elasticity. Student response percentages in parentheses, and correct answer in bold. (N = 67, SE = 6%.)

Which of the following is the best statement describing the relationship between ductility and yield strength?

(10%)	A metal with greater yield strength is more ductile.
(29%)	A metal with a greater yield strength is less ductile.
(10%)	A metal with greater yield strength tends to be more ductile.
(40%)	A metal with greater yield strength tends to be less ductile.
(10%)	**Ductility has no relation to yield strength.**

Figure 5.14: Example question demonstrating student confusion of the concepts of yield strength and ductility. Student response percentages in parentheses, and correct answer in bold. (N = 67, SE = 6%.)

The following metal pieces are cut from the same plate. Compare the yield strength of the pieces.

(60%) A has a higher yield strength than B
(8%) B has a higher yield strength than A
(32%) A and B have the same yield strength

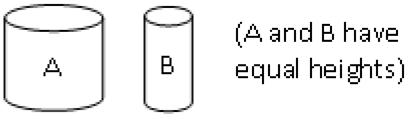


Figure 5.15: Example question demonstrating difficulty with the definition of yield strength. Student response percentages in parentheses, and correct answer in bold.
(N = 114, SE = 4%.)

Consider the stress-strain curves of two metals above. Which metal has a higher modulus of elasticity?

(46%) A has a higher yield strength than B
(54%) B has a higher yield strength than A
(0%) A and B have the same yield strength

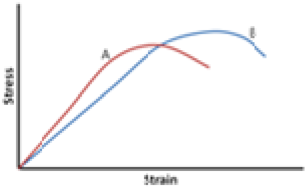


Figure 5.16: Example question demonstrating difficulty with interpreting stress-strain plots. Student response percentages in parentheses, and correct answer in bold.
(N = 116, SE = 6%.)

for defining mechanical properties. One place where this is manifest is in the common incorrect reasoning with questions involving yield strength, force and stress. In particular, students usually associate yield strength with force rather than stress. A dramatic example of this is demonstrated by student post instruction responses to the question in Figure 5.15. In this simple question, it seems that the majority of students believed that yield strength depended on cross sectional area, or put another way, that yield strength was defined in terms of force rather than force per unit area. In interviews, student responses were consistent with this: we found most incorrect students had considered only that the larger piece could withstand a greater force (rather than stress) without deforming. Furthermore, interviews and classroom observations revealed that many students used the terms force and stress interchangeably. When questioned further, most students did recognize the formal difference between the concepts stress and force. Nonetheless, they often failed to recognize that the two terms must be used carefully.

Difficulties relating stress-strain graphs and mechanical properties. The second kind of student difficulty with mechanical properties was in relating mechanical properties to a stress-strain plot. For example, Figure 5.16 presents results from a simple question comparing modulus of elasticity of two materials represented in two stress-strain curves. Over half of the students chose the curve that had the higher maximum value, rather than the curve with the steepest linear slope. This explanation that "higher position on graphs means more" was commonly found in interviews, and is similar to findings for kinematic graphs in physics courses (McDermott, Rosenquist & van Zee, 1987). Interestingly, some students thought that since the higher curve also had higher elongation until breaking, this was seen as additional evidence of higher elasticity. This example highlights the fact, as discussed earlier, that many students conflate concepts such as ductility, strength and elasticity, and this difficultly is manifest in (and confounded by) the reading of graphs. (Students difficulties with graphs and diagrams will be discussed more in Chapter 6.)

5.5.2 The Mechanical Properties Tutorial

The tutorial designed to address difficulties with material properties is shown in Appendix C. The tutorial provides various ways in which to apply and practice definitions and explanations of elastic deformation, Young's modulus, yield strength, and tensile strength. Special attention is given to terms students frequently confuse such as stiffness and strength or yield and tensile strength, and students were asked to explain the differences between these terms. Following a technique used in other tutorials (McDermott, Shaffer, and the PER Group at U. of Wash. (2002)), the tutorial also provides a set of quotes commonly made by students during interviews. Students are asked to comment on the correctness of the quotes. These "student dialog" questions are designed to demonstrate the necessity of precision in language and raise student awareness of common incorrectly stated definitions or generalizations, such as, "A tougher material is stronger," or, "A stiffer material is harder to break."

The mechanical properties tutorial also asks students to apply their mechanical properties definitions to stress-strain plots. This serves several purposes. It gives students experience deriving information from, and plotting information on, graphs which is itself a goal of instruction. It also provides a second way to think about the definitions and thus acts as both a check to students understanding and an additional way for students to distinguish exactly what parts of their written definitions were of importance. The easily separable dimensions of the graph (i.e. slope, height, peak, and line length) provide a clear visual aid for discussion, as a group or with a TA, of the exact differences in the properties and how one property does not necessarily affect another property.

5.6 Plastic Deformation and Strengthening

5.6.1 Student Difficulties with Strengthening

We found a number of student difficulties understanding plastic deformation and strengthening. Many of these difficulties pertain to students not having a good mental image of atomic structure and the process by which deformation in the material will take place, As the size of the grains increases, what happens to the strength of the metal?

(7%)	The number of defects increases so the strength increases.
(44%)	The number of defects increases so the strength decreases.
(8%)	The number of defects decreases so the strength increases.
(38%)	**The number of defects decreases so the strength decreases.**
(3%)	The strength of the metal is not affected by the size of the grains.

Figure 5.17: Student difficulties with grain size effects of strength. Student response percentages in parentheses, and correct answer in bold. (N = 64, SE = 6%)

i.e. dislocation motion. In addition, as mentioned in the mechanical properties section, some of the difficulties students had were caused by their mistaken ideas about mechanical properties being transferred to interfere in this new topic. Examples of both of these two difficulties can be seen in student responses to the multiple-choice questions in Figures 5.17, 5.18, and 5.19. (Note that these multiple-choice questions were created from student responses to more open ended questions. So, student responses represent actual student ideas.)

Students have a hard time with questions, even simple ones, that pertain to structure. We asked students in posttest interviews to: "draw a picture of a grain" and "draw what a metal looks like on the atomic level". Many students struggled with this activity. So, perhaps student responses in Table 5.17 should not be surprising. However, part of what is so interesting about the student responses is that they are wrong in both pieces of the question. First, they do not have a good understanding of a grain and thus treat the grain as the defect, i.e. increasing grain size implies increasing defects. Then, they do not have a good understanding of strength and thus treat defects as bad for strength, i.e. increasing defects implies decreasing strength.

Student difficulties with mechanical properties and how this effects student understanding of strength can be seen in both Figure 5.18 and 5.19. Many students say that a colder object is not as strong because it is more brittle. A colder object is more brittle; however, this property indicates a low toughness and not a low strength which is the property in question and which in a colder material is greater. This is yet another example of a misunderstanding that arises from every day and scientific definitions being at odds with each other.

In addition, both of these questions show the previously noted difficulty with confusion of strength and density, or bond length. And, question 5.19 shows the difficulty with confusion of strength and cross sectional area.

Two metal rods A and B are cut from the same plate and are at room temperature. The temperature of Rod A is lowered by 100 degrees Celsius and held at that temperature for a long time. Nothing is done to Rod. B.

(40%) Rod A has a lower yield strength than Rod B because it is more brittle.
(20%) Rod A has a higher yield strength than Rod B because it has shorter bond lengths.
(7%) Rod A has a higher yield strength than Rod B because it has a lower number of defects.
(18%) Rod A has a higher yield strength than Rod B because it has slower dislocation motion.
(17%) Rod A has equal yield strength to Rod B because temperature does not effect yield strength.

Figure 5.18: Student difficulties with temperature affects on strength. Student response percentages in parentheses, and correct answer in bold. (N = 64, SE = 5%)

Two thin metal rods are cut from the same plate. Rod A is pulled through a tapered hole smaller than the rods original diameter. Nothing is done to Rod. B.

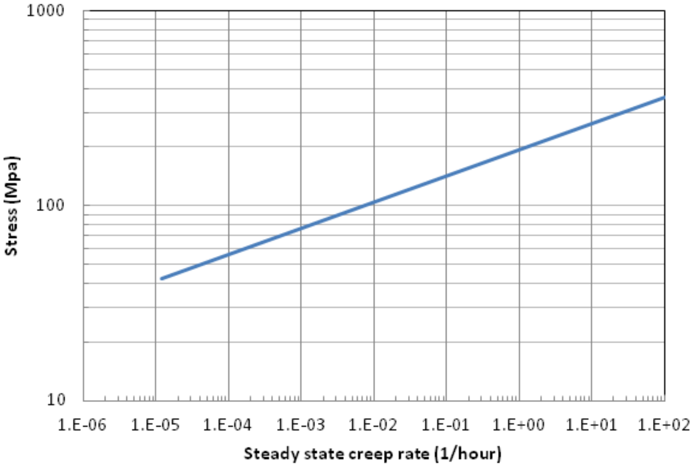
(33%) Rod A has a higher yield strength than Rod B because it has more defects.
(24%) Rod A has a higher yield strength than Rod B because it has become denser.
(29%) Rod A has a lower yield strength than Rod B because it is now skinnier.
(3%) Rod A has a lower yield strength than Rod B because it has already been under stress.
(10%) Rod A has equal yield strength to Rod B because they are the same material and have the same composition.

Figure 5.19: Student difficulties with coldworkings effect on strength. Student response percentages in parentheses, and correct answer in bold. (N = 64, SE = 6%)

5.6.2 The Plastic Deformation and Strengthening Tutorial

This tutorial starts by having students do a review activity drawing and describing grains and dislocations. While students do this in the crystal structures tutorial, many students still struggle with this topic so the review is helpful for them. Then, students work through a series of questions about different ways to strengthen metals. These are grain size, coldworking, and solid solution alloying. These questions are designed to elicit student thought and discussion about how and why the microstructure changes are effecting the properties of strength. Next, the tutorial takes students through a graphing exercise where they recreate the graphs seen in the text which show changes to a materials strength and ductility due to the amount of coldworking done. This is to help student better understand these graphs, which they are commonly asked to use for tests, and to reinforce their understanding of

The figure below presents a stress vs. steady state creep rate graph for a steel alloy at 925C. What is the creep rate when the stress is 70 MPa?



33% of students responded with answers between $2.9 * 10^{(-4)}$ and $6.1 * 10^{(-4)}$ which were considered correct.

Figure 5.20: Example question demonstrating difficulty with log plots.
(N = 102, SE = 5%.)

coldworking and its effect on a material.

5.7 Failure: Creep, Fatigue, Fracture, and Surfaces

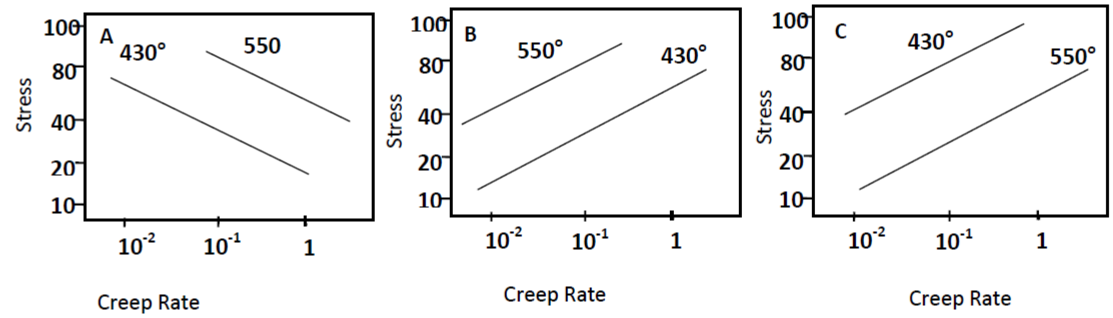
5.7.1 Student Difficulties with Failure

Less is known about student difficulties with this topic. It was not one of the original areas researched so less exploration was done in this area. A majority of student difficulties found were identified from existing test questions used by previous instructors. Because of this, one of the major difficulties found with the topics of creep and fatigue is an inability to use creep and fatigue graphs to answer questions.

The difficulties that students have with creep and fatigue graphs are varied. Many students have difficulty with the use of logarithm scales as part of one or more of the graph axis. (For example, see Figure 5.20) They also have difficulties applying simple conceptual knowledge to a graphical question such as the one shown in Figure 5.21 which simply asks students which line is at a higher temperature. When students are asked if creep is higher or lower with increasing temperature, a majority of them can given the correct answer. However, many of them cannot, or do not, transfer this knowledge to the graphical question.

In addition, students have a hard time recalling and understanding the causal mecha-

A metal is put under tension, and its creep rate as a function of applied tension stress is measured at two different temperatures, 4300C and 5500 C. Which pair of lines A, B, or C are possible graphs of the measurements.



(8%) A
(49%) B
(38%) C
(3%) None of the above are possible

Figure 5.21: Example question demonstrating difficulty with graphs for a simple creep question they would normally be able to answer. Student response percentages in parentheses, and correct answer in bold.
(N = 155, SE = 4%.)

nisms of each type of failure. This is true for topics that instructors would consider difficult, e.g. creep is diffusion related failure usually occuring at grain boundaries, and for topics that instructors would consider easy, e.g. that creep and fatigue occur at stresses below the yield strength of the material.

5.7.2 The Creep, Fatigue, and Fracture Tutorial

The first activity in the tutorial has students give explanations of both creep and fatigue. Students are asked to consider underlying causal mechanisms for the different failure types, and while no comparison activity is required, the close proximity of the two is expected to help students recall and distinguish the two types of failure. Special attention is paid to creep and its dependence on temperature and occurrence below the yield strength.

The next activity is not aimed at a student difficulty but at a goal for the course. One of the main goals for the course is student understanding of materials properties, tradeoffs, and selection. Failure is an excellent topic for addressing this goal. This activity asks students to consider two different systems (e.g. bike, airplane, etc.), how that system's components can be susceptible to creep and fatigue, and improvements that can be made to the system to reduce the risk of failure. TAs often have to provide students with some special support and encouragement to get them to think deeply about this, but interesting discussions can take place if it is done right.

Consider the Copper -Nickel phase diagram to the right. Which of the following is true about the composition of the solid alpha phase at point B in the graph?

(0%)	0% wt Ni
(7%)	Less than 50% wt Ni but greater than 0%
(15%)	50% wt Ni
(75%)	**Greater than 50% wt Ni but less than 100%**
(3%)	100% wt Ni

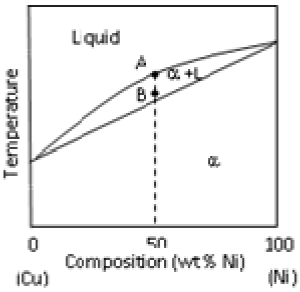


Figure 5.22: Example question demonstrating student difficulty with phase diagrams for a single solid phase. Student response percentages in parentheses, and correct answer in bold. (N = 64, SE = 6%.)

Students then work through two graphing and calculation activities - one for creep and one for fatigue. These activities require students to use the respective definitions and transfer them to this new abstracted graphical situation. It also gives them experience with log plots and reading from and creating graphs.

The final activity in this tutorial has students work through a fracture toughness and crack size calculation. Students often have difficulty reasoning with fracture toughness formulas and this gives them experience with one example.

5.8 Phases and Phase Diagrams

5.8.1 Student Difficulties with Phase Diagrams

Perhaps not unexpectedly, students have a number of difficulties with phase diagrams. These difficulties appear to arise from both an inability to understand the nature of the diagrammatic representation of phases and a lack of understanding of the nature of phases and concepts relating to phases such as the difference between composition of a phase and the phase fraction of an alloy.

In general, many students have significant difficulty extracting relevant information from phase diagrams, and performance decreases rapidly with increasing complexity of the diagram. For example, as shown in Figure 5.22, post instruction, over 75% of students can typically answer simple questions about binary phase diagrams involving solid solutions, i.e. only one solid phase. However, as shown in Figure 5.23, student performance dramatically decreases for questions about binary eutectic diagrams. Only 38% of students correctly picked fraction of α with almost as many choosing the composition of the α. The rest of the students were divided between the composition amounts for the solution as a whole. In interviews and classroom discussions with students it was clear that much of the difficulty

Consider the Copper-Nickel phase diagram to the right. Which of the following is true about the composition of the solid alpha phase at point B in the graph?

(34%) 6%
(38%) 26%
(10%) 28%
(15%) 72%
(0%) 94%

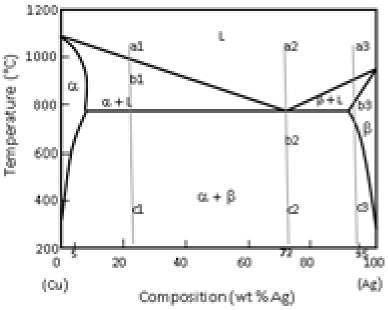


Figure 5.23: Example question demonstrating student difficulty interpreting eutectic phase diagrams. Student response percentages in parentheses, and correct answer in bold. (N = 212, SE = 3%.)

In the phase diagram above, what is the α phase?

(51%) It is only Copper Atoms.
(7%) It is only Silver Atoms.
(10%) It is a mixture of Copper and Silver atoms with a specific fixed wt.% of each.
(32%) It is a mixture of Copper and Silver atoms with a wt.% of each that depends on temperature.

Figure 5.24: Example question demonstrating student difficulty with the nature of a phase in binary phase diagrams. Student response percentages in parentheses, and correct answer in bold. (N= 155, SE = 4%.)

with this question arises from the complex nature of the graph itself and thus from students not knowing what to attend to. However, there are also a significant number of students who have difficulty either understanding the differences between the concepts of fraction and composition as well as the meaning of phase.

Student difficulties with understanding the meaning of phase is also demonstrated in the question in Figure 5.24, in which most student identified the phase as being comprised solely of one of the two elements in the alloy. Interestingly, many students who incorrectly believed in "pure phases" still successfully performed lever rule calculations, which inherently assume that the composition of phases is mixed. (This is consistent with previous finding by Demetry (2006).) Thus, it becomes clearer that part of the difficulty students have with phase fraction and phase compositions terminology is that they do not have a correct understanding of the nature of the α and β phases and thus what the graph itself is representing.

Refer to the TTT diagram to the right. Austenite at the eutectoid composition is quenched from 750C to 625C and held for 10 s. It is then quenched to 300C and held for 10 s. Finally, it is quenched to room temperature (25C). What is the final microstructure?

(5%) 100% austenite

(14%) 100% martensite

(21%) 50% pearlite + 50% austenite

(59%) 50% pearlite + 50% martensite

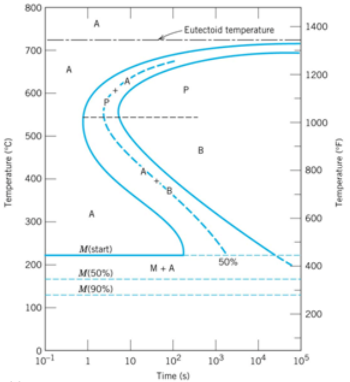


Figure 10.13 The complete isothermal transformation diagram for an iron–carbon alloy of eutectoid composition: A, austenite; B, bainite; M, martensite; P, pearlite.

Figure 5.25: Example question demonstrating student difficulty with the nonstandard diagram specific rules of TTT plots. Student response percentages in parentheses, and correct answer in bold. (N = 155, SE = 4%.)

5.8.2 The Phase Diagram Tutorial

The tutorial on phase diagrams is aimed at addressing student difficulties with the diagrams, including those described above, by guiding the students through a series of general questions about the nature of phases, the meaning of solubility for metallic alloys, and the meaning of the regions of the phase diagram. The tutorial also guides the students to describe the phases on both an atomic and macroscopic level at various temperatures and compositions as well as transformations that occur as an alloy of a given composition changes temperature. Exercises include drawing pictures of microstructure and calculations of composition and fraction. A sample of the tutorial's questions is shown in Appendix C.

5.9 TTT Diagrams

5.9.1 Student difficulties with TTT Diagrams

The major difficulty that students have with TTT diagrams is simply not knowing the rules that go with this nonstandard plot. As can be seen by student responses to the question in Figure 5.25 and follow-up interviews, students are unsure what to do to determine effects of multiple coolings and a large number of them ignored future cooling past the first one. Also, many students treat the plot like a graph and ignore the importance of the path taken to get to a certain point.

In addition, when students were asked relatively straight forward questions about the materials shown in the diagram, they often struggled. (For example, what does the p stand for? What is pearlite? Is it stronger than martensite?...) These difficulties can be seen

Which of the following will always be effective in distinguishing a difference between a sample of austenite and a sample of pearlite?

(12%) The weight percent of Carbon is different between the two materials
(36%) The phase(s) present in the two materials are different
(47%) Both a and b
(5%) There are no composition or phase differences between the two materials

Figure 5.26: Example question demonstrating student difficulty with the materials shown in a TTT plot. Student response percentages in parentheses, and correct answer in bold. (N = 155, SE = 4%.)

