CREATION OF A DIAGNOSTIC EXAM FOR INTRODUCTORY, UNDERGRADUATE ELECTRICITY AND MAGNETISM

By

Jeffrey D. Marx

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Physics

Approved by the
Examining Committee:

____
Dr. Jack M. Wilson, Thesis Advisor

____
Dr. Philip A. Casabella, Member

____
Dr. Michael J. Kalsher, Member

____
Dr. Alan S. Meltzer, Member

____
Dr. Wayne G. Roberge, Member

Rensselaer Polytechnic Institute
Troy, New York

July 1998
(For Graduation August 1998)

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

ACKNOWLEDGMENTS

First, I would like to thank my mother; without her, this would not have been possible.

I would like to acknowledge Sister Marie Cooper; she originally suggested to me the creation of a diagnostic exam for electricity and magnetism.

I am particularly grateful to Carole Flood for her patience, kindness, and understanding during what can be a lonely time for a significant other. In addition, I owe her a great debt for her crucial editing.

I would like to thank Bruce Laplante for his patient guidance on the particulars of document management and rhetoric, and for his tutelage regarding layout and formatting.

I would like to extend thanks to the physics educators that showed interest in my exam, and I would especially like to recognize those who offered meaningful comments and criticism.

Finally, I would like to thank the members of my committee: Dr. Jack Wilson, Dr. Philip Casabella, Dr. Alan Meltzer, and Dr. Wayne Roberge of Rensselaer Polytechnic Institute’s Physics department; and Dr. Michael Kalsher of Rensselaer Polytechnic Institute’s Philosophy, Psychology & Cognitive Science Department. They took a bold step by supporting the first Physics doctoral degree at RPI with a concentration in Physics Education. I would particularly like to thank my advisor, Dr. Jack Wilson, for offering me the latitude and resources to complete this project in a way that reflects its creator.

ABSTRACT

CREATION OF A DIAGNOSTIC EXAM FOR INTRODUCTORY, UNDERGRADUATE ELECTRICITY AND MAGNETISM

JEFFREY D. MARX

To fill the need for a tool testing introductory, undergraduates' knowledge of basic concepts in electricity and magnetism (EM), the investigator has developed a sixty-six item, multiple-choice diagnostic exam (DEEM). This thesis chronicles the iterative development of the DEEM.

The exam is intended to provide physics instructors with a psychometrically sound instrument that serves three general purposes: (1) gauge students' pre-instructional knowledge (baseline assessment), assess students' post-instructional achievement, and provide scores to determine students' conceptual learning gains; (2) yield results that are maximally diagnostic to instructors by highlighting students' most common misconceptions and reveling patterns of responses which are easily interpreted by physics instructors; and (3) provide the physics education research community with a tool to measure the relative effectiveness of various curricula.

The exam covers the following basic concepts of EM: forces on charged particles in electric or magnetic fields; properties of electric fields and magnetic fields; properties of electrostatic potential and potential energy; Maxwell's Equations; and induction, with Lenz's Law. Items comprising the DEEM do not require calculus, explicit calculations,

or memorization of fundamental constants; are pictorially based; and generally explore high-symmetry scenarios.

The most recent field-tested version of the DEEM (Version 3.0) yielded a coefficient alpha of 0.74 (post-test), a significant gain in score from pre-test to post-test, and no significant gender bias. The investigator established the exam's validity by surveying experienced instructors and widely used introductory physics textbooks during development of the DEEM. In addition, over 150 physics educators from around the world reviewed versions of this exam and offered comments and criticism on its content, emphasis, and layout.