Which of the following is the correct ranking of the metals, from greatest to least strength, for a fixed wt% of carbon?

(18%) Martensite, Austenite, Pearlite
(9%) Austenite, Martensite, Pearlite
(15%) Pearlite, Austeniste, Martensite
(16%) Pearlite, Martensite, Austenite
(39%) Martensite, Pearlite, Austenite

Figure 5.27: Example question demonstrating student difficulty with the strength of materials shown in a TTT plot. Student response percentages in parentheses, and correct answer in bold. (N = 155, SE = 4%.)

by students' responses to questions shown if Figures 5.26 and 5.27 of which only 40% of students could correctly answer post instruction.

5.9.2 The TTT Diagram Tutorial

The TTT tutorial takes students through the process of learning about the diagram. The first thing that students do is name and describe the A, B, P, and M seen on the diagram. Then it has students describe what is made, cooling times, and phase transformation times for four different paths on the diagram. These are designed to elicit student mistakes so that they can confront them through dialog with each other or the TA.

The next activity has students consider the microstructures that are formed during the different cooling paths and then rank the different paths by strength and ductility due to their knowledge of these microstructures. These questions are designed to help the students learn the microstructures and, more importantly, to reinforce the understanding that properties of materials are based on microstructure.

5.10 Additional Tutorials: Ceramics, Polymers, Composites, and Electronic Properties

The additional tutorials available were all used in class at least once. Ceramics and polymers were the only two included in the set of tutorials at the back and have both been used more than once. These two tutorials, like all of the others, are aimed at a couple of common difficulties students have with these topics and are based on the educational techniques of peer interactions and discussions, multiple representations, and drawing and visualization. They rely heavily on comparisons with metals - a topic students should already be familiar with. This serves to remind students about older material like the different mechanical property definitions, crystals structures, and stress-strain plots, but it also assists students in structuring and integrating this new material with what they already know.

5.11 Summary

In this chapter we reported on eight different areas of student difficulties with materials science and the tutorials that go with them. These areas are atomic bonding, crystal structure, diffusion, mechanical properties, strengthening mechanisms, failure, phase diagrams, and TTT plots. Student difficulties in each of these areas have been well established by student interviews as well as testing in both short-answer and multiple-choice format. Some of these difficulties are very concrete, elaborate, and systemic in the students' minds such as their difficulties with bond length and bond strength. Other difficulties are more superficial and represent a lack of knowledge more than a 'misconception'. However, both types of difficulties lead to failures in student learning and performance, and both types of difficulties can be helped by student participation in the tutorial sessions.

Chapter 6

The Course Goals and Pedagogy Used to Address Student Difficulties with These Goals

6.1 Three Main Course Goals and Tutorial Activities Addressing Them

So far, we discussed how the tutorials address specific content topics, such as mechanical properties or phase diagrams. However, there are also several general themes that run through the tutorials; these themes are aimed at goals the faculty and instructors identified for the course as a whole, as outlined in Table 4.4. In this section, we describe how the tutorials address these general course goals.

6.1.1 Addressing the Goal of Student Understanding of General Concepts, Basic Definitions, and Terminology

The main difficulty that students have in this area is that they commonly conflated similar terms and concepts. Examples discussed in the previous sections include confusion between force and energy of a bond, strength and elasticity of a material, and phase composition and phase fraction. Interestingly, after finding initial evidence of confusion between two given concepts or terms such as force and stress, we found that following further conversation many students could distinguish between the two concepts in question. Therefore in some cases, rather than finding that students could not make the distinction between the concepts in question, we found instead that students often simply did not distinguish between them. For example, after brief conversation, students would often easily grasp (or recall) the distinction between stress and force. While these are precise terms in the domain of engineering, students nonetheless often appeared to equate the concepts, or they often used the terms interchangeably.

Furthermore, in some cases it appeared that the confusion of terms was partially due to common language usage. For example, when referring to a property of a material in everyday language, stiff is often synonymous with strong or tough. However, these terms are not synonymous in materials science and have precise and critically different meanings.

Therefore, there appears to be two issues associated with student incorrect usage or application (interpreted as confusion) of similar terms and concepts. First, the students must learn the distinction between the concepts in question. This may or may not be difficult depending on for example whether the differences are subtle or whether both concepts are complicated. Second, it may be the case that in everyday experience the two concepts or terms in question are habitually used interchangeably, and even if a distinction is understood by the students, they may not recognize the need to distinguish between the concepts or terms. This second issue then involves the student learning that the distinction is important in some circumstances, especially in matters concerning material science.

The tutorials were designed to address these issues in a number of ways. Some tutorials explicitly ask students to provide a definition of a term in non-technical words. For example, the phase diagram tutorial starts with the question, “What is a phase?” For some students, this is a challenging question because they often use the everyday meaning of the word. In addition, commonly confounded terms were often directly confronted by asking students to describe the difference between, for example, stiffness and strength or stress and force. Another methods used was to provide quotes from students that used terms incorrectly (such as the student quote in exercise 10 in the mechanical properties tutorial shown in Appendix C), and student were asked to discuss the validity of the quote. Finally, many terms, such as strength, were often used over multiple tutorials. As part of the intervention, TAs were instructed to be aware these oft-confused terms, and assist students in recalling and defining these terms when students were stuck on a question. For example, a student might say, “We are stuck on #4, how does cold working strengthen a material?” The TA might respond by asking the student, “Last week we talked about tensile strength. What is tensile strength? Can you recall how it differed from yield strength?”

6.1.2 Addressing the Goal of Student Understanding of the Relation Between Structure, Processing, Properties

Achieving the understanding of the relation between structure processing and properties even at a basic level is complicated because it relies on multiple areas of student knowledge and the knowledge that bridges these different areas. While more tutorial development is needed on this topic, the tutorials in their current form do address important aspects. We found that many student difficulties in this area are rooted in either a lack of a conceptual (or visual) model or an incorrect model of the structure of materials on sub-macroscopic scales. This is true even after they finish the course. For example, as mentioned earlier, students often believe that bonds can be permanently stretched, and this interferes with learning of the proper description of plastic deformation (and yield strength).

In addition, students have difficulty producing visual models of the crystal structure nature of metals, the grains and dislocations (and how these differ), and the phases of metals at the microscopic level. This lack of correct visual models of materials on smaller scales hinders students' abilities to understand how processing can effect macroscopic properties of the material. For example, drawing a wire through an aperture, changes microstructure creating a greater number of dislocations, and thus increasing the materials strength.

To help address these difficulties, most tutorials include a section which requires students to draw pictures representing important atomic or microstructure-sized features and often has questions requiring students to analyze or use their picture. Examples of this can be seen in each of the tutorial exercises presented in previous sections. For example, in the mechanical properties tutorial students are asked to sketch what the atoms are doing before, during, and after elastic and plastic deformation. Then they are asked to use their picture to infer how density or atomic separation changes. Another example of this can be seen in the tutorial aimed at difficulties with crystal structure and defects where students are required to both draw a dislocation and give a written description for it, or in the failure tutorial where students are required to identify the types of failure seen in magnified fracture surfaces and then to describe the identifying features of the failure. This kind of exercise is consistent with work by Krause et al., which indicates that drawing exercises may generally be an effective teaching tool for addressing conceptual difficulties with materials science (Krause, Kelly, Tasooji, Corkins, Baker, and Purzer (2010)).

While the process of drawing may have specific instructional advantages, see the above paragraph, there is evidence that multiple representation in general are excellent teaching tools. Van Heuvelen is perphase the best known proponent of multiple representations in PER. For example, see his work with multiple representations in problem solving (Van Heuvelen (1991)) and the development of energy bar charts (Van Heuvelen and Zou (2001)). However, there are many other papers on this topic that examine, for example, expert and novice uses for representations, modeling theory and how it uses multiple representation, problem solving, etc (Dufresne, Gerace, and Leonard (1997); Hestenes (1987); Larkin, Mcdermott, Simon, and Simon (1980)).

6.1.3 Addressing to Goal of Student Ability to Interpret and Use Graphs and Diagrams

A wide variety of graphs and diagrams are used in the many topics covered in introductory materials science, and we have found that there are significant student difficulties with all of them. This includes the Lennard-Jones potential graph, concentration versus position plots used to describe diffusion, stress-strain plots, phase diagrams, log plots, and TTT plots. Interestingly, the understanding of graphs and diagrams was not initially stated by instructors as a goal for the course. However, when the level of student difficulties with them became apparent and was shown to the instructors, they all agreed that this was indeed a goal of the course but had been so systemic in the course materials and instruction as to be implicit.

The difficulties students have with graphs and diagrams differs widely. For relatively simple diagrams, such as concentration vs. position and stress-strain plots, students displayed slope-height confusion, similar to well-known student difficulties with kinematics graphs in physics (McDermott, Rosenquist, and van Zee (1987)). For example, Figure 5.9 shows students tended to base answers on height rather than slope for concentration vs. position graphs and questions about diffusion. Additionally, from the answers to Figure 5.9, it is clear that the students do not understand the meaning of the graph, or the relation between diffusion and concentration profiles. Another example is that in Figure 5.16 which presents results from a simple question comparing modulus of elasticity of two materials represented in two stress-strain curves. When this question is given pre-tutorial, over half of the students chose the curve that had the higher maximum value, rather than the curve with the steepest linear slope. This explanation that “higher position on graphs means more” was commonly found in interviews.

Perhaps as to be expected, students also had great difficulty with novel, unfamiliar graphs and diagrams, such as the Lenard-Jones bonding potential energy graph, phase diagrams, and isothermal transformation diagrams, i.e. TTT plots. Interviews revealed that the difficulty was two-fold. First, the students did not grasp the underlying concepts represented in these diagrams. For example, students responses to the question in Figure 5.24 show that many students have no concept of what is being represented by the biamary phase diagram and/or no attendance to or understanding of the x-axis shown in the diagram. Second, the students were unfamiliar and unpracticed in reading the diagrams and understanding the “rules” of the diagram. A very good example of this can be seen in student responses to the question shown in Figure 5.25 about TTT plots. Many of the students who chose all martensite do not understanding the importance of the path taken to a region treating the diagram like a functional representation. Also, many of the students who chose a mix of pearlite and austenite reported being unsure what to do with subsequent coolings and settling on their answer by default given that it was at least true of the first step.

Students’ difficulties with graphs and diagrams are often difficult, if not impossible, to separate from their difficulties with the concepts being represented. However, since the goal of our instruction is to improve both the tutorials use this to their advantage. Each tutorial uses graphs and plots in at least one activity. In these activities students derive information from or plot information on a graph or diagram. This helps students to understand the diagrams giving the students practice with each type of diagram, and its specific axis and rules. In addition, these activities often include a series of questions regarding the chart or graph which address common specific difficulties with a particular diagram. Often the task is designed to elicit and resolve that specific difficulty. For example, some students are not sure what is being represent by the Lenard-Jones potential. Questions #2 and 3 in the atomic bonding tutorial addresses this topic. Another example is the question on cooling times in the TTT tutorial which is simply testing if the student can read from the log x-axis.

On the other hand, each tutorial's diagram also helps students to learn the concepts represented in the diagram. The use of multiple representations as a teaching tool is discussed already in the previous section. Having students transform concepts into factors on a diagram and vice versa helps students to understand the concepts. A good example of this is in the mechanical properties tutorial. As discussed in Chapter 5 students have difficulties recalling the different mechanical properties and distinguishing them from each other. They also have difficulty with the fact that the properties are not correlated with each other. Question #12 on the mechanical properties tutorial requires students to transform their verbal definitions they have already been working with into graphical ones, and the disapearte parts of the graph which give the various properties can be used to illustrate how/why properties do not correlate.

Chapter 7

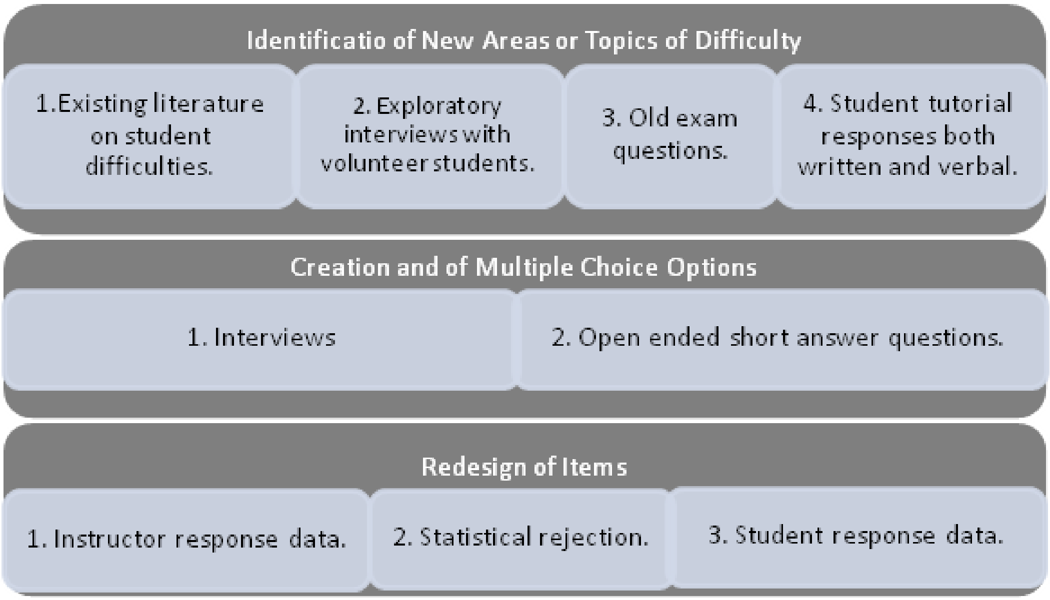
A Materials Science Conceptual Evaluation

7.1 Introduction

The creation of the Materials Science Conceptual Evaluation was the result of the construction of several multiple-choice items which were created to analyze areas of student difficulties for tutorial design and to assess student gains in these areas post tutorial. This ancestry has two main effects on the MSCE and the reported data. First, the test was item driven meaning that items were added, removed, or edited as needed. So, it was not so much the test that was developed as the items. Different versions of the MSCE test have several items that are identical. However, comparisons between the versions should be made carefully not only because the different versions are different but because they were given to students undergoing different research conditions. For example, tutorials were used/not used, recitation attendance was graded/not graded, and so forth. (See Tables 7.1 and 7.2 for a summary of differences between the versions and the student populations tested.) Second, the test is very broad in scope because it needed to cover the entire course. Because of this breadth, the term conceptual "Evaluation" is used. The test is not a concept "Inventory," which would have more variety of questions on a single topic, and it is not designed to be used as a concept inventory.

Here we report on the construct and content validity of the items. (See Chapter 2 for a discussion of test validity.) This is done mainly through a description of the process of item development and examples of items that were removed or revised. This process created items that were able: to assess the presence of certain topics of difficulty with materials science concepts; to correctly group students into their specific difficulty with each topic; to ensure that student ability on each topic was of value to instructors; and finally, to ensure that question wording was technically correct and clear both from an instructor and student view, which are not necessarily the same. In addition, we report on several post construction measures of validity which would indicate weak items or tests, and from

Figure 7.1: The main contributors to each stage of the item design and redesign process



which, nothing of concern was noted. These measures are correlations with student grade in the course, inter-item correlations and test reliability, and correlations with student use of tutorials - which are effective instructional materials as shown by results in Chapter 4.

7.2 Item Design Process

7.2.1 Initial Item Construction

There were several ways that new items were created for the MSCE and several ways that items were redesigned once they were created. A summary of these is presented in Table 7.1. The main contributors to the identification of new areas or topics of student difficulty were: existing literature on student difficulties, exploratory interviews with students, old exam questions, and student tutorial responses both written and verbal.

Once a topic or area was identified, subsequent student interviews and open ended test questions assessed the specifics of that difficulty and provided a general understanding of the number of students who exhibited that difficulty. Then, a multiple-choice question was created using common mistakes exhibited by the students in their short-answer responses as distractor options.

7.2.2 The Item Redesign, Validation, and Rejection Process

Once an item was created, it went through a "trail run" with a small number of students. During this time many items were validated by requiring students to both select an answer and then give a short written response explaining their thought process. This was used to assess the item's ability to group students into their actual idea state. All except for a few of the MSCE test items were created in this fashion. Subsequent rejection or redesign of these items was based on three main factors: instructor response data, statistical rejection, and student response data.

Instructor Response Data

The first way that items were reworded or removed was instructor response data. Three course instructors, who were faculty in the materials science department, analyzed the MSCE.0 items for: correctness, the importance of inclusion of each question on the MSCE test, and concern over students' response percentages. (These last two, importance and concern, were likert rankings 0 to 10 and were used as an indicator of instructors' views of item importance and value.) Any questions where these three things were not highly rated by a majority of the instructors were reevaluated and either removed or reworded. Additionally, items were reevaluated each quarter by a group of four or five teaching assistants for the course. Changes to question wordings were made based on their suggestions.

Statistical Rejection

Statistical rejection was the second way in which items were chosen to be removed or kept. Items with poor test correlations, as shown by low or negative point biserial coefficients, were eliminated from the test unless there was a specific reason for their continued inclusion such as high instructor approval of the item. Items with difficulties that were either too high or too low were also removed from the test, or they were reworded and tried a second time to ensure that they were indeed too easy or difficult as a concept and not simply as an item.

Examples of questions which were statistically rejected or reworded are shown in Figure 7.2 and Figure 7.3. The first of these was simply too easy of a question. While given an average score of 9 on inclusion in a concept evaluation, the score was 90%, and the point biserial was 0.027. This means that the items was giving no indication if a student was struggling with the concepts of diffusion. The second question received borderline approval by instructors for its importance on a concept inventory. (It was ranked 29th out of the 43 questions for inclusion.) However, the point biserial, -0.280, on this question was by far the lowest of all the items. This implies that students who did well on the test as a whole were likely to do poorly on this question.

Which of the following will increase the diffusion rate of atoms in a metal?

(90%) Increase the temperature of the metal.
(5%) Increase the activation energy of the metal.
(2%) Increase the size of the atoms.
(3%) None of the above will necessarily increase the diffusion rate.

Figure 7.2: A too easy diffusion question. (N = 47, STE = 4%.)

Shown below are the Potential Energy curves for three different materials. Circle the answer with the correct rankings of the three materials' stiffness (i.e. their Young's moduli, E).

(18%) $A > B > C$
(30%) $C > B = A$
(11%) $C < B = A$
(41%) $C > B > A$

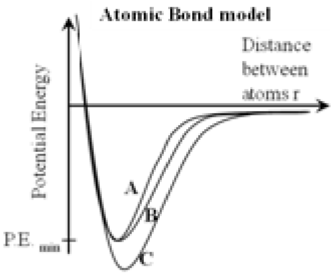


Figure 7.3: Example of a statistically rejected item. The Point Biserial was -0.0280 which is bad. This question was rejected for future MSCE test versions. (N = 47, STE = 5%.)

Student Response Data

Student response data was the main way that questions were validated and the third way for items to be rejected or, more often, reworded. As reported above, many items were validated by requiring students to both select an answer and then give a short written response explaining their thought process. In addition, much more informal analysis was done via group posttest interviews on later versions of the MSCE. After students took the test they would discuss the answers with each other in small groups in the presence of a test administrator. Questions which students indicated were "tricky" or "confusing," or if students were able to answer the question correctly in their discussions but had not been able to answer it correctly before, were reconsidered and reworded or removed as necessary.

7.3 Statistical Validity Measures on the MSCE Tests: Averages, Reliability, and Final Grade Correlations

Table 7.1 gives brief descriptions of the different MSCE test versions and how they differ from the previous version. Table 7.2 reports the number of items on each test version, the population tested with each version, and the different tests' average score, standard deviation, reliability (KR 20), and correlation with final exam grade. As can be seen, there are no statistical abnormalities in this data that would suggest invalid tests. The averages on

Table 7.1: Summary of changes made between versions.

Test Version	Brief Comparison of Version Items
MSCE.0	This is the longest version of the test. The 43 items, which had been created through an iterative process of collection and validation through interviews and testing , were given to a small population of students to collect response statistics. Then these items were given to three faculty to evaluate. Instructors rated, 0 to 10, the importance of inclusion of each question on the MSCE and rated, 0 to 10, their concern over students' response percentages. In addition, they indicated the correct answers and made comments about the question wording and content.
MSCE.1	This version of the test is a subset of the questions of version MSCE.0 based on faculty evaluations and statistical reliability and validity.
MSCE.2	This version of the test included changes to the wording of four questions in MSCE.1. Students were indicating these items were confusing, and/or we felt these items were not properly identifying which students understood the concept and which did not. In addition, two questions were removed from MSCE.1 in favor of two different questions from the MSCE.0 version which were more central to the course material.
MSCE.3	This is the newest version of the MSCE. It has 31 items. It differs from MSCE.2 in the diffusion, phase diagram, and creep questions. Three of the diffusion questions testing direction of the flux and change of concentration with time were dropped in favor of two other questions testing the same topic. An option was added to the phase diagram questions so that all possible solid phases are available answer choices, and a small change was made to two of the pictures to make the phases look more realistic. Lastly, one of the creep questions was changed so that its wording was more clear.

the different tests were neither at floor or ceiling for any of the different student groups. The reliability was low but within normal ranges, and given the previously discussed status of the test as a concept evaluation, it would not be expected to have a especially high reliability because the items are testing different topics. (Also, because the testing environment is not high-stakes, the reliability is likely to be lower than if it was high-stakes.) In addition, each quarter shows reasonable correlations with grade on the final exam, a measure of student ability in materials science.

7.4 Gains in Item and Test Score with Tutorial Use

While different versions of MSCE were given in different quarters, an analysis of the available data shows that students do better on the MSCE with tutorial use in class. And, the more students use the tutorials the better they do. The MSCE.2 was given as flex homework

Table 7.2: A summary of MSCE test versions and data collected on each.

		Population Tested							
Test Version	Number of items	N	Tutorials	Attendance	Recitations Graded	Average	STD	KR 20	Grade-Score correl.
MSCE.0	43 items	N = 47	Not used	30% (est.)	Ungraded	44%	12%	0.562	0.290
MSCE.1	32 itmes	N = 102	Used	36%	Ungraded	60%	15%	0.745	0.387
MSCE.2	32 items	N = 221	Used	62%	Ungraded	56%	18%	0.803	0.623
		N = 155	Not used	30% (est.)	Ungraded	47%	15%	0.735	Unavailable
MSCE.3	31 items	N = 199	Used	93%	Graded	68%	17%	0.791	0.443

in two quarters, one quarter that was using the tutorials in nongraded recitation sections and one that was not using them. In the quarter where tutorials were not used, the course format was identical, but the recitations were used for mini-lectures, homework help, and/or short handouts with questions or readings for the students. Students were encouraged to work in groups, but this was not structured into the class so several students would work alone. The average score on the MSCE.2 was 47% without tutorials and 56% with tutorials. So, students scored 9%, or 0.53 standard deviations, better on the MSCE.2 in the quarter the tutorials were used ($t(360) = 24.10$, $p < 0.001$).

In addition, Figure 7.4 shows increased student score with increased recitation attendance. The MSCE.1 and the MSCE.2 were both given in quarters with nongraded recitation attendance and tutorial use. As discussed previously, this lead to variable attendance and thus variable exposure to tutorials. Different versions of the MSCE were used in the two quarters. This makes across quarter comparison problematic; however, within quarter comparisons are not compromised. Figure 7.4 shows there is a linearly increasing relationship between recitation attendance, thus tutorial use, and MSCE score. For each recitation attended, there was on average a 0.12 standard deviation increase in score on the MSCE.2 and a 0.18 standard deviation increase in score on the MSCE.1 (This is very similar to the data seen in Chapter 4, but final exam score was reported there. Also, unlike in Chapter 4, there is no way to control for the possibility that better students go to recitation more often.) See Figure 7.4 for a reporting of the exact percentages.

Another interesting result, which is shown in Figure 7.4, is an increase in students' scores when all students are "required" to attend recitation. MSCE.3 was given in a quarter where recitation attendance was graded which resulted in 93% attendance. Because nearly all students attended recitation, it does not make sense to bin students by number of recitations attended. However, 93% attendance implies that most students came to at least eight of the recitations. The average score of 67% on the MSCE.3 for this quarter places the data in a near perfect fit on the graph. In addition, the two quarters where no tutorials were used are also a near perfect fit (47% score on the MSCE.2 and 44% score on the MSCE.0). This may be because the different versions of the MSCE are not significantly different, or it may be a statistical fluke of several factors coming together at once. The two data points are included for interest. (It should be noted that, when student responses to the subset of questions that are the same across all four versions of the test are analyzed in this same way, there is very little actual changes to the reported data in Figure 7.4. This suggests that indeed the different versions, while different, are approximately of equal difficulty for the students.)

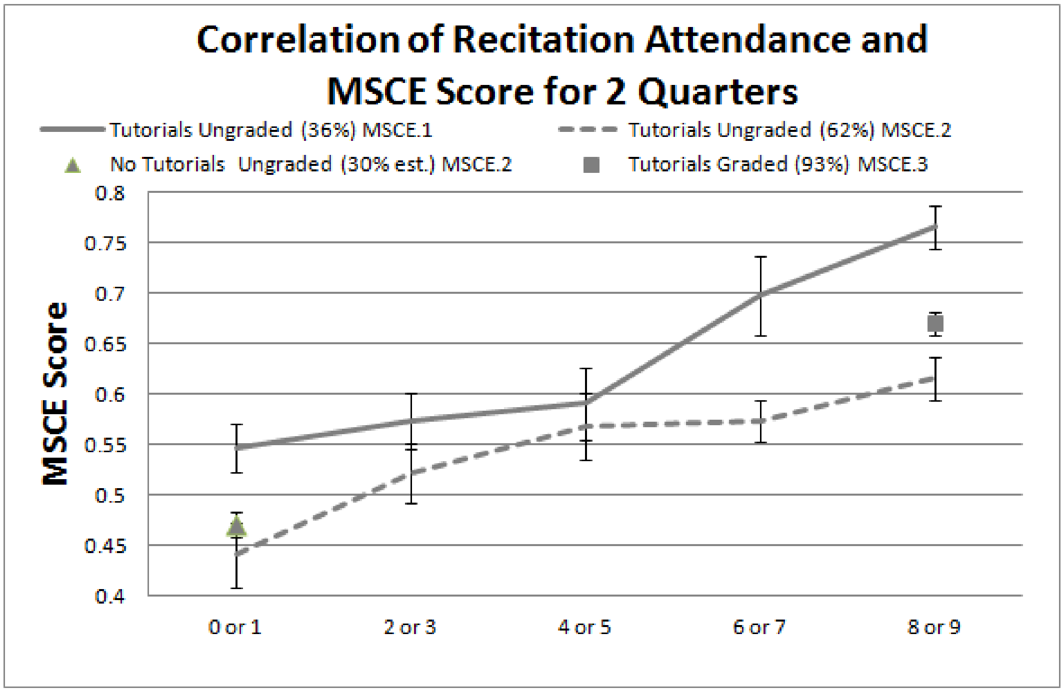


Figure 7.4: Gains in MSCE score with increased tutorial exposure. For each recitation attended, there was a 0.12 standard deviation increase in score on the MSCE.2, dotted line, and a 0.18 standard deviation increase in score on the MSCE.1, solid line. Since the test is changing for the different data sets, a strictly rigorous science view would say attend only to the slopes of the two lines. For reader interest, two data points are included that compare scores in a quarter without tutorials and a quarter with graded tutorials, but careful inferences should be drawn from these two data points. (See the text for a larger discussion.)

7.5 MSCE Score and Reported Student Testing Effort

MSCE tests were all given during flex homework sessions. With the exception of the session in which the MSCE.3 test was given, students were told that their score on the MSCE test would not affect their grade in the course. In addition, students were not encouraged, and thus did not, study or review old material prior to the assessment. They were encouraged to try their hardest and treat the test as a practice final in which they could assess their preparedness for the upcoming final, and they were told that they would be allowed to compare answers and receive feedback directly after the test. In principle, this resulted in almost all of the students completing the test and giving a good faith effort. However, there was concern that students might not try hard if it did not effect their grade. To control for this, a posttest exit survey question asked students to rank from 1 to 5 their effort. 1 was, ‘I guessed randomly on most of the questions,’ and 5 was, ‘I tried my hardest and took an

Test	Tutorials	Recitation attend.	Percentage of students ranking themselves in each leichert number of reported effort. 1 is No effort .... 5 is Tried their Best.				
			1	2	3	4	5
MSCE.2	NOT Used	Voluntary	5%	12%	38%	38%	9%
MSCE.1	Used	Voluntary	2%	9%	36%	43%	10%
MSCE.2	Used	Voluntary	2%	7%	25%	34%	32%
MSCE.3	Used	Graded	1%	4%	28%	42%	27%

Table 7.3: Self reported effort on the MSCE test shows an effect of quarter. ($\chi^2(11) = 92.945, p < 0.001$).

educated guess when unsure of the answer'. There are three main results of this data. 1) different quarters show differing amounts of effort on the flex test (yet another reason for caution in between quarter comparisons). 2) Students who tried harder did do better on the MSCE tests. 3) When factoring out effort on the test, there is still a main effect of recitation attendance in improved MSCE score.

[The MSCE.3 test was given differently because recitations were being graded. Recitations were graded on a 75% participation and 25% correctness standard. The group worksheets were collected, graded, and returned to students each week. Students were told that their correctness on the test would factor into the 25% of each recitation grade which was usually reserved for correctness. However, because they were working alone and usually they worked in groups and with open notes, they were told that grading would be much more leniently than usual.]

Different quarters show differing amounts of effort on the MSCE test version. Both inspection of Table 7.3 and a chi squared test for independence reviles this, ($\chi^2(11) = 92.945, p < 0.001$). It should be noted that even separating out the quarters that are unusual and looking only at the two voluntary-attendance-and-tutorials-used quarters still shows a significant chi squared test for independence ($\chi^2(3) = 19.830, p = 0.001$). This suggests that while student effort may be tied to the perceived importance of the flex sessions, by either the similarity of the flex sessions to the regular course recitation sessions or by the presence of grading, there are other factors involved.

Because there are differences in effort between quarters, the obvious question is, "Do students do better when they try harder?" The answer is generally yes. In all four quarters, student who reported a higher effort scored better. Interestingly however, the few students who reported 1, i.e. Did not try at all, usually scored better than students who reported 2 and 3 on effort and students who reported 5, i.e. Tried their Best, scored worse in half the quarters than those who indicated a 4 for effort. For this reason, the data was binned and reported in Figure 7.5 as high effort, 5 or 4, and low effort, 3 or less. Students who

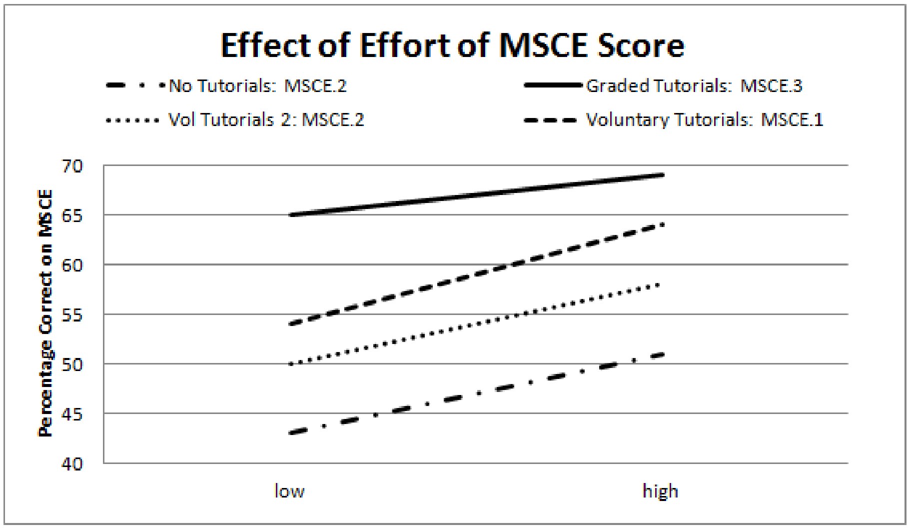


Figure 7.5: Effect of effort on MSCE score. High effort is a reported 5 or 4 and low effort is reported 3 or less, on a likert 1 to 5 scale. Students scored an average of 7.5%, or 0.44 standard deviations better, for high vs. low effort.

reported high effort scored an average of 7.5%, or 0.44 standard deviations, better on the MSCE test version they took.

At this point an alert and skeptical reader may be thinking that the correlation seen between recitation attendance and better MSCE score was caused by, or highly influenced by, effort. This is logical: students who go to recitation more often may be more likely to care about the class and thus may be more likely to try harder on nongraded activities. But, analysis of the data shows that there is a main effect of both recitation attendance and effort on the MSCE test for both quarters. In addition, there is only interaction between attendance and effort in one of the two quarters, which makes this line of logical reasoning about attendance and effort questionable. This data is recreated in Figure 7.6.

The quarter MSCE.1 was given the data shows an interaction as well as a main effect. The main effect of attendance in recitation - thus tutorial use: (F(4, 92) = 5.120, $p = 0.001$). The main effect of effort: (F(1, 92) = 6.568, $p = 0.012$). The interaction effect: (F(4, 92) = 2.665, p = 0.035). The quarter MSCE.2 was given the data shows only two main effects and no significant interaction. The main effect of attendance in recitation - thus tutorial use: (F(4, 201) = 6.184, $p < 0.001$). The main effect of effort: (F(1, 201) = 8.480, $p = 0.004$). No significant interaction effect: (F(4, 201) = 0.158, $p = 0.959$). Thus, tutorial attendance has a possitive effect on MSCE score which is independant of reported effort on the test. And, reported effort on the test has a positive effect on MSCE score which is independant of tutorial use. There may be an interaction of the two on MSCE score but the signal is unclear.

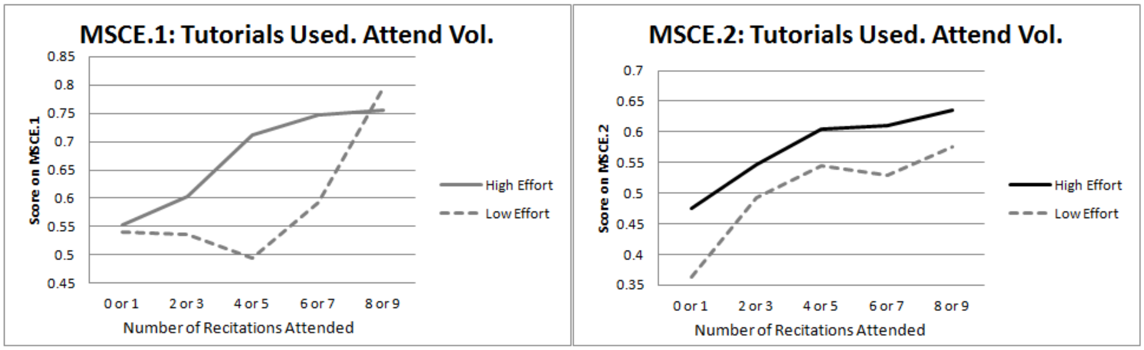


Figure 7.6: Interplay of tutorial use and effort on MSCE score. MSCE.2 shows no interaction but a main effect of both tutorials and effort. MSCE.1 shows a main effect of both and an interaction.

It is also interesting to note that when tutorial attendance and reported effort are considered independantly to MSCE score. In neither of the two quarters shown in Figure 7.6 is there a significant difference in reported effort on the MSCE test for students who attend recitations more often (F4, 97) = 0.914, $p = 0.459$ and F(4, 206) = 1.031, $p = 0.392$). This suggests that students who attend recitation more often either do not actually expend more effort in the course or simply do not report that they expend more effort than a student who does not go to recitations as often.

7.6 Teaching Assistant Score on the MSCE Tests

The last comparison presented here on construct validity is teaching assistant responses. This data is shown in Figure 7.7. The main results from this data are that in all cases the teaching assistants scored at least 15% better on a pretest than the students were able to on a posttest, and in all quarters considered, the TAs did much better on a posttest than on the pretest, between 20% to 25% better. Thus, this data is consistent with the expectations of a valid test.

7.7 Psychometric Properties of MSCE.3: Item Analysis.

Reported here are response percentages on the newest and, we believe, best version of the MSCE test, MSCE.3. This test is shown in full in Appendix B and has several questions in common with the data reported in Chapter 5 on student difficulties. The evaluation was given to students in flex homework sessions, discussed in the data collection methods section in Chapter 4. Unlike the data reported in the student difficulties section, the data shown here was collected in a quarter where almost all, 93% on average, students attended recitation with the tutorials. The test was given the day after the second midterm which

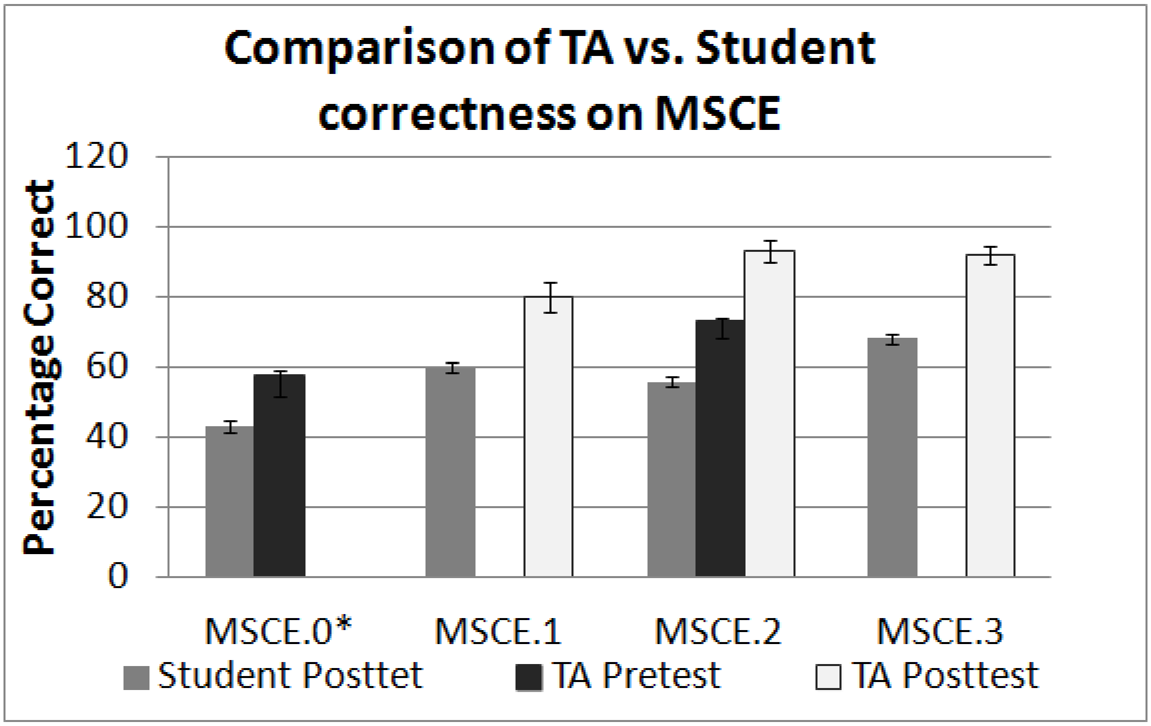


Figure 7.7: Teaching Assistant MSCE test score.
∗ A subset of 35 of the MSCE.0 items was given to the TAs. This set of 35 items was edited further to the MSCE.1. The MSCE test the TAs took as a pretest was the previous quarter's posttest, and they took was their quarters posttest. Gains were calculated by comparison of match questions. Although, similar gains were seen comparing across MSCE versions too. TA gains: 25% ± 9% (MSCE.0∗ to MSCE.1) and 20% ± 12% (MSCE.2 to MSCE.3).

may have contributed to the higher than normal scores on the test.

7.8 Summary

Here I reported on the creation and validation of a Materials Science Conceptual Evaluation (MSCE). This assessment was created from a set of test items designed to identify areas of difficulties students had with the introductory materials science course content. Thus, a majority of items were separately validated before inclusion in the MSCE test version.

The tests items were created by the following process: 1. identify a difficulty through interviews, literature searches, and old test questions; 2. create a multiple-choice item from short-answer and interview responses; 3. test students to make sure the multiple-choice item is a difficulty and is clear to the students.

Then, the content validity of the test items was established by three course instructors,

Table 7.4: Posttest response percentages on the 31 item MSCE given for Autumn Quarter 2011 (N = 199). All students received a small amount of participation credit for attending recitation. The average attendance was 93%. Correct answers are bold.