“To ensure the best possible test, one must plan for it.” – E. J. Furst

TEST CONSTRUCTION PROCESS AND THEORY

This dissertation is a testament to the process of test construction and the theory that guides it. Many authors offer outlines regarding the procedure one should follow when creating a multiple-choice exam. The investigator used one work in particular (Crocker and Algina, 1986) as a reference for the steps to be considered in test construction. Their list succinctly summarizes the stages of test creation and serves as a framework from which this chapter, and thesis, is built:

• Identify the primary purpose of the exam
○ Review the literature
• Define the domain
• [Prepare test specifications]
○ Conduct interviews
• Construct a pool of items
• Have items reviewed
• [Hold preliminary item try-outs]
• Field-test
• Determine the statistical properties of the items
• Determine reliability and validity of the final form
• Develop guidelines for administration
○ Disseminate the final version

The investigator took the liberty of modifying Crocker and Algina’s series slightly. Items preceded by a solid bullet, •, are directly from their set. The text of the items blocked off by brackets, [ ], represent steps incorporated into other stages. Bulleted items preceded by an open bullet, ⚬, indicate steps that were included as part of this work but were omitted or not explicitly mentioned by Crocker and Algina.

One may infer that the architect of an exam should strictly adhere to the order delineated above. This suggests, for example, that once the literature review is complete and the test designer defines the domain, a single journal article is never consulted again. Fortunately, this is not the case; the test construction process is not linear. The designer can have confidence in the exam only after each of these stages has been addressed, and then readdressed in the context of other stages.

The sections of this chapter generally serve two purposes. The first is to introduce and describe each stage of test development. The second is to relate these processes and theories specifically to the creation of the Diagnostic Exam for Introductory Electricity and Magnetism (DEEM). In most sections, the introduction and descriptions are brief, and the reader is referred to a later chapter that delves into the topic more thoroughly.

1.1 Purpose

The exam's statement of purpose must precisely render its uses through a narrow and well-defined set of provisos. If an exam's purpose strongly states the intended use and the meaning of its scores, then the probability of the exam being widely employed by the community for which it was intended will only increase (Crocker and Algina, 1986). One can strengthen the statement of purpose by convincing educators it addresses the following two questions. First, what decisions do the test results help make? Second, what data are needed to aid this decision process (Furst, 1958)? Once the architect of the exam has addressed these questions, the purpose will be properly defined.

Prospective purposes for the DEEM include: measuring overall achievement and progress of individual students, relating students' response patterns to misconceptions in electricity and magnetism, and gauging the effectiveness of particular teaching styles or class formats (Furst, 1958). The next chapter will answer the aforementioned questions while laying out in detail the purposes for which the DEEM was created.

1.2 Literature Review

The literature review is an integral part of any dissertation or research project. The literature review for this thesis served four distinct ends. First, it helped define the exam's domain by exposing the investigator to educational research in undergraduate, introductory electricity and magnetism. (Section 1.3 will provide more insight into the meaning of the term "domain".) Second, it prevented duplication of previous work and provided guidance through inspection of similar research. Third, it provided information on secondary issues that could influence this project, such as physics educators' opinions of multiple-choice tests. Finally, the literature review yielded insight into the usefulness of novel work to the community. Chapter 3 will present a full survey of literature pertinent to shaping the DEEM.

1.3 Define the Domain

Concurrent with determining the purpose of the exam, a test author must consider the exact domain boundaries of the exam. The domain is an abstraction that represents a particular set of ideas or knowledge. The DEEM's domain are basic concepts of electricity and magnetism. More precisely, the domain of this exam is a particular sub-set of electricity and magnetism (E&M) typically included in a first- or second-year introductory, undergraduate electricity and magnetism class. Chapter 4 will be devoted to establishing exactly what concepts comprise the domain of this exam. Chapter 4 will also clarify the mathematical and conceptual level of the questions on the exam.

It is not sufficient for a test creator to simply state the concepts defining the domain. One should rely on multiple sources to help establish it further. The investigator engaged in five separate approaches to better delimit the domain: conducted interviews, searched literature, observed directly students that represented the potential pool of examinees, sought expert advice, and ascertained instructional objectives (Crocker and Algina, 1986).

The interviews limited the domain by eliminating questions understood by nearly all students and expanded the domain by including questions originally thought to be trivial. Chapter 5 will describe in detail the interview process. As mentioned in the previous section, a literature search served to define the domain by exposing the investigator to previous scholarly works regarding both the concepts on the exam and the subjects' perceptions of those concepts. Field studies further enhanced the investigator's understanding of the domain by immersing the investigator in real-world environments. For example, the investigator gained insight both by teaching E&M and by observing it being taught to a student body that represented the ultimate exam audience. Finally, expert advice on content also established domain boundaries. Experts, some of whom are potential administrators of the exam, can offer comments on how well it meets instructional objectives of a typical class (Furst, 1958). Chapters 6 and 9 will address expert review of the exam's content.

As a final note to this section, stage six (Prepare test specifications) suggests that a test designer should consider how particular parts of the domain will be emphasized on the exam. The investigator incorporated this procedure into defining the domain as well as the item construction process.

1.4 Interviews

The investigator conducted interviews with subjects that represented the pool of test takers. These interviews served many purposes. As was mentioned in section 1.2, the investigator focused the domain through interviews. For this project, interviews also served as a testing ground for potential questions on the exam. This provided the investigator a legitimate means to bypass stage 8 (Hold preliminary item try-outs). The interviews also served a valuable role in other aspects of the DEEM's creation; they helped clarify the wording of questions, established what assumptions must be explicitly described in the exam, and removed ambiguities from diagrams. For the investigator, the interviews offered a wealth of information and insight into students' misconceptions of electricity and magnetism. Particularly, the interviews furnished the investigator with a range of plausible wrong answers (also known as *distractors* or *foils*) to accompany the correct answer (also known as the *key*). Chapter 5 will discuss the role of the interview in creating this exam and a description of how the interviews were conducted.

1.5 Construction of Items

After the investigator outlined what concepts would be covered on the exam, the next step was to write items that addressed those concepts. As was mentioned in section 1.2, stage 6 (Prepare test specifications) was incorporated here, as well. The investigator determined the number of questions to be devoted to each topic and used this to approximate the total number of exam questions.

The exam's purpose addresses exactly how the exam is to be used; this, in-turn, determines what type of exam will be created. The investigator decided the DEEM would best serve its purpose as an achievement test. Achievement test item formats can be divided into two major categories: items for which the examinee generates the response (subjective items) and items for which the examinee chooses from a list of possible responses (objective items). The format of the exam has broad repercussions for the audience it will reach (Popham, 1981). The investigator determined the DEEM should be a multiple-choice exam. Chapter 3, which is a description of the literature search, will present a number of pros and cons of particular formats and offer proof that the investigator's choice of an objective exam format is sound.

Generally, an item on a multiple-choice exam consists of two parts: the stem (the question or statement of the problem) and the options (answers). The options are broken down into the key and distractors. For the DEEM, two more parts were added as possible components of an item or items: pre-stems and extended stems. A pre-stem is a statement before the stem that gives instructions to the examinee based on choices from the previous question. An extended stem is a broad set of statements that apply to a specific set of questions.

An item writer must consider what is necessary to ensure stems and options are meaningful and sound, including: item difficulty, number of options, clarity, and remaining within the bounds of the domain. Chapter 7 will address details and issues concerning item creation, as well as the pre-stems and extended stems.

1.6 Item Review

Expert review of exam items helps to ensure the content validity is strong, the items are clear, and the diagrams are appropriate. Although this is a rather informal process, it can help circumvent "test creation blindness" an author might suffer after having labored over numerous items, options, and figures. The investigator offered three versions of the DEEM to physics education researchers by notifying the community of the exam's existence through an announcement on relevant physics education listserves (phys-l, physlrnr, and physhare). Educators who desired a copy of the exam were informed to visit the investigator's web-site (currently - http://ciue.rpi.edu/jeff/) for information regarding the exam and instructions for acquiring it. Over one hundred and fifty physics professors and high school teachers from around the world requested copies of the exam. Chapter 9 includes a discussion of insights the investigator gained from review of the exam.