Item	Pt-Bis.	a	b	c	d	e	Item	Pt-Bis.	a	b	c	d	e
#1	0.288	5	14	8	**72**	...	#16	0.318	16	10	**74**	...	...
#2	0.3	23	5	0	**72**	...	#17	0.342	8	3	6	**84**	0
#3	0.315	4	**73**	10	13	...	#18	0.454	22	3	**75**	...	...
#4	0.267	7	6	**87**	...	...	#19	0.369	4	8	**73**	16	...
#5	0.374	**47**	23	8	22	...	#20	0.391	27	**52**	8	12	1
#6	0.47	**67**	4	26	3	...	#21	0.392	28	14	0	**58**	...
#7	0.404	11	4	**65**	12	8	#22	0.369	5	4	12	**79**	1
#8	0.238	26	**66**	5	1	3	#23	0.371	3	18	3	2	**75**
#9	0.443	31	10	8	**39**	13	#24	0.399	**66**	10	6	8	10
#10	0.41	4	19	10	**67**	...	#25	0.337	0	9	8	**83**	0
#11	0.359	**53**	7	6	34	...	#26	0.431	2	2	**75**	15	7
#12	0.464	3	**74**	22	...	...	#27	0.261	2	0	**93**	3	3
#13	0.486	**77**	1	6	8	8	#28	0.376	11	30	**57**	3	...
#14	0.485	**82**	17	1	...	...	#29	0.341	7	12	18	**63**	...
#15	0.297	2	**94**	5	1	0	#30	0.346	5	**46**	41	8	...
							#31	0.41	18	9	20	5	**47**

$Avg.Score = 67.7\% \pm 16.5\%$

KR 20 = 0.721

Correlation of MSCE score w/ final grade = 0.421

and materials scientists, assessing items based on importance and correctness of the answer options and question. This data was analyzed and items with low importance were not included on the MSCE. In addition, items were reworded as needed to make them technically sound and clearer based on instructor comments and/or concerns.

After that, the construct validity was established through a series of think-aloud interviews and explain your answers short responses while the student was taking the MSCE test.

Finally, a series of statistical validation measures were taken to ensure the items were valid and the test was internally reliable. There are several statistical measures reported here to ensure the validity of the MSCE. Correlations with final exam score, point biserial coefficients, reliability of the test as a whole (KR 20), gains with increased tutorial participation, gains with increased student effort on the test, and lastly better performance by more expert students, i.e. the TAs.

Each of these validity measures could have provided evidence of an invalid test were they to not show the expected results, but on none of these measures was there any indication of an invalid test. Thus, we conclude that to the best of our knowledge the MSCE is a valid and appropriate measure of student conceptual understanding of the introductory materials science course.

Chapter 8
SUMMARY AND CONCLUSIONS

Here I present each of my goals stated in the introduction and a summary of the work addressing each goal.

8.1 Goal 1

8.1.1 Statement of Goal

Goal 1 was to develop a multiple-choice instrument to assess student understanding of the directional relationships between force, velocity, and acceleration in one dimension. This instrument answers questions like: "Does a student who understands that velocity and acceleration are not directionally related also understand that force and velocity are not directionally related?"

8.1.2 Summary of Reported Data Addressing that Goal

In this thesis, I presented the process by which we developed an instrument to systematically investigate student conceptual understanding of the relationships between the directions of net force, velocity, and acceleration in one dimension. Unlike previous work, this instrument can simultaneously study all six possible conditional relations between force, velocity, and acceleration in order to obtain a coherent picture of student understanding of the relations between all three concepts. A variety of evidence demonstrating the reliability of the instrument and the validity for all three traditionally separated validity types - content, construct, and criterion is presented.

The content validity of the FVA test was established by the following process. We showed a group of three other PER researchers, and a few graduate students, several questions on the FVA test. We described the goals of the test, showed them some student response data, and explained the inferences we were drawing from the data. These researchers were asked to give feedback on the question wording and projects goals. From this, we received only one concern about content validity. This was that we needed to ensure that the wording was clearly referring to an instant in time for each item. This was put into action and all FVA test data reported in Part I has this edit. In this way the content validity of the FVA test was established.

The construct validity of the FVA test was established by a series of interviews and student testing. First, a series of student think-aloud interviews, as the students completed the test questions, established that students were responding in a manner consistent with their thought process while taking the test. Second, student scores on three different quizzes - 'Always, Sometimes, Never', 'Possible, Not Possible', and Not Enough Information - were compared to the fva scores in order to establish that the answer choice format used on the fva tests was the most reliable form available. Third, five additional quizzes were created in order to establish that the question type and not the question context was driving the student responses. Three of these quizzes were fixed question types and multiple context quizzes, which showed that student scores on the fva items were in the middle of the range of context related variations.Two of these quizzes were changing question types which held the fva context fixed so that a $\vec{v} \rightarrow \vec{f}$ question about a boat on a lake could be compared to a $\vec{f} \rightarrow \vec{v}$ or $\vec{a} \rightarrow \vec{v}$ question about a boat on a lake. These two quizzes also supported the conclusions made about how question type affects student response. Lastly, a series of test data and statistical measures showed that students in increasingly higher course levels and pre/post instruction within a course had the expected gains in fva test scores.

The criterion validity of the FVA test was established through correlations with final exam score and FCI score. For each of these measures, the expected gains were seen.

8.2 Goal 2

8.2.1 Statement of Goal

Goal 2 was to use this instrument to gain a better understanding of student knowledge and how that knowledge changes with student level and instruction.

8.2.2 Summary of Reported Data Addressing that Goal

An analysis of over 650 student responses from three different course levels revealed three main findings.

First, we consistently found evidence of an intermediate, partially correct level of understanding of the relations between force, velocity, and acceleration held by up to 30% of the students pre or post instruction. This is in addition to finding that a significant number of students answer consistently with the well-known and common student "misconception" that the vector quantities should always point in the same direction. Specifically we found two intermediate models. The first model is the belief that two vectors, such as force and velocity, need not be aligned, but they may also be pointed in opposite directions, but one cannot be zero ("Cannot-be-Zero" model). The second model is the belief that the two vectors need not be aligned, though one of them could be zero but not pointed in the opposite direction as the other ("Cannot-be-Opposite" model). Furthermore, we found that about half of the students who improved their understanding of the relations between the directions of force, velocity, and acceleration did so by evolving through these partially correct "states". Roughly speaking, from pre to post test in the honors physics sections, we found that about half of the students did not change their responses, about one-quarter changed from the misconception answered to the correct answer, and about one-quarter either changed from misconception answer to the partially correct answer or from the partially correct answer to the correct answer.

Second, we found an asymmetry in student responses to conditional relations. That is, students often treated questions that probe the concept *motion implies acceleration* differently than the concept *acceleration implies motion.* Likewise they often treated questions about *motion implies force* differently than *force implies motion*, and perhaps surprisingly, they often treated questions about *Force implies acceleration* differently than *acceleration implies force.* The differences are reflected in the response frequencies of each answer choice. For example for $\vec{v} \rightarrow \vec{a}$ versus $\vec{a} \rightarrow \vec{v}$ questions, there were differences in the number of correct and misconception responses as well as in the kinds of partially correct responses (Cannot-be-Opposite versus Cannot-be-Zero). For the $\vec{v} \rightarrow \vec{F}$ versus $\vec{F} \rightarrow \vec{v}$ questions, there were no differences in the correct and misconception response frequencies, but there were differences in the partially correct Cannot-be-Opposite versus Cannot-be-Zero responses.

Third, we found evidence of specific hierarchies in correct responses to different question types. The evidence included both within-student scores at one point in time and within-student gains in scores from pre to post. For example, we found evidence that if students correctly answered $\vec{F} \rightarrow \vec{v}$ questions, then they were very likely to also correctly answer $\vec{v} \rightarrow \vec{a}$ questions, but not the converse. Further, if $\vec{v} \rightarrow \vec{a}$ questions were answered incorrectly, then it was very likely that $\vec{F} \rightarrow \vec{v}$ questions were also answered incorrectly (but not the converse). Likewise we found that for a given student, gains in $\vec{F} \rightarrow \vec{v}$ scores most likely occurred in the presence of gains in $\vec{v} \rightarrow \vec{a}$ scores, but the not converse. These findings are indeed interesting and suggest that it may be required to understand the relationship between the direction of velocity and acceleration in order to understand the relationship between the direction of force and velocity. However, to more firmly establish a possible causal link between understanding these relations, one must first be careful to explicitly define what is meant by *understand*, and, second, a controlled intervention (for example manipulating the amount of velocity-acceleration instruction) is needed.

8.3 Goal 3

8.3.1 Statement of Goal

Goal 3 was to develop an effective set of group-work conceptual worksheets for use in recitations in the introductory materials science course.

8.3.2 Summary of Reported Data Addressing that Goal

I reported in Chapter 4 on the process and results of applying a the tutorial design process, which has proven to be successful for a number of physics topics, to the design of similar tutorials for a university-level introductory materials science and engineering course. The process involved the identification of instructional goals, the identification of specific student difficulties, the iterative design of the tutorials, their implementation in recitations, and follow up assessment of their effectiveness as teaching tools.

The project, which involved over 1000 students, included extensive interviewing, testing, and iterative classroom implementation over a period of three years. It yielded 9 field-tested 48 minute tutorials in which students work together in small groups on the tutorials in the presence of teaching assistants who assess and facilitate student progress.

These tutorials were all assessed based on final exam scores, and it was shown that even accounting for the fact that slightly “better” students tended to attend recitations more often, there was a significant valued-added effect of the recitations on final exam performance. Specifically, students gained on average 0.1 standard deviations on their final exam for each of the tutorials used. These results suggest that these recitation methods and materials are effective in teaching students the difficult and important conceptual materials which they were designed to address. Furthermore, since this process was initially designed for physics courses yet is also successful for an engineering course, this implies that this process may be successful for a wide range of STEM courses.

8.4 Goal 4

8.4.1 Statement of Goal

Goal 4 was to develop a multiple-choice assessment instrument to test the effectiveness of these conceptual worksheets, or ‘tutorials’, and which can generally be used in coordination with the materials concept inventory (MCI) to assess student understanding in materials science.

8.4.2 Summary of Reported Data Addressing that Goal

The content validity of the MSCE test was established by the following process. Three course instructors, and materials scientists, assessed four aspects of the items on the MSCE test: the importance of each items inclusion in the MSCE test (reported as a number 1 to 10 on a likert scale); the correct answer for each item; the level of concern for each items reported student response data (reported as a number 1 to 10 on a Likert scale); and finally, any comments they had about question wording or content. This data was analyzed and items with low importance were dropped from the MSCE. In addition, items were reworded as needed to make them technically sound and clearer based on instructor comments and concerns.

The construct validity of the test was completed through three main ways. First, a series of student think-aloud interviews as the students completed the test questions. In addition, because of the length of the test and the time consuming nature of think-aloud interviews, a set of "choose-an-answer and explain-your-choice" questions were given to several students. Second, it was given to teaching assistants for the course to assess how "experts" performed. And, finally, it was given to students undergoing different amounts of tutorial exposure. (Tutorials had already been shown to be effective teaching tools.) These responses were analyzed to make sure that gains in score were consistant with expected results. (Note: such validation assessments were done before, during, and after item creation. The before and during interview and testing were used to update and edit the items on the test, and the post were used as a validation measure.)

The criterion validity of the test was completed by evaluating correlations with final exam scores. In no case was a comparison made which suggested that the test was invalid. This would have been the case if we expected to see a statistically significant positive or negative correlation and did not see one. Thus, the validity of the MSCE was establish.

8.5 The Main Points

In this thesis I reported on two main areas of research and design. These two areas are "Identifying Student Difficulties and Misconceptions with a Specific Physics Concept", namely the directional relationships between net force, velocity, and acceleration in one dimension, and "Identifying and Addressing Student Difficulties with Materials Science Concepts." There are two main points that should be taken away from this thesis.

The first: even in a very well known and highly studied area, such as student confusion of net force, velocity, and acceleration, a carefully constructed, specific test instrument is able to provide previously unknown and highly interesting information on student response patterns which allows for an analysis of hierarchies in how students think about, reason with, and learn the topic.

The second: the tutorial design process known to be very effective in physics was able to be applied effectively to a materials science course. The nonspecific way in which the process was applied suggests that this could be done for several areas of science education with similar positive results.

8.6 Future Research

There are several areas of additional research that this work would suggest.

1. Conduct a larger test of the instructional implications suggested by the FVA results.

Here we presented results from a very small training that had limited effect. A larger course wide training could be very informative. For example, we showed that students learn the $\vec{v} \to \vec{a}$ relationship before the $\vec{v} \to \vec{F}$. If the traditional order of instruction was altered so that students learned force first or so that the traditional kinematics formulas were presented with force and not acceleration, this pattern might very well change. Such research would begin to answer questions about the *cause* for the pattern that was seen and carefully described here.

2. Develop a study of student difficulties with the directional relationships between net force, velocity, and acceleration in two or three dimensions.

Beginning exploratory data in this has been collected. It is unclear yet if there is a consistent and large number of students choosing relationships which are not along the direction of the vector. If only a small percentage of students choose these on axis answers than while interesting this would probably not be a large project.

3. Analyze student difficulties with vector magnitudes and object size.

For example, do students have more difficulty with the net force and velocity relationship when the object is very heavy or when the velocity is very large?

4. Test the tutorials design method in other STEM courses.

This materials science research suggests that the process of tutorial design could be applicable to many classes, but this has not been tested. In addition, research and classification of the types of classes it works best in, or does not work in, would be interesting.

5. Develop tutorials for a semester long materials science course.

Application of this process to other topics that a semester long class would address or to higher level courses is the next step in materials science course design.

6. Examine student difficulties with log plots and online instructional modules to teach this topic better.

7. Study student difficulties reasoning with multiple variables.

There were several areas in the tutorials where student difficulty reasoning with multiple variables were seen. This difficulty was sometimes not realizing a variable was, or was not, affecting a system, but there were also difficulties with reasoning when all the variables were identified. These difficulties can be seen in physics too, and I believe it is generally an issue for STEM education. Research into the nature of this problem and better instructional techniques, especially ones that could lead to transfer to new areas, would be highly useful to stem educators.

This is a short list of very briefly described ideas. However, it is clear that there are many interesting areas of research that could be persued in the future which are related to what has already been done in this thesis.

BIBLIOGRAPHY

Alonzo, A. C. and Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education*, **93**, 389–421.

Ambrose, B. S. (2004). Investigating student understanding in intermediate mechanics: Identifying the need for a tutorial approach to instruction. *American Journal of Physics*, **72**, 453–459.

Bao, L. and Redish, E. F. (2002). Understanding probabilistic interpretations of physical systems: A prerequisite to learning quantum physics. *American Journal of Physics*, **70**, 210–217.

Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics*, **62**, 750–762.

Beichner, R. J. (2009). An Introduction to Physics Education Research. In Getting Started in PER, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2009), Reviews in PER Vol. 2, ¡http://www.per-central.org/items/detail.cfm?ID=8806¿.

Callister, W. D. (2007). Materials Science and Engineering: an Introduction. New York: Wiley.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., and Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, **13**, 145–182.

Chi, M. T. H., De Leeuw, N., Chiu, M., and LaVancher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science*, **18**, 439–477.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences*, **14**(2), 161–199.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, **50**, 66–71.

Demetry, C. (2006). Use of Formative Assessment to Probe Student Conceptions of the Lever Rule. Proceedings of the American Society of Engineering Education Annual Conference & Exposition.

Dufresne, R. J., Gerace, W. J., and Leonard, W. J. (1997). Solving physics problems with multiple representations, *Physics Teacher*, **35**, 270-275.

Ding, L., Chabay, R.,, Sherwood, B., and Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics - Physics Education Research*, **2**, 010105.

Enderstein, L. G. and Spargo, P. E. (1996). Beliefs regarding force and motion: a longitudinal and cross-cultural study of South African school pupils. *International Journal of Science Education*, **18**, 479–492.

Enderstein, L. G. and Spargo, P. E. (1998). The effect of context, culture and learning on the selection of alternative options in similar situations by South African pupils. *International Journal of Science Education*, **20**, 711–736.

Engelhardt, P. V. (2009). An Introduction to Classical Test Theory as Applied to Conceptual Multiple-choice Tests. In Getting Started in PER, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2009), Reviews in PER Vol. 2, ¡http://www.per-central.org/items/detail.cfm?ID=8807¿.

Felder, R. (1995). A longitudinal study of engineering student performance and retention, IV. Instructional methods and student responses to them. *Journal of Engineering Education*, **84**(4), 361–367.

Finkelstein, N. D. and Pollock, S. J. (2005). Replicating and understanding successful innovations: Implementing tutorials in introductory physics. *Physics Review Special Topics - Physics Education Research*, **1**, 010101.

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.

Halloun, I. A. and Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics* **53**, 1043–1055.

Heckler, A. F. and Rosenblatt, R. (2010). Student understanding of atomic bonds and their relation to mechanical properties of metals in an introductory materials science engineering course. Proceedings of the American Society of Engineering Education Annual Conference & Exposition.

Heckler, A. F. and Rosenblatt, R. (2011). Student difficulties with basic concepts in Introductory Materials Science. Proceedings of the Annual ASEE/IEEE Frontiers in Education Conference.

Paula R.L. Heron, Peter S. Shaffer, and Lillian C. McDermott. Identifying and Addressing Student Conceptual Difficulties: An Example from Introductory Physics.

Heron, P. R. L., Loverude, M. E., Shaffer, P. S., and McDermott, L. C. (2003). Helping students develop an understanding of Archimedes' principle. II. Development of research-based instructional materials. *American Journal of Physics*, **71**, 1188–1195.

Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, **55**, 440-454.

Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, **30**, 141–166.

Kitto, K.L. (2007). Analyzing What Students Write about Materials - Another Strategy for Developing Conceptual Knowledge in a Materials Engineering Course. Proceedings of the Annual ASEE/IEEE Frontiers in Education Conference.

Kitto, K. L. (2008). Developing and Assessing Conceptual Understanding in Materials Engineering Using Written Research Papers and Oral Poster Presentations. Proceedings of the Annual ASEE/IEEE Frontiers in Education Conference.

Krause, S., Decker, J. C., Niska, J., and Alford, T. (2002). A Materials Concept Inventory for introductory materials engineering courses, *National Educators Workshop Update 2002*, **17**, 1–8.

Krause, S. J., Decker, J. C., and Griffin, R. (2003). Using a Materials Concept Inventory to assess conceptual gain in introductory materials science courses. Proceedings of the Annual ASEE/IEEE Frontiers in Education Conference.

Krause, S. J., Tasooji, A., and Griffin, R. (2004). Origins of Misconceptions in a Materials Concept Inventory From Student Focus Groups. Proceedings of the American Society of Engineering Education Annual Conference & Exposition.

Krause, S., Kelly, J., Tasooji, A., Corkins, J., Baker, D., and Purzer, S. (2010a). Effect of Pedagogy on Conceptual Change in an Introductory Materials Science Course. *International Journal of Engineering Education*, **20**, 868–879.

Krause, S., Kelly, J., Triplett, J., Eller, A. and Baker, D. (2010b) Uncovering and addressing some common types of misconceptions in introductory materials science and engineering courses. *Journal of Materials Education*, **32**, 255–272.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, **15**, 11–24.

Larkin, J., McDermott, J., Simon, D. P., and Simon, H. A. (1980). Expert and Novice Performance in Solving Physics Problems. *Science*, **208**(4450), 1335–1342.

Lindell, R., Peak, E., and Foster, T. M. (2007). Are they all created equal? A comparison of different concept inventory development methodologies. 2006 Physics Education Research Conference **883**, 14 17.

Loverude, M. E., Heron, P. R. L., and Kautz, C. H. (2010). Identifying and addressing student difficulties with hydrostatic pressure. *American Journal of Physics*, **78**, 75–85.

McDermott, L. C., Rosenquist, M. L., and van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics*, **55**, 503–513.

McDermott, L. C. (1991). Millikan Lecture 1990: What we teach and what is learned - Closing the gap. *American Journal of Physics*, **59**, 301–315.

McDermott, L.C. (2001). Physics Education Research-The Key to Student Learning. *American Journal of Physics*, **69**, 1127–1137.

McDermott, L. C., Shaffer, P. S., and the Physics Education Group at the U. of Washington. (2002). Tutorials in Introductory Physics, Upper Saddle River, NJ: Prentice Hall.

McDermott, L. C., Shaffer, P. S., and Somers, M. D. (2004). Research as a guide for teaching introductory mechanics: An illustration in the context of the Atwood's machine. *American Journal of Physics*, **62**, 46–55.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, **50**, 741–749.

National Research Council (2007). Taking science to school. Washington, DC: The National Academies Press.

Palmer, D. (1997). The effect of context on students' reasoning about forces. *International Journal of Science Education*, **19**, 681–696.

Pollock, S. J. and Finkelstein, N. D. (2008). Sustaining education reforms in introductory physics. *Physics Review Special Topics - Physics Education Research*, **4**, 010110.

Posner, G. J., Strike, K. A., Hewson, P. W., and Gertzog, W. A. (1982). Accommodation of a Scientic Conception: Toward a Theory of Conceptual Change. *Science Education*, **66**, 211–227.

Reif, F. and Allen, S. (1992). Cognition for Interpreting Scientific Concepts: A study of acceleration. *Cognition and Instruction*, **9**, 1–44.

Rosenblatt, R. and Heckler, A. F. (2010). Student understanding of the mechanical properties of metals in an introductory materials science engineering course. Proceedings of the Annual Conference of the American Society of Engineering Education.

Rosenblatt, R., Heckler, A. F., and Flores, K. (2011). Tutorials for an Introductory Materials Engineering Course. Proceedings of the Annual ASEE/IEEE Frontiers in Education Conference.