1.7 Field-Test

The field-test of an exam is a full-scale administration to a large pool of examinees that represent the intended audience of the test. It generates data to be analyzed for biases, difficulty, and patterns of response. It also helps establish guidelines for administration. Guidelines for administration of the DEEM are detailed in Chapter 8. Chapter 9 describes the three administrations of the exam to undergraduates enrolled in Physics II at Rensselaer Polytechnic Institute in the Fall semester of 1996 and the Spring and Fall semesters of 1997.

1.8 Item Statistics

The statistical analysis of an exam determines properties of the test as a whole and each of the items. The statistical analysis of the DEEM was divided into three parts: test analysis, item statistics, and gender biases. The analysis of the entire exam uncovered properties of the distribution of grades, gains, and reliability. An item analysis revealed items' discriminatory power, option frequencies, and difficulty. The test for gender biases determined if males or females performed better on the exam. In an attempt to detect gender biases on an item-by-item basis, the investigator employed four techniques: phi coefficient, differences in item difficulties, delta plot, and a signed Scheuneman statistic.

Chapters 10 through 12 will examine the analyses of the test, items, and gender biases, respectively. Unless otherwise noted, all of the reported statistics are for Version 3.0 of the DEEM; this is the closest form to the final version of the exam, and the administration of this version yielded the largest number of matched students from pre-test to post-test with the highest quality of scores. Chapter 9 will discuss each of the administrations of the exam.

1.9 Reliability and Validity

Reliability and validity appear near the end of the list as it was presented in the opening of this chapter; however, a test creator must be concerned with them during every stage of test creation. Briefly, reliability is a quantitative measure of the effectiveness of a test, while validity (particularly for diagnostic tools) tends to be a more subjective measure of its content.

If the exam is to be a meaningful measurement instrument, it must be consistent, or reliable (Gregory, 1992). In other words, the score one would receive on one trial of the exam would closely match a re-test score on the same test, assuming no knowledge could be gained simply by taking the exam. Chapter 6 will present the theory of reliability. Tests for reliability will be presented in Chapter 11. Techniques for improving reliability will appear in the aforementioned chapters, as well as Chapter 7.

Validity is a measure of how well the test measures what it is intended for (American Psychological Association, et. al., 1985). There are several types of validity; including: criterion-related validity and representative (or content) validity. The DEEM's purpose is to serve as a type of achievement test, so its representative validity must be strong. Discussion of the theory of validity will be confined to Chapter 6. Section 9.3 will examine expert review of the DEEM. It is important for one to note that validity rests on reliability. A reliable exam does not necessarily have to be valid, but a valid exam must be reliable.

1.10 Guidelines For Administration

The test creator must form guidelines for administration to ensure the exam is administered as intended. If the same conditions are not met for each administration, or, at the very least if the test environment can not be quantified, then the meaning of the results becomes questionable. Instructors have three basic obligations pertaining to the administration of the DEEM: limiting exam time, administering the exam at prescribed points in the course, and motivating students. The first two obligations are easy to follow and quantify; the last is not.

When the investigator distributes the DEEM to the physics education community, a description of these three obligations will be included to guaranty a consistent testing environment. This will help to ensure accurate and uniform publishing of results in the future. Chapter 8 will present administration guidelines for the DEEM.

1.11 Disseminating the Final Version

This exam is intended to serve the physics education community to aid in its assessment of introductory-level courses and students. The investigator will disseminate the final version of this exam to the education community by submitting it to an appropriate journal, most likely “The Physics Teacher” or “The American Journal of Physics.”

1.12 Final Notes

The introduction to this chapter, and the eleven sections that followed, presented the basics components of test-creation; however, two additional elements need to be addressed: *Creation of the Final Version of the Exam* (Chapter 13) and *Conclusions* (Chapter 14). Chapter 13 will chronicle the investigator’s item-by-item analysis of the DEEM. This analysis is necessary to create an even more reliable and content-valid diagnostic tool, which is based on all that was discussed and discovered up to that point. Chapter 14 will close this dissertation with summaries, findings, and recommendations.