Rosenblatt, R., Sayre, E. C., and Heckler, A. F. (2008). Toward a Comprehensive Picture of Student Understanding of Force, Velocity, and Acceleration. In Proceedings of the 2008 PER Conference, AIP, Melville, NY, 183–186.

Rosenblatt, R., Sayre, E. C., and Heckler, A. F. (2009). Modeling students' conceptual understanding of force, velocity, and acceleration. In Proceedings of the 2009 PER Conference, AIP, Melville, NY, 245–248.

Rebello, S. N. and Zollman, D. A. (2004). The effect of distracters on student performance on the force concept inventory. *American Journal of Physics*, **72**, 116–125.

Scherr, R. E., Shaffer, P. S., and Vokos, S. (2002). The challenge of changing deeply held student beliefs about the relativity of simultaneity. *American Journal of Physics*, **70**, 1238–1248.

Shaffer, P. S. and McDermott, L. C. (1992). Research as guide for curriculum development: an example from electricity, Part II: Design of instructional strategies. *American Journal of Physics*, **60**, 1003–1013.

Shaffer, P. S. and McDermott, L. C. (2005). A research-based approach to improving student understanding of the vector nature of kinematical concepts. American Journal of Physics. 73, 921-931. Singh, C. (2008). Interactive learning tutorials on quantum mechanics. *American Journal of Physics*, **76**, 400–405.

Singh, C. (2008). Interactive learning tutorials on quantum mechanics. *American Journal of Physics.* **76**, 400-405.

Streveler, R. A., Litzinger, T. A., Miller, R. L., and Steif, P. S. (2008). Learning Conceptual Knowledge in the Engineering Sciences: Overview and Future Research Directions. *Journal of Engineering Education*, **97**(3), 279–294.

Taber, K. S. (2006). Beyond Constructivism: the Progressive Research Programme into Learning Science. *Studies in Science Education*, **42**, 125–184.

Tamir, P. (1990). Justifying the selection of answers in multiple choice items. *International Journal of Science Education*, **12**, 563–573.

Thornton, R. K. and 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.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, **10**, 159–169.

Trowbridge, D. E. and McDermott, L. C. (1981). Investigation of student understanding of the concept of acceleration in one dimension. *American Journal of Physics*, **49**, 242–253.

Tversky, A. and Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In Progress in Social Psychology, edited by M. Fishbein, (Erlbaum, Hillsdale, NJ, 1980), 49–72.

Van Heuvelen, A. and Zou, X. (2001). Multiple representations of workenergy processes. *American Journal of Physics*, **69**, 184–194.

Van Heuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. *American Journal of Physics*, **59**, 891–897.

Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education*, **1**, 205–221.

Wang, J. and Bao, L. (2010). Analyzing Force Concept Inventory with Item Response Theory. *American Journal of Physics*, **78**, 1064-1070.

Wittmann, M. C., Redish, E. F., and Steinberg, R. N. (2003). Understanding and Addressing Student Reasoning about Sound. *International Journal of Science Education*, **25**(8), 991-1013.

Wittmann, M. C., Steinberg, R. N., and Redish, E. F. (2005). Activity-Based Tutorials. New York: Wiley.

Wittmann, M. C., Morgan, J. T., and Bao, L. (2005). Addressing student models of energy loss in quantum tunneling. *European Journal of Physics*, **26**, 939–950.

Wosilait, K., Heron, P. R. L., Shaffer, P. S., and McDermott, L. C. (1998). Development and assessment of a research-based tutorial on light and shadow. *American Journal of Physics*, **66**, 906–913.

Wosilait, K., Heron, P. R. L., Shaffer, P. S., and McDermott, L. C. (1998). Addressing student difficulties in applying a wave model to the interference and diffraction of light. *American Journal of Physics*, **67**, S5–S15.

Appendix A
The FVA Test

1. At exactly 2:31PM, a boat is moving to the north on a lake. Which statement best describes the forces on the boat at this time?
   (a) there may be several forces to the north and to the south acting on the boat, but the forces to the north are larger.
   (b) there may be several forces to the north and to the south acting on the boat, but the forces to the south are larger.
   (c) there may be several forces to the north and to the south acting on the boat, but the forces to the south are equal in magnitude to those to the north.
   (d) both a and b are possible.
   (e) both a and c are possible.
   (f) a, b, and c are possible.

2. A car is moving to the right and speeding up. Which statement best describes the acceleration of the car at this instant?
   (a) the car's acceleration is to the right.
   (b) the car's acceleration is to the left.
   (c) the car's acceleration is zero.
   (d) both a and b are possible.
   (e) both a and c are possible.
   (f) a, b, and c are possible.

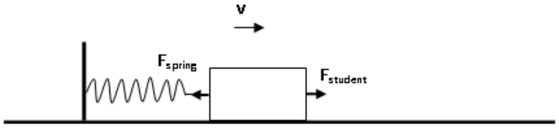
3. A student and a dog are playing tug of war with a rubber toy. If at a particular time the student is pulling on the toy to the right and the dog is pulling to the left with an equal force, which statement best describes the motion of the toy at this time?
   (a) it is moving toward the dog.
   (b) it is moving toward the student.
   (c) it is not moving.
   (d) both a and b are possible.
   (e) a, b, and c are possible.

4. A car is on a hill and the direction of its acceleration is uphill. Which statement best describes the motion of the car at that time?
   (a) it is moving uphill.
   (b) it is to moving downhill.
   (c) it is not moving.
   (d) both a and b are possible.
   (e) both a and c are possible.
   (f) a, b, and c are possible.

5. A group of workers is pushing on a car in a driveway. There may be several forces on the car but those toward the street are greater. Which statement best describes the acceleration of the car at this instant?
   (a) its acceleration is toward the street.
   (b) its acceleration is away from the street.
   (c) its acceleration is zero.
   (d) both a and b are possible.
   (e) both a and c are possible.
   (f) a, b and c are possible.

6. A wagon is rolling east along the sidewalk. What can you say about the acceleration of the wagon at this time?
   - (a) it is accelerating east.
   - (b) it is accelerating west.
   - (c) it is not accelerating.
   - (d) both a and b are possible.
   - (e) both a and c are possible.
   - (f) a, b, and c are possible.

7. At exactly $t0 = 3.0sec$, a student is pulling with a force Fstudent on a box which is connected to a spring, as in the diagram below. The spring is exerting a force Fspring on the box. At this exact time, the box is moving toward the right and slowing down. Assuming the friction is negligible, which statement best describes the forces at this time?



   - (a) Fstudent > Fspring.
   - (b) Fstudent < Fspring.
   - (c) Fstudent = Fspring.
   - (d) both a and b are possible.
   - (e) both a and c are possible.
   - (f) both b and c are possible.
   - (g) a, b and c are possible.

8. The direction of acceleration of an object is to the right. What is the most you can say about the motion of the object at this time?
   - (a) it is moving to the right and it's speed is increasing.
   - (b) it is moving to the right and it's speed is decreasing.
   - (c) it is to moving to the left.
   - (d) both a and b are possible.
   - (e) both a and c are possible.
   - (f) a, b, and c are possible.

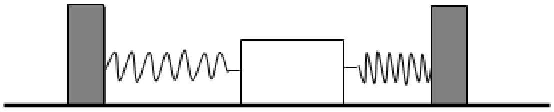
9. As a wood block slides past a tick-mark on a smooth level surface, it has a velocity of 2m/s to the right. There is a small amount of friction between the block and the surface, and eventually the block comes to rest. What are the forces acting on the block a few moments after it passes the mark (circle all that apply)?
   - (a) weight (down).
   - (b) normal force (up).
   - (c) force of block (right).
   - (d) friction (left).
   - (e) there are no forces on the block.

10. A force sensor is attached inside a soccer ball that is used during a match. The force sensor measures the forces acting on the ball. At a randomly chosen instant during the game, the sensor detects that there is only one horizontal force on the ball, and that force is directed toward the home-team goal. Which statement best describes the motion of the ball at this instant?
    (a) the ball is moving toward the home team goal.
    (b) the ball is moving away from the home team goal.
    (c) the ball is not moving.
    (d) both a and b are possible.
    (e) both a and c are possible.
    (f) a, b, and c are possible.

11. At a particular instant of time during a kickball game, a ball on the playground is accelerating to the right. What can you say about the forces on the ball at this time?
    (a) there may be several forces to the right and to the left acting on the ball, but the forces to the right are larger.
    (b) there may be several forces to the right and to the left acting on the ball, but the forces to the left are larger.
    (c) there may be several forces to the right and to the left acting on the ball, but the forces to the right are equal in magnitude to those to the left.
    (d) both a and b are possible.
    (e) both a and c are possible.
    (f) a, b, and c are possible.

12. A boy rolls a ball toward the east on level ground into the wind. The ball rolls eastward against the wind and slows down, and after a short time the ball stops and rolls westward (with the wind) and starts to speed up. At the moment the ball turns around, the velocity is zero– which statement best describes the forces on the ball at this moment?
    (a) there may be several forces to the east and to the west acting on the ball, but the forces to the east are larger.
    (b) there may be several forces to the east and to the west acting on the ball, but the forces to the west are larger.
    (c) there may be several forces to the east and to the west acting on the ball, but the forces to the west are equal in magnitude to those to the east.
    (d) both a and b are possible.
    (e) both a and c are possible.
    (f) both b and c are possible.
    (g) a, b, and c are possible.

13. A block is attached between two springs as in the diagram below, and oscillates back and forth. At an instant of time depicted in the diagram, the acceleration of the block is to the left. Which statement best describes the motion of the block at this instant?



    (a) it is moving left.
    (b) it is to moving right.
    (c) it is not moving.
    (d) both a and b are possible.
    (e) both a and c are possible.
    (f) both b and c are possible.
    (g) a, b, and c are possible.

14. At a particular instant of time, there are several forces acting on an object in both the positive and negative direction, but the forces in the negative direction (to the left) are greater. Which statement best describes the motion of the object at this instant?
    (a) it is moving to the right.
    (b) it is moving to the left.
    (c) it is not moving.
    (d) both a and b are possible.
    (e) both b and c are possible.
    (f) a, b, and c are possible.

15. A child is playing with a toy car. At one instant, the acceleration of the toy car is to the north. Which statement best describes the toy car's motion?
    (a) its speed is increasing.
    (b) its speed is decreasing.
    (c) both a and b are possible.

16. At exactly 10:02 A.M., a man is pushing to the right on a box with a force, F. There is also a friction force f between the box and the floor. If at that exact moment, the box is moving to the right, which statement best describes the forces on the box at that time?
    (a) $F > f$.
    (b) $F < f$.
    (c) $F = f$.
    (d) both a and b are possible.
    (e) both a and c are possible.
    (f) a, b and c are possible.

17. A train going from Columbus to Cleveland passes a telephone pole. What can you say about the acceleration of the train when it passes the pole?
    (a) it is accelerating toward Cleveland.
    (b) it is accelerating to toward Columbus.
    (c) it is not accelerating.
    (d) both a and b are possible.
    (e) both a and c are possible.
    (f) a, b, and c are possible.

Appendix B

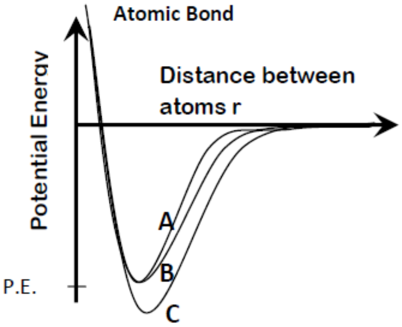
Materials Science Conceptual Evaluation (MSCE.3)

1. What is the definition of the atomic packing factor?
   - (a) It is the number of atoms in a unit cell.
   - (b) It is the volume of a unit cell that is occupied by atoms.
   - (c) It is how tightly packed atoms are.
   - (d) It is the fraction of a unit cell that is occupied by atoms.

2. Material A is denser than Material B. How does Material A's melting temperature compare to material B's.
   - (a) Material A has a higher melting temperature than Material B.
   - (b) Material A has a lower melting temperature than Material B.
   - (c) Material A has an equal melting temperature than Material B.
   - (d) a, b, and c are all possible.

3. Shown below are the Potential Energy curves for three different materials. Circle the answer with the correct rankings of the three materials' melting temperatures.

a) $T_A > T_B > T_C$
b) $T_A = T_B < T_C$
c) $T_A = T_B > T_C$
d) $T_A < T_B < T_C$



4. What is the close packed direction for body centered cubic crystal?
   - (a) Along the diagonal of a cube face
   - (b) Along the edge of a cube
   - (c) Along a diagonal through the cube center

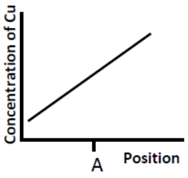
5. What is the constraining relation between the side length of the unit cell, a, and the radius of an atom, r, for a body centered cubic crystal?
   - (a) $\sqrt{a^2 + a^2 + a^2}$
   - (b) $\sqrt{a^2 + \sqrt{2}a^2}$
   - (c) $\sqrt{\sqrt{2}a^2 + 2a^2}$
   - (d) $\sqrt{a^2 + a^2}$

6. A material is made from 80% wt Mg and 20% wt of Cu. Given that Cu has a greater atomic weight, in grams per mole, than Mg, what is the atomic percent of Cu in the material?
   - (a) less than 20%
   - (b) equal to 20%
   - (c) greater than 20%
   - (d) Not enough information

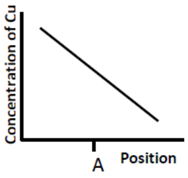
7. Which of the following materials do you expect to behave isotropically, i.e. to have properties which are independent of special orientation?
   - (a) A perfect single crystal of copper.
   - (b) A polycrystalline sheet of copper with grains aligned by rolling.
   - (c) A polycrystalline piece of copper with randomly arranged grains.
   - (d) Both a & b
   - (e) Both b & c

8. For which of the following concentration vs. position graphs is the net diffusion of copper atoms at point A in the positive x-direction, i.e. to the right as shown?

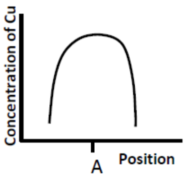
a)



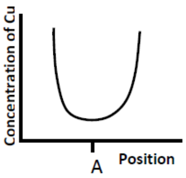
b)



c)

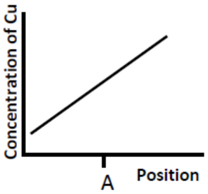


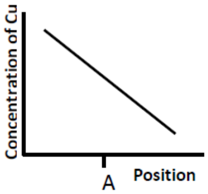
d)

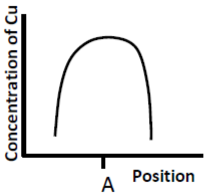


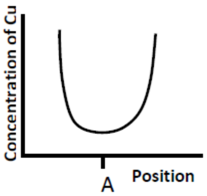
e) None of the above

9. For which of the following concentration vs. position graphs will the concentration of copper atoms be increasing with time at point A?

a) 

b) 

c) 

d) 

e) None of the above

10. Consider a single particular Copper atom in a sample that is 100% Copper at 300^{o}C. Which of the following is correct?

   (a) The Copper atom will not diffuse through the sample because the atoms are fixed in a crystal lattice.
   (b) The Copper atom will not diffuse through the sample because the sample is 100% Cu and therefore movement of atoms is not energetically preferred.
   (c) The Copper atom will diffuse through the sample because Copper atoms are not in a crystal lattice.
   (d) The Copper atom will diffuse through the sample because of thermal energy.

11. What is the Young's modulus of elasticity or "stiffness" of a material?

   (a) A measure of a material's resistance to elastic strain when under stress.
   (b) A measure of a material's ability to return to its original shape after a load is applied.
   (c) A measure of a material's ability to stretch or deform without breaking.
   (d) A measure of a material's ability to withstand an applied stress without permanently deforming.

12. Two pieces of metal A and B are the same size and shape but Metal A has a greater yield strength than Metal B. Which of the following statements is always true?

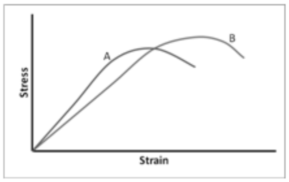
   (a) Metal A will have greater elastic strain than Metal B for a given applied load.
   (b) Metal A can withstand a greater applied load than Metal B before permanent deformation occurs.
   (c) Both a and b are true

13. Two pieces of metal, A and B, are the same size and shape but Metal A has a greater yield strength than Metal B. Which of the following statement is always true?

    (a) Metal A will permanently deform at a greater stress than Metal B
    (b) Metal A will have a greater tensile strength than Metal B
    (c) Metal A will have a greater young's modulus of elasticity than Metal B
    (d) Both a & b
    (e) a, b, & c are all true

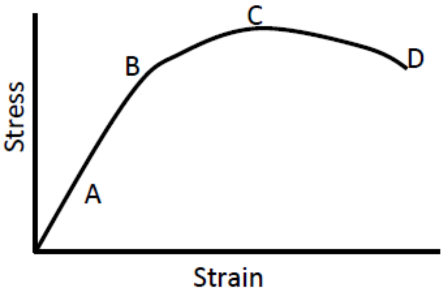
14. Which metal has a higher modulus of elasticity ?

    a) A has a higher modulus.
    b) B has a higher modulus
    c) The modulus of A is equal to that of B



15. To the right is a stress-strain curve for a metal. Which part of the curve defines the yield strength of the metal?

    a) A
    b) B
    c) C
    d) D
    e) All of the above



16. The following metal pieces are cut from the same plate. Compare the yield strength of the pieces.

    a) A has a higher yield strength than B
    b) B has a higher yield strength than A
    c) A and B have the same yield strength

(A and B have equal heights)

17. Several samples of the same kind of metal are processed in different ways to produce different grain sizes in each sample. As the size of the grains increases, how does this affect the strength of the sample?

    (a) The number of grain boundaries increases so the strength increases.
    (b) The number of grain boundaries increases so the strength decreases.
    (c) The number of grain boundaries decreases so the strength increases.
    (d) The number of grain boundaries decreases so the strength decreases.
    (e) The strength of the metal is not affected by the size of the grains.

18. A load is applied to a metal bar. The bar undergoes plastic deformation so that it is permanently elongated. After the load is removed, how does the average atomic separation of the atoms in the metal compare with the average separation before plastic deformation?

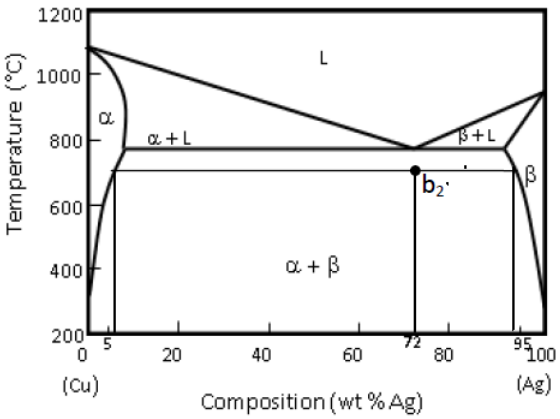
    (a) The average separation after is larger
    (b) The average separation after is smaller
    (c) The average separation after is equal to what it was before

19. Metal A has a higher yield strength than metal B. Which of the following is the best statement describing the ductility of the metals?

    (a) Metal A will be more ductile
    (b) Metal B will be more ductile
    (c) Both a) and b) are possible
    (d) Ductility depends only on tensile strength, rather than yield strength

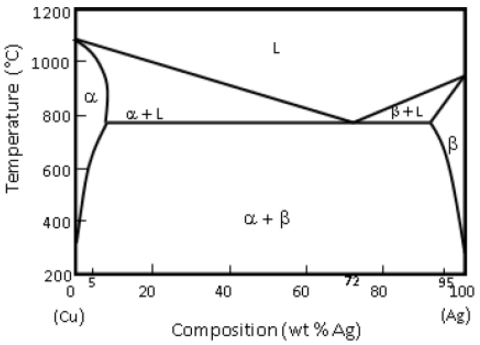
20. At point b2 what is the fraction of the alloy that is α?

    a) 6%
    b) 26%
    c) 28%
    d) 72%
    e) 94%

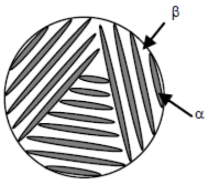


21. In the phase diagram to the right, what is the α phase?

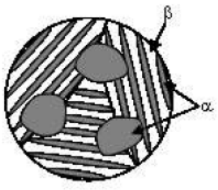
a) It is only Copper Atoms.
b) It is only Silver Atoms.
c) It is a mixture of Copper and Silver atoms with a specific fixed solubility.
d) It is a mixture of Copper and Silver atoms with a solubility that depends on temperature.