“The only failure a man ought to fear is failure in cleaving to the purpose he sees to be the best” – George Eliot

PURPOSE

The statement of purpose for an exam defines what the exam is intended to measure and how those measurements will be utilized. A test designer must choose a purpose and state it with the utmost care. A designer who mindfully observes this statement of purpose will produce an exam with well-defined boundaries and a sound arrangement. If the designer clearly conveys the purpose to the intended instructional community, then the likelihood the exam will enjoy widespread use increases.

An exam may have more than one purpose, and each purpose may carry a different weight. And although each purpose must be distinct, as a whole they are not necessarily divergent.

Before the investigator began creating the DEEM, several possible purposes were readily apparent. Previous diagnostic exams offered clues to the DEEM’s potential and limitations. In other areas of physics, diagnostic exams are employed to infer student knowledge of a particular domain (Hestenes and Wells, 1992; McDermott and Shaffer, 1992). They are also used to evaluate the effectiveness of different curriculum styles (Hake, 1998). The investigator drew from these and from traditional applications of diagnostic exams to establish the six basic purposes of the DEEM.

1) Statement of Purpose

The DEEM has six fundamental purposes. The first three dictate how the exam should be employed as a direct measure of an undergraduate's achievement in particular areas of electricity and magnetism, before and after instruction. The fourth purpose offers the exam as a tool to gauge changes in a student's understanding of electricity and magnetism. The fifth purpose affirms that the exam can be used to gain insight into students' pre- and post-instructional erroneous notions of electricity and magnetism. The final purpose states that large numbers of average scores, or gains, can be used to provide assessment of different teaching environments.

Purpose One

The DEEM, as a content-valid achievement test, will enable instructors to measure their students' understanding of basic concepts of electricity and magnetism included in the exam. They will accomplish this by analyzing a student's score and response patterns from the post-test. The instructor will administer the post-test after covering all of the material on the exam.

Purpose Two

The exam will establish a baseline of students' knowledge as they enter their first introductory, college-level electricity and magnetism class. To meet this purpose, the exam must be administered to the students preceding current instruction of the concepts contained in the exam. The information gleaned from the pre-test will furnish perspective on students' knowledge of E&M prior to instruction. An instructor may use this information to formulate assumptions regarding the inceptive instructional level and about which topics need to be stressed or can be omitted.

Purpose Three

The results from the exam, both pre- and post-test, will be readily interpretable by the instructor. This is to say, the questions and options will openly reveal a pattern of student thought which physics instructors can link to specific deficits in students' comprehension of E&M concepts. To accomplish this, the stems and options were chosen so they do not "split conceptual hairs," or cause confusion over what the student means with a particular option choice.

Purpose Four

The DEEM will determine individual student gains based on the level of accomplishment from the pre-test to the post-test. Any instructor recognizes that students begin and end a course at different levels of achievement. Since the students will take the same exam before and after instruction (see *Purpose One* and *Purpose Two*), the data needed to quantify their learning gains will be available. An instructor who tracks the *percent gain* of a student will be more informed and can use this information to decide if the student's conceptual knowledge gains are sufficient, or if there needs to be more emphasis in a particular area. (Section 10.5 will present the definition of percentage gain.)

Purpose Five

The exam will provide insight into students' misconceptions regarding basic concepts of E&M. To do so, the exam includes options that are plausible distractors. As a pre-test, the exam will furnish information on students' pre-instructional misconceptions. Students may carry over these misconceptions from a previous course, say in mechanics, or they may have developed from some "common sense" notions. As a post-test, the exam will reveal which pre-test misconceptions remain and which new misconceptions were generated over the course of the semester. An instructor can use the pre-test data to eliminate misconceptions before they become ingrained in the students' cognitive schema of E&M. And an instructor can use the post-test data to create new ways of presenting material to avoid having future students complete the class with the same old misconceptions.

Purpose Six

After the exam has experienced widespread use, the physics education community can employ it to compare and contrast student learning gains in different curricula. Once there is a large body of data to draw from, researchers can confidently use average percentage gains to determine what types of curricula more effectively aid students in learning the basic concepts of electricity and magnetism covered on this exam.

2) Summary

In summary, the investigator has laid out these six purposes as a guideline for the instructional community for using and interpreting this exam. It is certainly possible for a physics instructor to use this exam for other purposes, however, those results must be examined carefully. For example, college-level students are the exam's intended audience, but, it is conceivable the DEEM could be administered to high school students as well. In that case, the exam's administrator would need to take steps to ensure that the validity and reliability of the exam remain intact for a pool of students at the high school level.

Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information on it." – Samuel Johnson

LITERATURE REVIEW

The purpose of this chapter is to offer background on the research related to multiple-choice exams. Section 3) will describe several cases of multiple-choice exams employed in physics education research to measure students' learning gains in introductory courses. Section 3) will present views on the appropriateness and effectiveness of multiple-choice exams in physics education. Section 8) will review literature on the creation of diagnostic exams, both inside and outside the physics education community. Section 9) will offer suggestions from the literature regarding possible grading strategies for multiple-choice exams. Section 10) will overview previously published multiple-choice exams in the field of introductory electricity and magnetism. Section 11) will introduce research on students' misconceptions of the basic concepts of electricity and magnetism. Finally, section 12) will briefly discuss gender issues in science and multiple-choice testing.

3) Multiple-Choice Assessment in Physics Education

Assessment plays a large role in shaping physics education. Many instructors want to explore new pedagogical territory to discover novel methods for presenting material. Multiple-choice assessment tools provide a means to objectively measure changes in students' understanding, or to compare groups of students. Nearly all of the multiple-choice exams mentioned in this chapter were part of larger studies to evaluate student learning. The authors generally used the information gleaned from assessment to suggest improvements to the relevant instructional community. The next few paragraphs will site specific examples of the role of assessment in physics education.

One of the largest paradigm shifts in physics education in recent years has been the introduction of computers into the classroom. At the University of Maryland, “active-engagement tutorials,” using microcomputer-based laboratories (MBLs), were introduced to replace the traditional problem sessions (Redish, et. al., 1997). To establish the effectiveness of styles of learning associated with these laboratories, the authors used questions from the *Force Concept Inventory* (FCI) on the final exam. They chose a sub-set of questions from the FCI, which were directed at the specific Newtonian concepts students need to conceptually navigate the laboratories. The results of the final exam showed that students enrolled in the sections using the MBLs scored significantly higher on the multiple-choice questions as compared to students in the traditional recitation sections. Both groups of students also answered free-response questions on the final exam which revealed that, although the MBL students had a firmer grasp of Newtonian physics than students in traditional sections, in both groups there was “still room for improvement.”

Another use of computers in physics instruction are *Interactive Lecture Demonstrations* (ILDs). ILDs differ from MBLs in that students do not use the computer themselves to carry out an experiment; rather the instructor performs a kinematics or dynamics demonstration in front of the class, and a computer is used to display position, velocity, acceleration, and/or force in real time. Students become engaged in this method because, before the demonstration is run, the instructor asks them to make predictions on a sheet of paper about the motion and forces, and then discuss their predictions with their classmates. After the demonstration, the instructor requests the students to copy the results displayed on a screen and compare their predications with the actual physical outcome. To evaluate the effectiveness of this method of instruction, its creators examined student gains on the *Force and Motion Concept Evaluation* (FMCE) (Sokoloff and Thornton, 1996). The FMCE is a multiple-choice test used to evaluate students' understanding of the Newtonian concepts of force and how force affects motion. Using the FMCE, the authors determined that students exposed to the ILDs significantly improved internalization of Newtonian concepts over students who were not exposed to the ILDs.