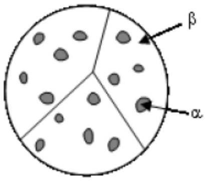
22. Which of the following sketches best illustrates the microstructure that will develop when a 85 wt% Ag alloy is cooled slowly from 1000C to 250C? (Refer to the phase diagram to the right of #21)



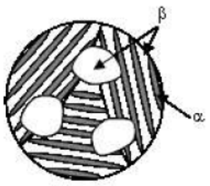
Microstructure A



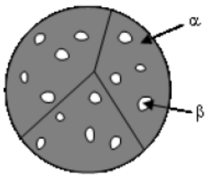
Microstructure B



Microstructure C

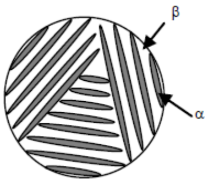


Microstructure D

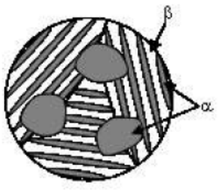


Microstructure E

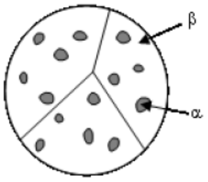
23. Which of the following sketches best illustrates the microstructure that will develop when a 6 wt% Ag alloy is cooled slowly from 1200C to 250C? (Refer to the phase diagram to the right of #21)



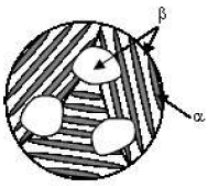
Microstructure A



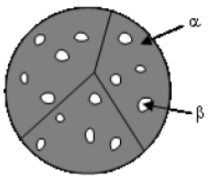
Microstructure B



Microstructure C



Microstructure D

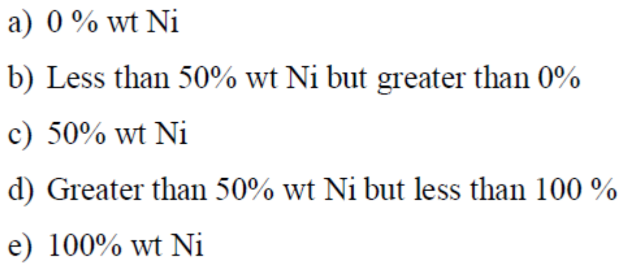


Microstructure E

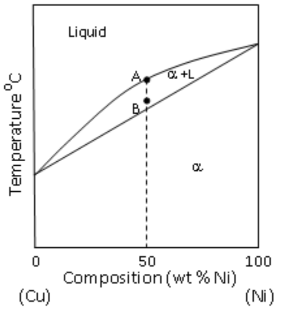
24. Two thin metal rods are cut from the same plate. Rod A is pulled through a tapered hole smaller than the rods original diameter. Nothing is done to Rod B. Which of the following is true?

    (a) Rod A has a higher yield strength than Rod B because it has more defects.
    (b) Rod A has a higher yield strength than Rod B because it has become denser.
    (c) Rod A has a lower yield strength than Rod B because it is now skinnier.
    (d) Rod A has a lower yield strength than Rod B because it has already been under stress.
    (e) Rod A has equal yield strength to Rod B because they are the same material and have the same composition.

25. Consider the Copper -Nickel phase diagram to the right. Which of the following is true about the composition of the solid alpha phase at point B in the graph?



a) 0 % wt Ni
b) Less than 50% wt Ni but greater than 0%
c) 50% wt Ni
d) Greater than 50% wt Ni but less than 100 %
e) 100% wt Ni

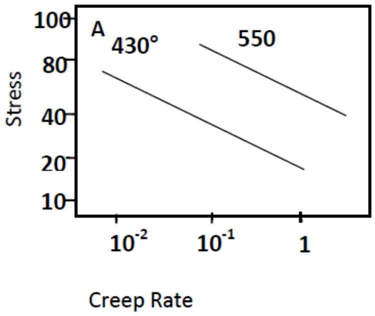


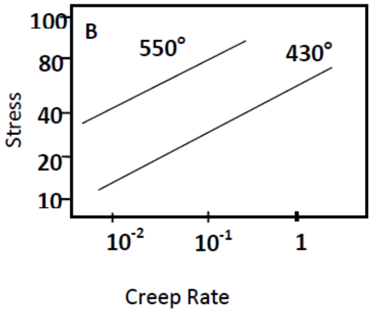
26. An engineer is designing a metal component of a machine and is considering the fatigue lifetime of the component. The component will be subject to repeated loads under tension. Which of the following statements is true?

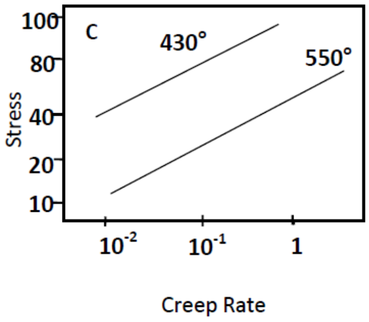
    (a) As long as the maximum stress is less than the yield strength, failure of the component will not be a problem.
    (b) As long as the maximum stress is less than the tensile strength, failure of the component will not be a problem.
    (c) Increasing the cross sectional area of the component will reduce the stress and increase the fatigue lifetime.
    (d) Both a and c are true
    (e) Both b and c are true

27. An engineer is designing a metal component for a machine. The component will be subject to a constant low level of stress but relatively high temperatures, up to half of the melting temperature of the component. Which of the following statements about the creep lifetime of the component is true?
    - (a) Because the stress on the component is low, i.e. significantly less than the yield strength, creep of the component will not be a problem.
    - (b) Because the stress is constant, unchanging, creep of the component will not be a problem.
    - (c) Because the temperature is high, creep of the component will be a problem.
    - (d) Both a) and b) are true.
    - (e) None of the above are true.

28. A metal is put under tension, and its creep rate as a function of applied tension stress is measured at two different temperatures, 430^{o}C and 55^{o}C. Which pair of lines A, B, or C are possible graphs of the measurements.
    - (a) A
    - (b) B
    - (c) C
    - (d) None of the above are possible







29. Refer to the TTT diagram to the right. Austenite at the eutectoid composition is quenched from 750^{o}C to 625^{o}C and held for 10 s. It is then quenched to 300^{o}C and held for 10 s. Finally, it is quenched to room temperature (25^{o}C). What is the final microstructure?

a) 100% austenite

b) 100% martensite

c) 50% pearlite + 50% austenite

d) 50% pearlite + 50% martensite

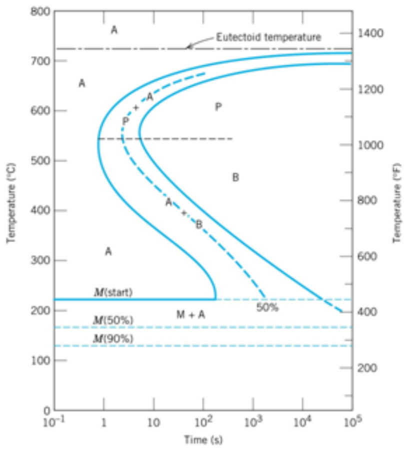


FIGURE 10.13 The complete isothermal transformation diagram for an iron–carbon alloy of eutectoid composition: A, austenite; B, bainite; M, martensite; P, pearlite.

30. Which of the following will always be effective in distinguishing a difference between a sample of austenite and a sample of pearlite?

    (a) The weight percent of Carbon is different between the two materials

    (b) The phase(s) present in the two materials are different

    (c) Both a & b

    (d) There are no composition or phase differences between the two materials

31. Which of the following is the correct ranking of the metals, from greatest to least strength, for a fixed wt% of carbon?

    (a) Martensite, Austenite, Pearlite

    (b) Austenite, Martensite, Pearlite

    (c) Pearlite, Austeniste, Martensite

    (d) Pearlite, Martensite, Austenite

    (e) Martensite, Pearlite, Austenite

Appendix C

MATERIAL SCIENCE TUTORIALS

Here we recreate the tutorials used. Because of the difference in margins of this thesis and the actual tutorial pages, the spacing here is imperfect and does not represent the real amount of room left for students pictures and responses on the tutorials.

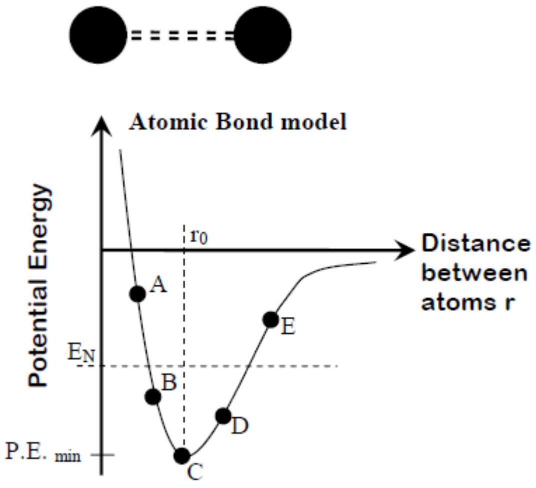
The tutorials are

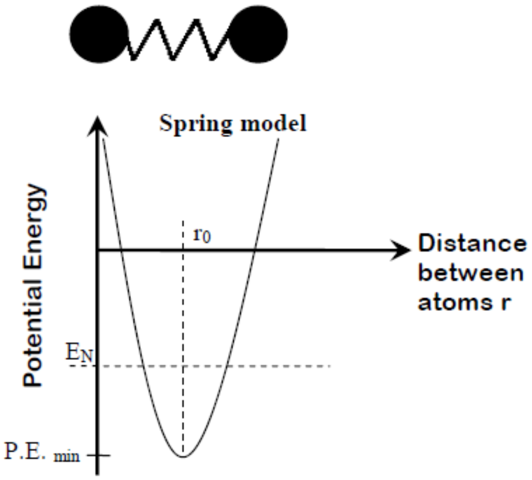
Bonding Properties of Atoms Name(s)________________

Pre Activity:

1. When a metal is heated it expands, why does this happen?

A model of atomic bonding: The symmetric spring potential vs. the real asymmetric potential.





Activity I: Understanding the graphs

2. For points A, C, and D draw pictures representing the two atoms and their atomic separations.

A

C

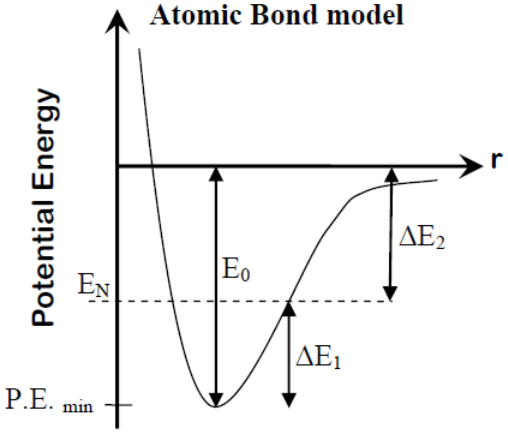
D

3. For the points A-E which points are allowed when the system has total energy EN? Give an explanation for why points would be allowed or not allowed.

4. What is the average atomic separation in the spring model case? What is the average atomic separation in the “atomic bond” case? (Greater than, less than, or equal to r0?)

5. What happens to the average separation for the spring model and atomic bond model when the total energy increases?

6. If the total energy of the system is EN, give a simple description of what each of these ΔE energies are.



E_0 = Bond Energy = $\Delta E_1 + \Delta E_2$

ΔE_1

ΔE_2

Activity II: Using the Graphs to Explain/Describe Properties of Bonding

7. When you heat up a material, what happens to the total energy between two adjacent atoms? When you heat up a material, what happens to the average separation between atoms for the spring model and atomic bond model? Why?

8. On the atomic bond potential energy graph, indicate the total energy needed to break the represented atomic bond. (This bond energy is related to the melting temperature of a system. Although due to effects of entropy the two are not proportional to each other.)

9. What is the total energy needed to break the bond in the spring model?

10. If I increase the temperature of the material, what happens to its bond energy and melting temperature? Why?

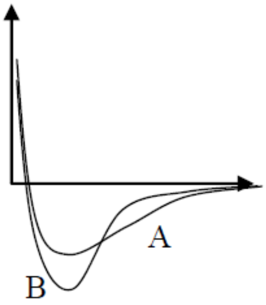
Activity III: Draw the Graphs and Explain your Picture.

11. How would the potential energy curve for a material with a greater bond energy be different? (Either a written explanation or a comparison picture with two labeled curves is sufficient.) B.E.:

12. What about a material with a higher coefficient of thermal expansion? (Either a written explanation or a comparison picture with two labeled curves is sufficient.) Hint: $\alpha \sim (\Delta \text{ r})/(\Delta \text{ T})$:

13. What about a material with a higher modulus of elasticity? (Either a written explanation or a comparison picture with two labeled curves is sufficient.) Hint: A stiffer material, higher modulus, is a stiffer spring. P.E.spring $= \frac{1}{2} * k * x^2$ try graphing a few different k to see what happens. Modulus:

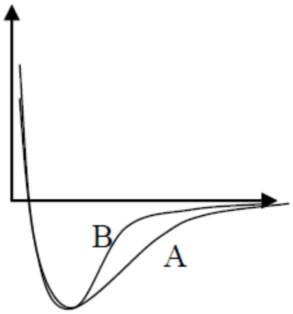
14. Concept check: For each situation rank $(<, >, or =)$ the two curves A & B by the requested properties.



B.E.:

Modulus:

α:



B.E:

Modulus:

α:

Activity IV: Looking Forward.

15. Material A is denser than Material B. How does Material A's melting temperature compare to material B's? Explain.

Activity V: Looking Backward.

16. When a metal is heated it expands, why does this happen? How has your answer to this question changed?

Properties of Crystals and Defects **Name(s)**____________

Pre Activity:

1. Material A has a smaller average atomic separation than Material B, i.e. the atoms in Material A are on average closer together. Which of the following statements is true?

a) Material A has greater mass density.
b) Material B has greater mass density.
c) Material A or Material B could have the greater mass density.

Activity I: Crystals and Atomic Packing Factor

2. Give an explanation of the atomic packing factor which another student who has never had materials science would understand.

3. Draw a body-centered cubic (BCC) unit cell.

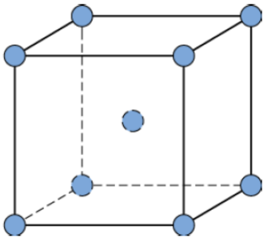
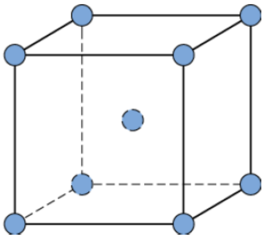
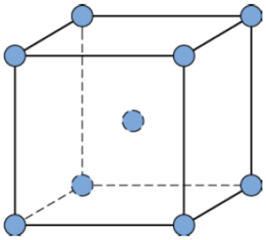
4. Determine the relation between the length of a side of the unit cell, a, and the radius of the atoms, r. (Please show your work.)

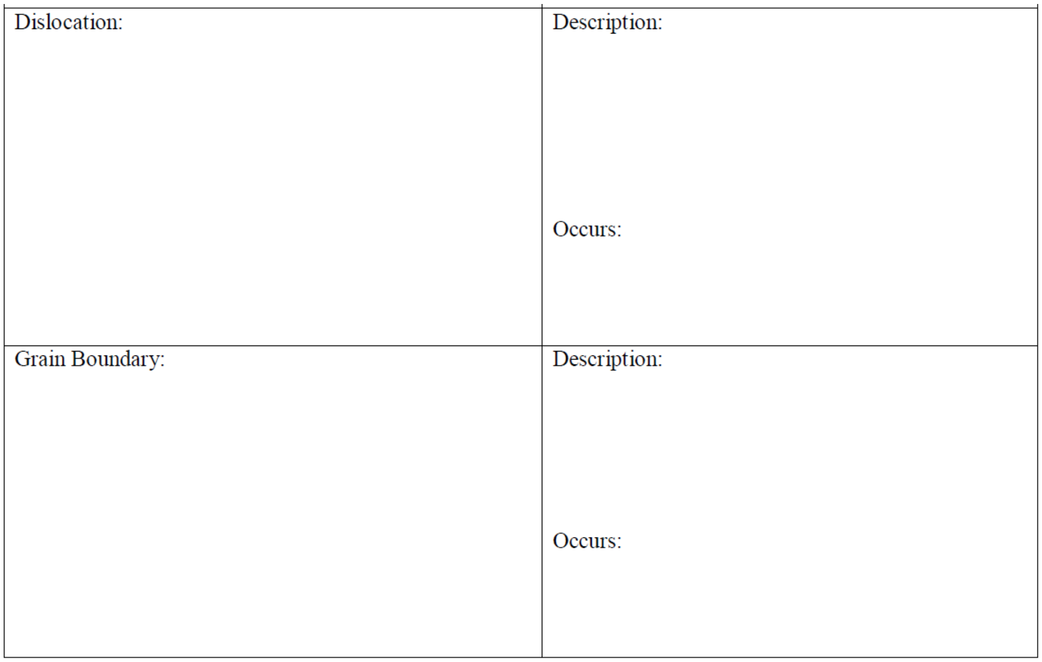
5. Using your relation between a and r find the atomic packing factor for a bcc unit cell. (Again please show your work.)
Hints: atomic packing factor = Volume of atoms in a unit cell / Volume of a unit cell.
Volume of atoms = Number of atoms*$4/3 * \pi * r^3$

6. Imagine that you fill this room with close-packed basketballs. Then you fill the room with close-packed marbles. How does the amount of empty space in the room when it is filled with basketballs compare to that when it is filled with marbles?

Activity II: Point Defects.

7. “Sketch” each defect and then briefly describe each defect and where it could occur in the unit cell.

Sketch of each defect at an atomic level	**Written description for each defect & where it can occur**
Vacancy: 	Description: Occurs:
Self Interstitial: 	Description: Occurs:
Interstitial Impurity: 	Description: Occurs:



Dislocation:	Description: Occurs:
Grain Boundary:	Description: Occurs:

Activity III: Looking Forward/Challenge Questions

8.

a) Is a perfect crystal isotropic? Explain.

b) If a metal is polycrystalline, how can the metal's properties be non-directional?

c) Can a polycrystalline metal be directional in its properties? Explain.

Post Activity

9. Material A has a greater mass density than Material B. Which of the following statements is true?

a) Material A has smaller average atomic separation.

b) Material B has smaller average atomic separation.

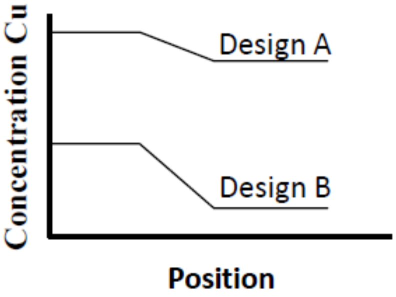
c) Material A or Material B could have smaller average atomic separation.

Explain your reasoning:

Diffusion **Name(s)**__________________

Pre Activity:

1. For which Design is the Diffusion Flux of Cu the greatest? Explain your response.



Activity I: Fick's Laws

$$J = -D\frac{\partial C}{\partial x} \qquad \frac{dC}{dt} = D\frac{\partial^2 C}{\partial x^2}$$

2. Give an explanation of the diffusion flux, J, which another student who has never had materials science would understand.

3. In Fick's First Law, why is J, the diffusion flux, proportional to the concentration gradient, (i.e. the change in concentration with position) rather than being proportional to the concentration, C?

4. What does the minus sign mean physically? What would happen if the minus were to become a plus sign?

5. In steady state diffusion, what quantity is "steady"?

6. Are atoms moving in steady state diffusion?

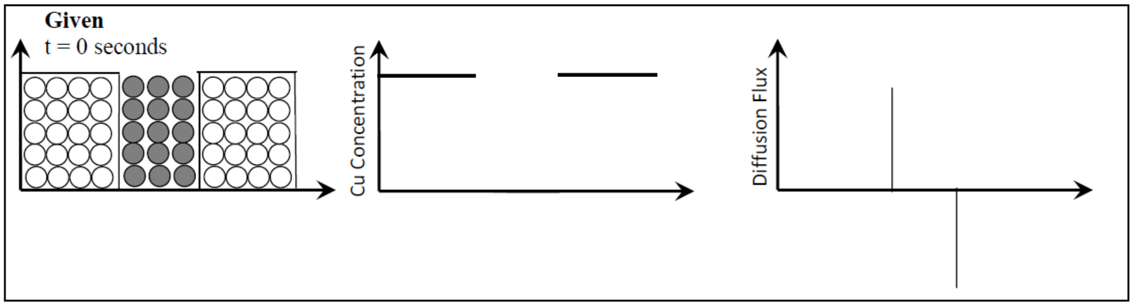
7. Is there anything that is changing in steady state diffusion?

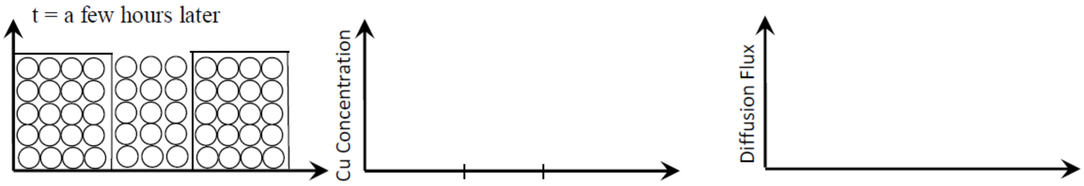
8. How does the diffusion flux, J, change with position when diffusion is in steady state?

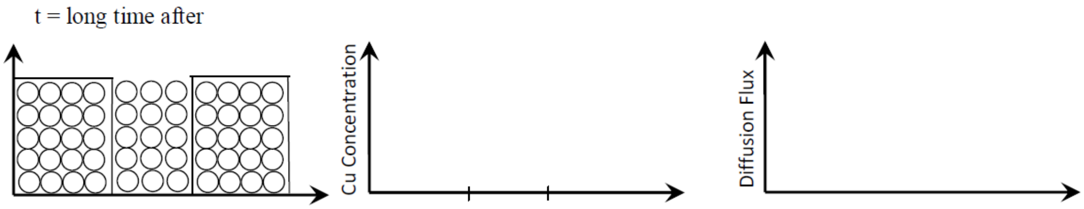
9. Give a physical example of steady state diffusion and draw the Concentration graphs and the Diffusion Flux graph as a function of position.

Activity II: Representing Diffusion by Drawing and Graphing

10. A slab of Ni atoms (gray) are perfectly sandwiched between two slabs of Cu atoms (white), as shown on the first graph marked "t = 0 seconds". Given are the graphs for the Cu concentration and the Cu diffusion flux as a function of position at t = 0 seconds. Draw the graphs for t = a few hours later (non-steady state) and t = long time after (equilibrium), show how the atoms would be distributed (i.e. color in Ni atoms) and draw the graphs for the Cu concentration and the Cu diffusion flux as a function of position.







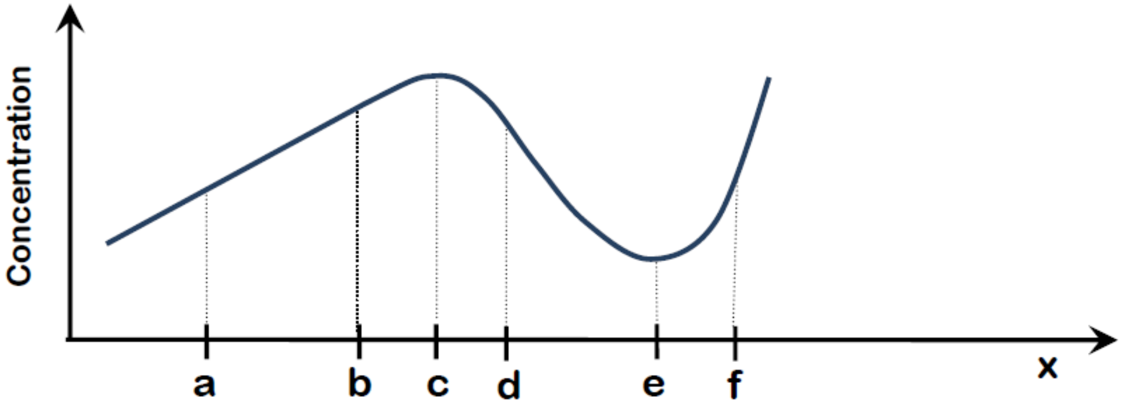
Activity III: Post Activity

Consider the graph below of concentration vs. position.

11. Rank, smallest to largest, the magnitude of the diffusion flux at each of the six points (a-f).

12. Represent the direction of the diffusion flux at each of the six points below by drawing arrows on the graph.

13. Is the concentration at point c increasing or decreasing with time? How do you know this?



Mechanical Properties **Name(s)**________________

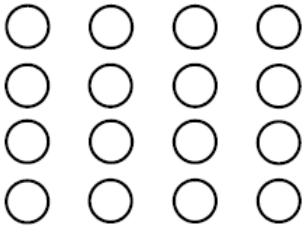
Pre Activity:

1. What is the difference between a material's strength and a material's stiffness?

Activity I: Elastic Deformation

2. A tensile stress is applied to a metal bar such that is deforms elastically. Given is a simplified view of the atoms in the bar before deformation. Sketch what the atoms will look like for conditions b) and c).

Before deformation



b) During deformation while under stress

c) After the stress is released

3. While the bar is under stress, is its volume different than before being deformed? Explain.

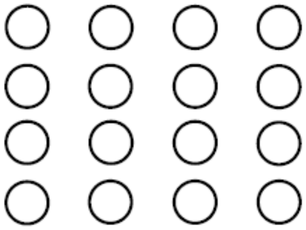
4. Does the density increase, decrease, or remain the same during elastic deformation? Explain.

Activity II: Yield Strength and Plastic Deformation.

5. How is the strength of a metal defined? (Please give description that a 2nd year engineering student who has not taken this class yet would understand.)

6. What is the difference between yield strength and tensile strength?

7. A tensile stress is applied to a metal bar such that it deforms plastically. Given is a simplified view of the atoms in the bar before deformation. Sketch what the atoms will look like for conditions b) and c).
Before deformation



b) During deformation while under stress

c) After the stress is released

8. After the stress is released, is the bar's volume different than before being deformed? Explain.

9. Does the density increase, decrease, or remain the same after plastic deformation? Explain.

Activity III: Student Dialog Questions

Below are comments from students. For each comment, indicate whether you think that each student is correct, incorrect, or some of both. Then give a quick explanation for your choice by indicating the strengths and/or weaknesses of the student's statement.

10. Student A says: "Steel has a yield strength of 180 MPa and Nickel only has a yield strength of 130 MPa. Steel is therefore stronger and more force is needed to break it."

Correct, Partially Correct, Incorrect

Explanation:

11. Student B says: "A metal with a greater yield strength will deform less before it breaks and therefore is less ductile and more brittle than a weaker metal."

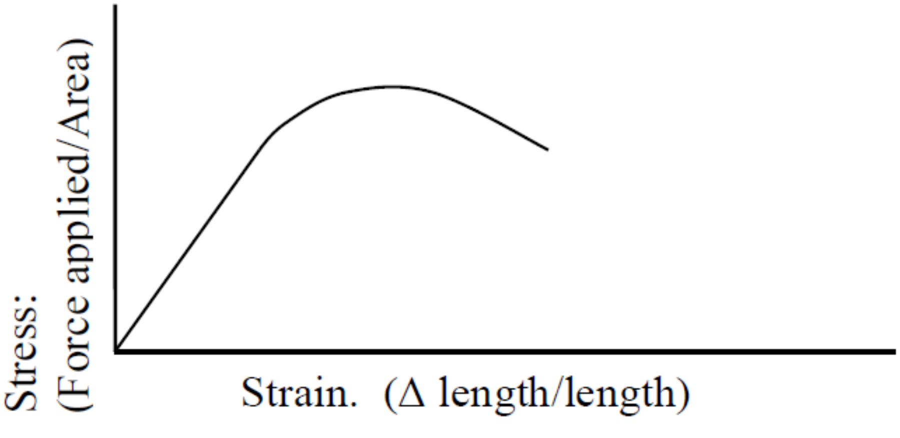
Correct, Partially Correct, Incorrect

Explanation:

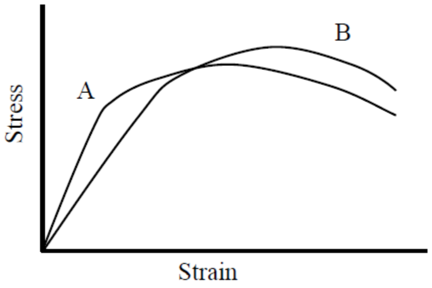
Activity IV: Stress-Strain Plots

Information about strength and Young's modulus can be obtained from a stress vs. strain curve. A stress-strain curve for a metal is drawn below.

12. Indicate the features which would characterize the Young's modulus, yield strength, tensile strength, ductility, and toughness.



13. Rank the two curves:



Modulus, E: **A** (> , < , =) **B**
Yield strength: **A** (> , < , =) **B**
Tensile strength: **A** (> , < , =) **B**
Ductility: **A** (> , < , =) **B**

Activity V: Challenge Question

14. The following metal pieces are cut from the same plate. Compare the yield strength of the pieces.

(A and B have equal heights)

a) A has a higher yield strength than B.
b) B has a higher yield strength than A.
c) A and B have the same yield strength.

Explain.

Activity VI: Looking Backward

15. When you put a metal under tension, different changes occur in the metal in the elastic and plastic regions as seen by the stress vs. strain curve and the atomic pictures you drew. Please discuss the differences between a material's strength and a material's stiffness on both a microscopic and macroscopic scale. Then summarize your ideas here. (We are not looking for definitions, and your response should be more sophisticated than in #1.)

Plastic Deformation and Strengthening Name(s)___________

Pre Activity:

1. Description of Grain and Grain Boundary:

Representative Sketch -

2. Description of Dislocation:

Representative Sketch -

Activity I: Strengthening Mechanisms

3. Two samples of metal, A and B differ only in their grain size. Sample A has a larger grain size than sample B.

a) Which piece has a larger yield strength?

b) Explain your answer choice for part a.

4. Two thin metal rods are cut from the same plate. Rod A is pulled through a tapered hole smaller than the rods original diameter. Nothing is done to Rod. B.

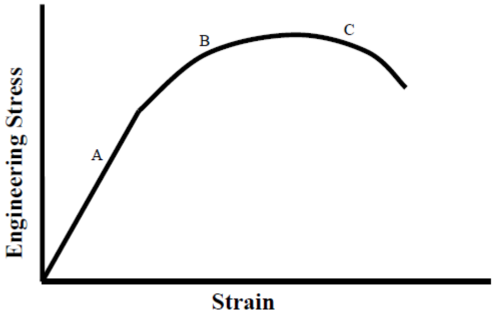
a) On an atomic and grain size level, is Piece A different than Piece B? Briefly explain how it is different or why it is not different.

b) Does Piece A have a larger, smaller, or the same yield strength as Piece B? Briefly explain.

5. Pure Copper has a yield strength of about 70 MPa and pure Aluminum has a yield strength of about 35 MPa. Will a Copper-Aluminum alloy of 5% wt Al have a yield strength that is less than or more than 70 MPa? Explain.

Activity II: Strain Hardening and Cold Working

6. Draw on the graph what would happen if you release the stress at point B and then reapplied stress until the rod breaks.



7. Imagine placing a rod under tensile stress until it reaches point B, and then releasing the tension. Does the rod now have a lower, equal, or greater yield strength than before it was placed under tension? Explain.

8. [See next page]

Activity III: Looking Forward

9. These activities suggest that when we work a piece of metal it gets stronger. However, you also know that stress on metal over time can cause the metal to break and therefore the stress must have "weakened" the metal. Talk with your group members and try to come up with a few possible reasons to explain this discrepancy. (You will learn more about how parts can fail like this next week!)

Activity IV: Looking Backward

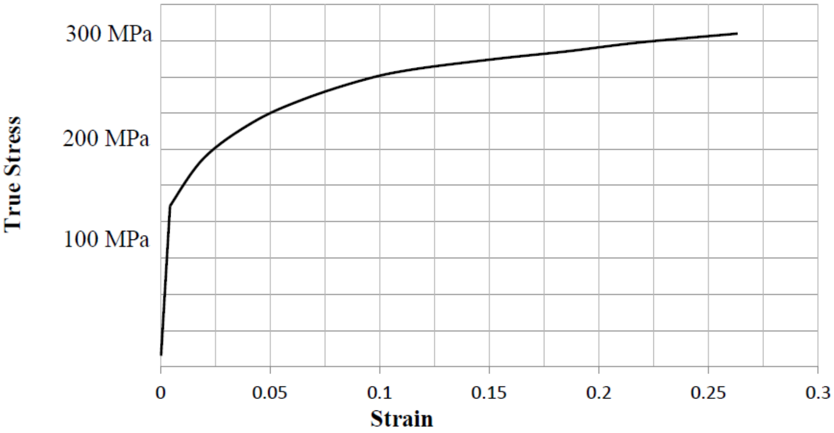
10. Student B says: "The stronger a metal is the greater its yield strength and the more force you must apply to it before it permanently deforms."

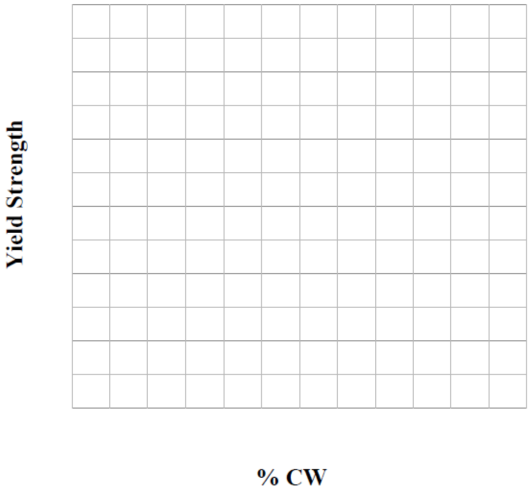
Circle: Correct Incorrect

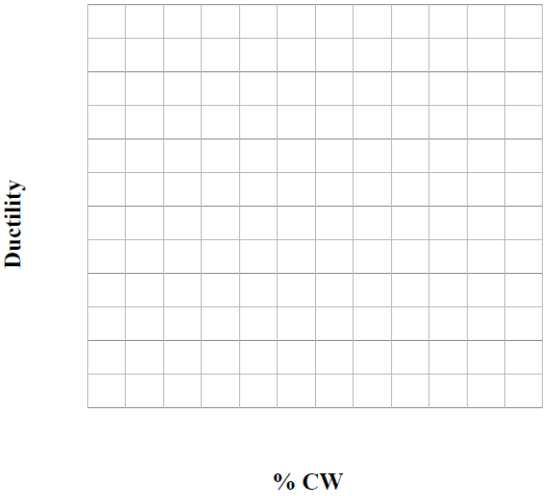
Explain:

8. Five pieces of metal are cut from the same plate. One of the pieces is put under tension until it breaks, and its true stress strain curve is plotted below. The other four pieces are put under incrementally increasing amounts of plastic strain before being released from tension. (The second piece is put under tension until it is at 5% plastic elongation and then the tension is released. The third piece is put under tension until it is at 10% plastic elongation and then the tension is released…..)

Using the true stress stain curve calculate the new yield strength and ductility for each of the 4 worked pieces. Then plot them on the graphs provided below.







Failure: Creep, Fatigue, and Fracture

Name(s)____________

Activity I : Conceptualizing Creep And Fatigue

1. a) What is creep? Give an explanation that a 2nd year engineering student who has not taken this class yet would understand. (Note: There are several necessary factors for creep to occur. These should be part of your explanation as well as an explanation of the underlying atomic/grain size causes.)

b) Why is it dependent on temperature?

c) Does it occur if the applied stress is less than the yield stress?

2. What is fatigue? Give an explanation that a 2nd year engineering student who has not taken this class yet would understand. (Again note that underlying mechanisms and necessary factors should be part of this explanation.)

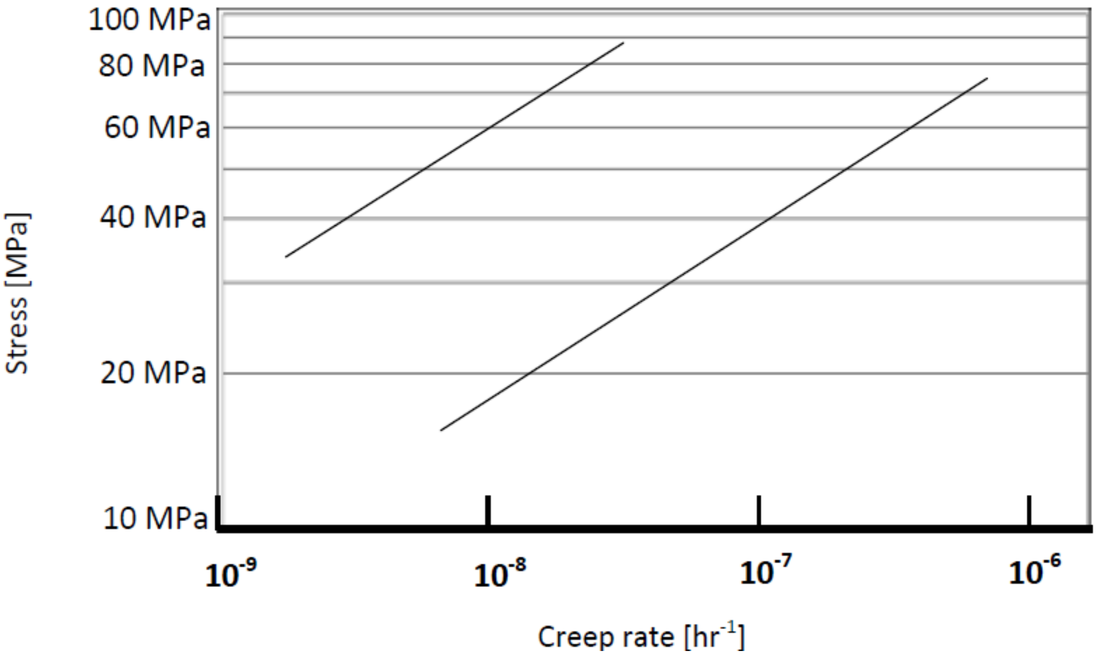
3. Pick 2 of the following engineering systems: bridge, airplane, bike, artificial joint, or make up a specific case. Describe specific examples of components/parts of these systems where creep and/or fatigue will be an issue, why they are issues, and options that you might have to improve the life time of the components.

System 1:	System 2:
Component	Component
Issues	Issues
Improvements	Improvements

Activity II: Graphical Understanding of Creep & Fatigue

4.

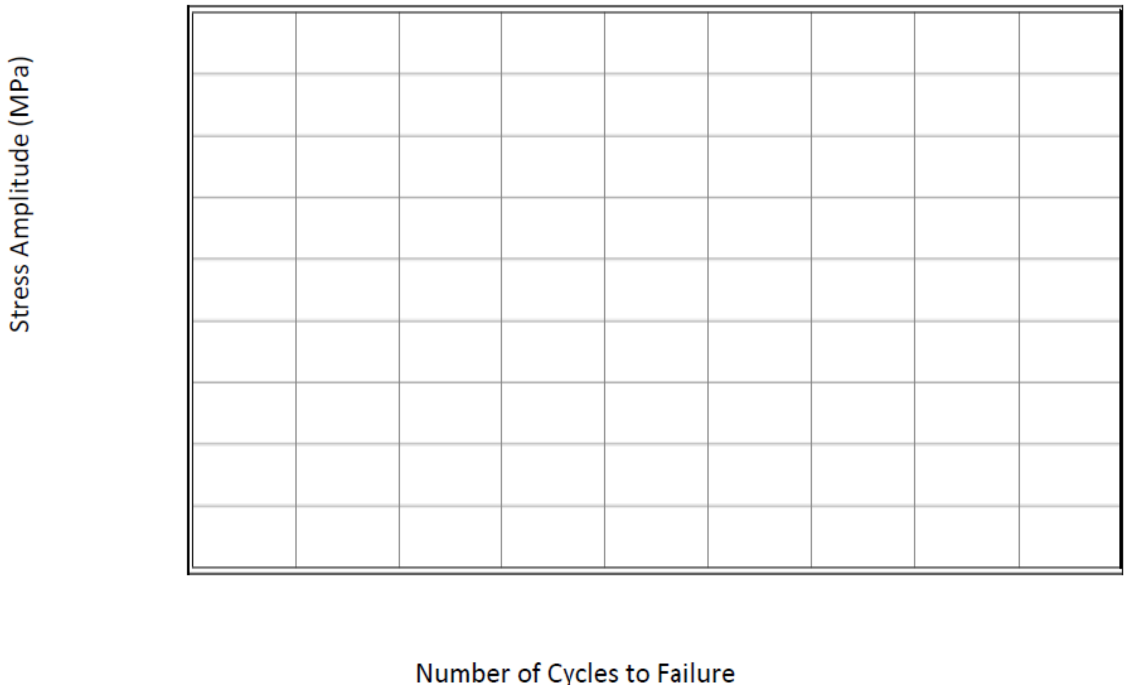
a) One of the lines in the plot below represents the creep rate (hr-1) for a metal at 500^{o}C and the other the creep rate at 400^{o}C. However, the labels have been lost. First identify which curve corresponds to which temperature. If the temperature is 500^{o}C and you apply a stress of 60 MPa to your piece of metal, indicate the value of the creep rate on the graph. Then give its numerical value.



b) You are told that your piece of metal has a yield strength of 100MPa and a length 150 cm. What will be the new length of your piece after 10 hrs of applying 60 MPa of stress to it at 500^{o}C?