Not all assessments have been of novel instructional techniques. A landmark work in the development of multiple-choice assessments in introductory physics used a multiple-choice exam to determine a baseline for undergraduates' knowledge of mechanics (Halloun and Hestenes, 1985a). This work had two facets. First, the authors described their reliability and validity studies used in the production of the mechanics diagnostic exam. Second, they used their exam to determine what effect students' initial knowledge of mechanics had on performance in introductory physics classes. They concluded that one's initial knowledge has a large effect on how well one does in an introductory physics class. Furthermore, they determined that "conventional instruction" only has a small impact on students' views of the Newtonian world.

One of the broadest surveys of the effectiveness of interactive engagement-style courses (IE) versus traditional courses (T) used the FCI and the *Mechanics Baseline* (MB) (a multiple-choice exam designed to test students' quantitative understanding of mechanics) as the tools for acquiring data (Hake, 1998). The author compiled FCI data of over six thousand students, in 62 courses (14 T, and 48 IE). He also collected data from thirty of these courses that administered the MB. The results showed average gains on the FCI to be 0.48±0.14 in IE courses and 0.23±0.04 in T courses. (An in-depth discussion of gains will be covered in Chapter 10.) The results from the MB suggested that IE courses improved problem-solving skills, as well.

A final example of the role of assessment in physics education comes not from mechanics, but rather from electric circuits (McDermott and Shaffer, 1992; Shaffer and McDermott, 1992). The authors offered a detailed description of the role of research in curriculum development. Part of their assessment portfolio was to present students with "real-life" drawings of circuits containing batteries, wires, and bulbs and to ask them to pick out the idealized circuit diagram that represented the drawing. The authors determined that students were not developing a deep "functional understanding that enables them to apply the basic electrical concepts." Furthermore, they reported that difficulties with electric circuits present before instruction were present at the end of the course. This suggested that the students were not challenged to really understand the material on their own, a process necessary for them to learn.

4) Appropriateness and Effectiveness of Multiple-Choice Tests

The debate surrounding multiple-choice exams essentially centers on arguments about how much can be inferred from the results of these exams. The first two subsections below offer, as an example of this debate, letters to the editor of the *American Journal of Physics* from May 1984 to April 1986. Section 6) reviews papers that investigated the effectiveness and reliability of multiple-choice exams. Section 5) reviews the debate surrounding the FCI's dimensionality.

5) AJP letters – Opposed to Multiple-Choice Exams

A debate over the effectiveness and appropriateness of multiple-choice exams was sparked by a letter in the *American Journal of Physics* (AJP) (Iona, 1984) reacting to a work that employed multiple-choice questions to investigate students' understanding of circuits (Cohen, Eylon, and Ganiel, 1983). Iona raised issues about the assumptions of some of the multiple-choice questions on their exam. Particularly, Iona contended that if one held to the assumptions made by the exam's authors, then the "correct" answers are not consistent with what an expert would consider "physically correct." This is always an issue for test creators; there are limitations for any assumptions. An exam that carries some assumptions too far may dramatically clash with reality.

In that same issue of the AJP, Cohen, et. al. were given the chance to respond. They answered the allegations by stating that some of the problems with the wording of questions resulted from a poor translation from the original Hebrew. They also asserted that in every realm of physics one must work with models. They then claimed that the model they presented and the answers they deemed correct were consistent. This small exchange started a volley of letters concerning the merits of multiple-choice exams. The next paper by Bork was particularly critical of multiple-choice exams.

In October 1994, Bork offered his opinion that Iona had missed the mark on what should be done about the problems of multiple-choice exams (Bork, 1994). He stated "…that multiple-choice questions *should never be used*." He then went on to list five reasons to support his claim. First, multiple-choice questions encourage guessing. Second, they are not characteristic of the real world. Third, they are not a friendly tactic for human beings. He supports this with the statement, "Students invariably consider them in somewhat a derogatory fashion, connected with the fact that guessing is involved." Fourth