5. A 2 mm diameter rod of an aluminum alloy (2014-T6) is loaded in tension and compression. The number of cycles to failure for various forces is given in the table. Fill in the table and complete the graph.

Tension and Compression Force (N)	Cycles to Failure	Stress Amplitude (Mpa)	Log$_{10}$ (Cycles) : (10^x) = Cycles
1200	10000		
1000	15000		
800	100000		
600	630000		
400	6300000		

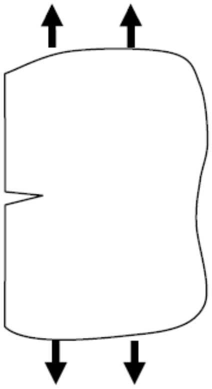


6. You are supposed to make a 50 cm long rod of 2014-T6 aluminum which must last up to at least 3*105 cycles of loading at 1200N of tension and compression force. What is the smallest diameter rod you should use?

Activity III: Failure

7. Which type of "failure" would be more acceptable from a safety standpoint, yielding or fracture? Explain your reasoning.

8. A very large sheet of a titanium alloy is loaded in tension and has a crack emanating from one edge as shown. (Assume that the crack is much smaller than the sheet.) The alloy has yield strength 830 MPa and fracture toughness 55 MPa$\sqrt{m}$. How long can the edge crack be to ensure that the component will yield before it fractures?



9. If the crack is 0.1 mm longer than calculated in Question 8, what will happen? At what stress will failure occur?

Activity IV: Reflection
(If your group has time, this can be a helpful learning tool.)

10. As a group discuss how creep and fatigue are different and similar. Take notes on some of your ideas.

Phases and Phase Diagrams

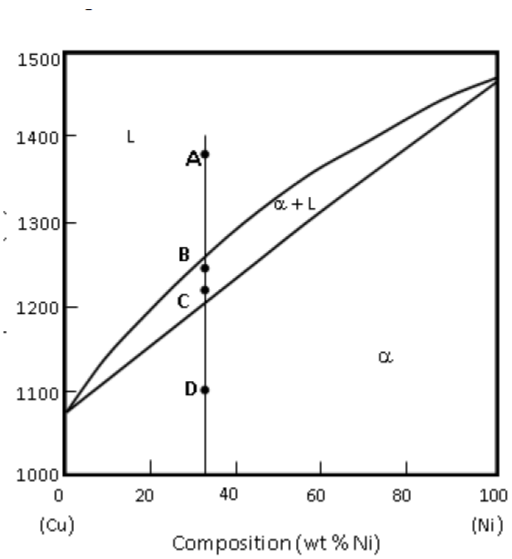
Name(s)____________

Activity I : Phases and Solid Solution Binary Alloys

1. What is a phase? Write a definition that a 2nd year engineering student who has not taken this class yet would understand.

2. Give 3 or 4 examples.

3. a) For each point draw the microstructure.



A	**B**
C	**D**

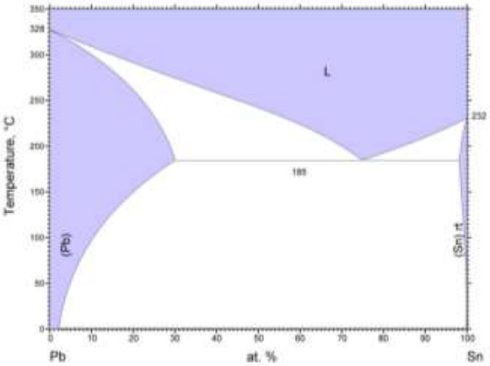
b) Is the % wt composition of Ni in the alpha phase the same for pt B and C? Explain

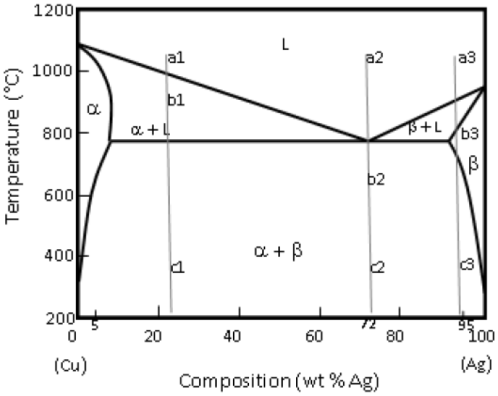
Activity II: Binary Eutectic Alloys

4. Some binary alloys (alloys with two constituents) have only one solid phase such, as the Copper-Nickel system. Others have two solid phases, such as the Lead-Tin system.

a) Why is this?

b) Under what conditions will a Lead Tin alloy have only one phase?





5. Draw pictures of the microstructure of this Copper-Silver alloy as you slowly cool the solution.

At 97% Ag

a3:	**b3:**	**c3:**

At 72% Ag

a2:	**b2:**	**c2:**

At 24% Ag

a1:	**b1:**	**c1:**

Activity III: Using the Phase Diagram to Determine Phase Composition and Amounts.

6. At point b1 in the above diagram, estimate the composition of the α?

7. At point b1 in the above diagram, estimate the composition of the Liquid?

8. At point b1 in the above diagram, use your estimates above to calculate the fraction of the microstructure that is α?

9. At point b1 in the above diagram, use your estimates above to calculate the fraction of the microstructure that is Liquid?

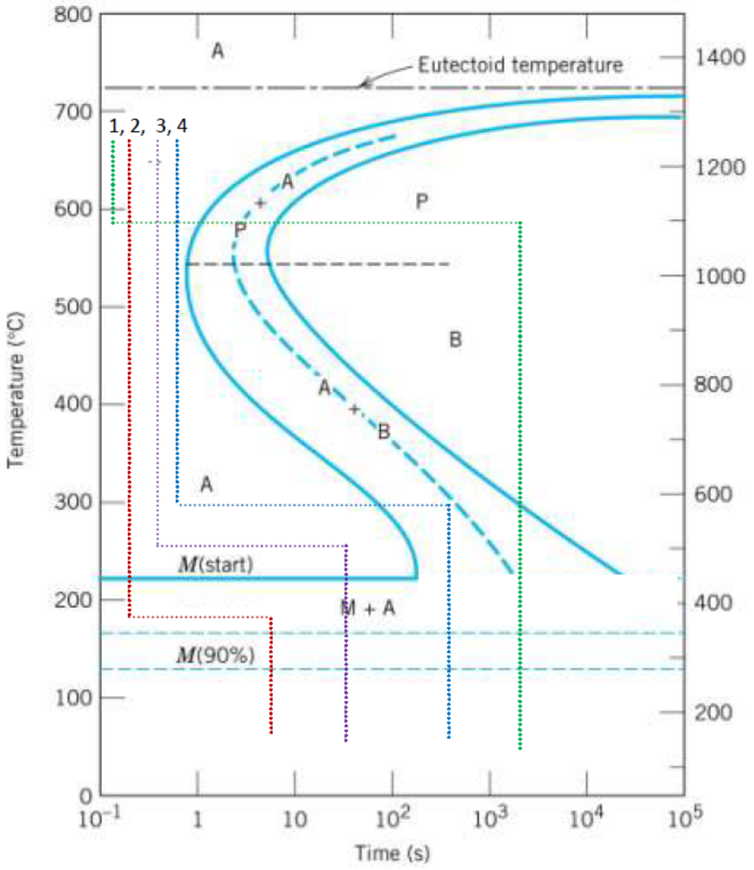
Activity IV: Looking Backward

10. When water is at a certain temperature, a small change to the temperature is all that is needed to go from all solid to all liquid. Does the same thing happen to binary alloys? Explain.

11. In the Cu-Ag eutectic phase diagram on page 2, what is alpha? Draw a picture of alpha at the atomic level.

12. What is the difference between phase <u>composition</u> and phase <u>fraction</u>?

FIGURE 10.13 The complete isothermal transformation diagram for an iron–carbon alloy of eutectoid composition:



TTT Diagrams Name(s)________________

Pre Activity:

Using your knowledge about ways to strengthen materials and your knowledge about how phases are formed when a material is cooled, discuss with your group what you would expect to happen to the strength of an alloy when it is cooled down quickly vs. slowly. Also, consider how the carbon content changes for fine pearlite, course pearlite, and spherodite and how this effects the strength of the materials.

Activity I: Reading a T-T-T Diagram

1. What are the meanings of A, P, B, and M? (For each, give a description of what phases are present as well as a name.)

A:

P:

B:

M:

2. For each line of the 4 dotted lines on the T-T-T plot, describe what happens as you cool the material. (What are the micro-constituents that form?)

1.

3.

2.

4.

3. Rank the four cooling paths based on the time required to start the first phase transition. (Shortest to longest).

4. Estimate the length of time required for the total phase transition of each cooling path and rank them shortest to longest. (Note: the time is from the start to the end of transition which is not necessarily the same as start to end of cooling.)

Activity II: Properties of Eutectic Iron-Carbon Alloy.

5. Draw a picture of the final microstructure for each cooling path and label your drawing as needed, e.g. label the alpha and Fe3C in the picture if you draw pearlite.

1. 3.

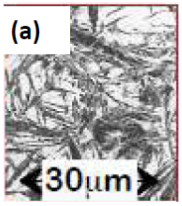
2. 4.

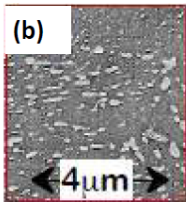
6. Rank the four material conditions based on final strength after following the different cooling paths. (Strongest to weakest)

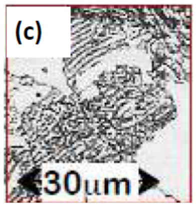
7. Rank the four material conditions in order of decreasing ductility.

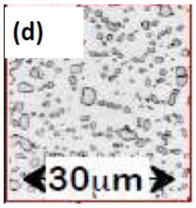
8. Below are four actual photos of microstructure.

a) Match each microstructure with its name

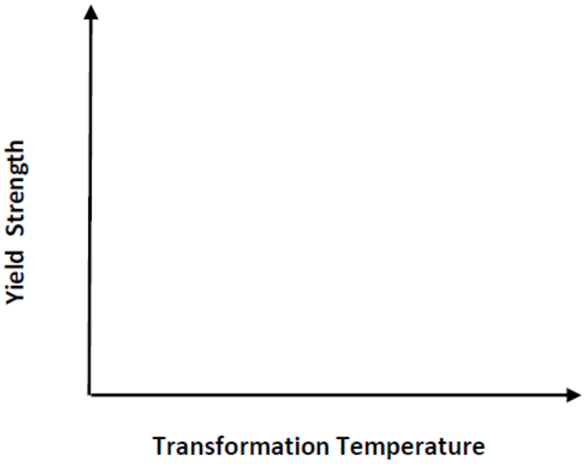








b) On the graph below plot the approximate yield strength and the transformation temperatures, which can be estimated from the T-T-T diagram, for martensite, bainite, fine pearlite, and course pearlite. (Trends, not numerical accuracy, are what we want you to be considering.)



Activity III: Looking Forward

9. For each of the following areas, how are metals and ceramics different? How are they similar?

Mechanical Behaviors:

Diffusion:

Phase Diagrams:

Activity IV: Looking Backward

10. In these examples, the stronger a material is the less ductile it is. However, we know that this does not hold in general for all materials. Why is it true here? (Describe and explain the atomic behavior which gives rise to these trends in macroscopic properties.)

Ceramics Name(s)________________

Pre Activity:

Give at least 3 examples of ceramic materials that you currently use every day. Try to be creative in your choices.

Activity I: Comparing Properties of Ceramics and Metals

Below is a list of material properties.

a. Compare Ceramics' properties to those of most Metals.

b. For each property come up with a causal mechanism for the difference between the ceramic and metallic property. e.g.: microstructure, defects, bonding, structure.

c. If there are important exceptions to these rules note them.

1. Fracture Strength (For compression and Tension)

a.

b.

c.

2. Ductility

a.

b.

c.

3. Young's Modulus

a.

b.

c.

4. Diffusivity at a fixed temperature

a.

b.

c.

5. Creep Resistance

a.

b.

c.

6. Conductivity (Electrical and Thermal)

a.

b.

c.

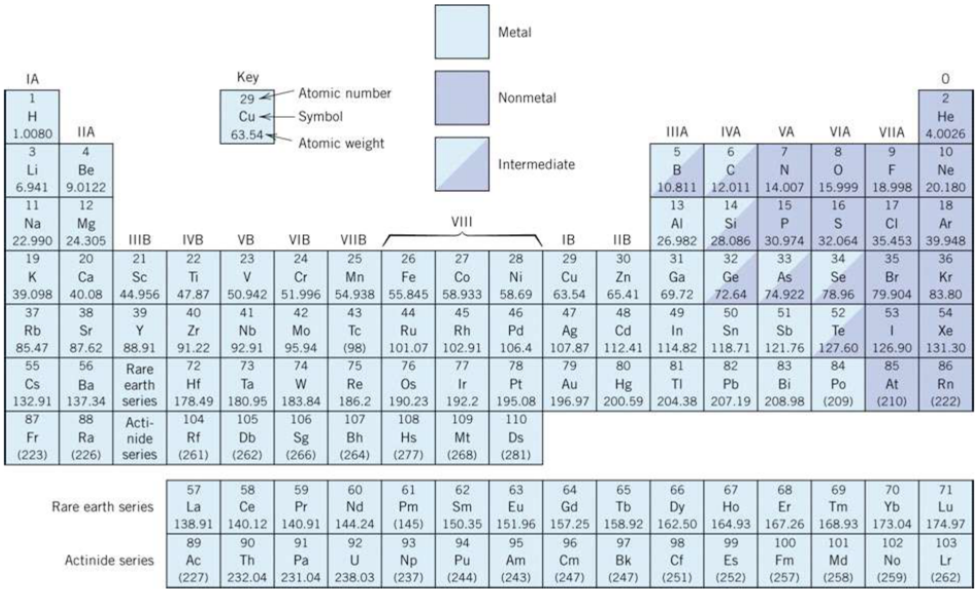
Activity II: Composition of Ceramics

Ceramic impurities are determined by charge neutrality. For the additions of impurities listed below determine allowed changes.

7. $Si^{(4+)}$ is substitutionally added to CaO →

8. $Al^{(3+)}$ is substitutionally added to TiO_2 →

9. $Al^{(3+)}$ is substitutionally added to CaO →



Key
29 ← Atomic number
Cu ← Symbol
63.54 ← Atomic weight

Metal

Nonmetal

Intermediate

IA	IIA	IIIB	IVB	VB	VIB	VIIB	VIII			IB	IIB	IIIA	IVA	VA	VIA	VIIA	0
1 H 1.0080																	2 He 4.0026
3 Li 6.941	4 Be 9.0122											5 B 10.811	6 C 12.011	7 N 14.007	8 O 15.999	9 F 18.998	10 Ne 20.180
11 Na 22.990	12 Mg 24.305											13 Al 26.982	14 Si 28.086	15 P 30.974	16 S 32.064	17 Cl 35.453	18 Ar 39.948
19 K 39.098	20 Ca 40.08	21 Sc 44.956	22 Ti 47.87	23 V 50.942	24 Cr 51.996	25 Mn 54.938	26 Fe 55.845	27 Co 58.933	28 Ni 58.69	29 Cu 63.54	30 Zn 65.41	31 Ga 69.72	32 Ge 72.64	33 As 74.922	34 Se 78.96	35 Br 79.904	36 Kr 83.80
37 Rb 85.47	38 Sr 87.62	39 Y 88.91	40 Zr 91.22	41 Nb 92.91	42 Mo 95.94	43 Tc (98)	44 Ru 101.07	45 Rh 102.91	46 Pd 106.4	47 Ag 107.87	48 Cd 112.41	49 In 114.82	50 Sn 118.71	51 Sb 121.76	52 Te 127.60	53 I 126.90	54 Xe 131.30
55 Cs 132.91	56 Ba 137.34	Rare earth series	72 Hf 178.49	73 Ta 180.95	74 W 183.84	75 Re 186.2	76 Os 190.23	77 Ir 192.2	78 Pt 195.08	79 Au 196.97	80 Hg 200.59	81 Tl 204.38	82 Pb 207.19	83 Bi 208.98	84 Po (209)	85 At (210)	86 Rn (222)
87 Fr (223)	88 Ra (226)	Acti-nide series	104 Rf (261)	105 Db (262)	106 Sg (266)	107 Bh (264)	108 Hs (277)	109 Mt (268)	110 Ds (281)								

Rare earth series	57 La 138.91	58 Ce 140.12	59 Pr 140.91	60 Nd 144.24	61 Pm (145)	62 Sm 150.35	63 Eu 151.96	64 Gd 157.25	65 Tb 158.92	66 Dy 162.50	67 Ho 164.93	68 Er 167.26	69 Tm 168.93	70 Yb 173.04	71 Lu 174.97
Actinide series	89 Ac (227)	90 Th 232.04	91 Pa 231.04	92 U 238.03	93 Np (237)	94 Pu (244)	95 Am (243)	96 Cm (247)	97 Bk (247)	98 Cf (251)	99 Es (252)	100 Fm (257)	101 Md (258)	102 No (259)	103 Lr (262)

Callister, 7th edition. The Periodic Table of the Elements

Indicate correct/incorrect for each of the following student statements? Then, if it is incorrect, correct it.

10. Be is substitutionally added to $Li_2O \rightarrow$ One Be is added for each Li.

11. Ca is substitutionally added to $TiO_2 \rightarrow$ One Additional Ca for each Ti

12. $Sr^{(2+)}$ is substitutionally added to $CaF_2 \rightarrow$ One Sr for each Ca

Activity III: Structure of Ceramics

13. A cation to anion radius ratio ($\frac{rc}{ra}$) of 0.732 - 1.0 will result in a structure with coordination number 8. As a group discuss and derive the lower bound of 0.732 using some simple hard sphere geometry arguments. What happens when $\frac{rc}{ra} > 1.0$?

Activity IV: Applications of Ceramics

14. Given your comparisons of properties in activity I, identify 3 situations where it would be advantageous to use ceramics instead of metals and 3 situations in which it is not. Try to consider multiple properties that are important for a given application and be creative in your answers.

Use Ceramics	Do Not Use Ceramics
A)	D)
B)	E)
C)	F)

Acitivity VI: Looking Backward

15. For each of the following areas, metals and ceramics have differences and similarities in their behavior. Give one or two key reasons for the main similarities and differences. (You have probably already done most of this work. The idea here is to summarize the main properties of ceramics.)

Mechanical Behaviors

Similar: Different:

Diffusion

Similar: Different:

Phase Diagrams

Similar: Different:

Polymers **Name(s)**______________________

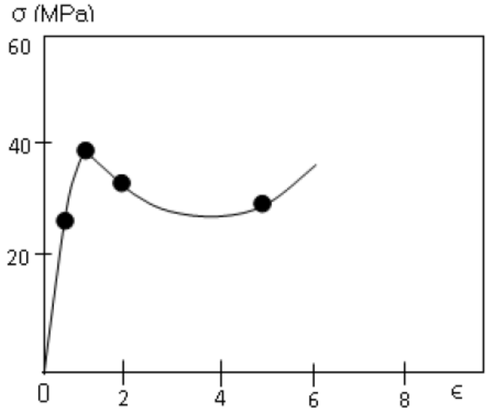
1. Polyethylene (C2H4) forms a covalently bonded polymer chain with 109^o between the main carbon bonds in the chain.

a) Draw an atomic picture of the bonding of a few sets of carbon and hydrogen atoms.

b) A polymer chain is often drawn as a random looking squiggly line. Why does the line look random if the carbon atoms are all 109^o apart?

2. The stress strain response of a thermoplastic is sketched below. For each dot on the curve state what is happening to the crystalline and amorphous regions of the polymer and draw a representative sketch.

Start:



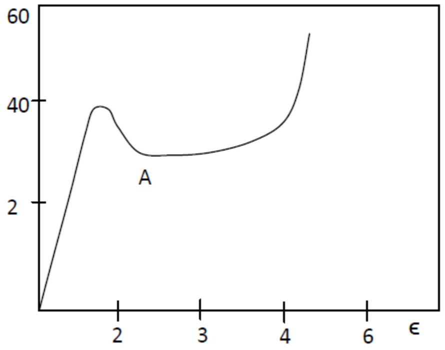
Phase 1:

Phase 2:

Phase 3:

Near Failure/Phase 4:

3. The stress strain plot for a polymer is shown below. What causes the increase in the measured stress seen after point A? Explain.



4. When a thermoplastic undergoes a drawing process, the Elastic Modulus, the Ductility, and the Yield Strength of the material all change. Indicate whether each increases or decreases and explain why drawing does this to the material.

Elastic Modulus:

Ductility:

Yield Strength:

5. Sketch the stress vs. strain behavior for thermosets, thermoplastics, and elastomers. Then note the main differences in the plots and explain the causes for these differences.

Thermoset:

Thermoplastics:

Elastomers: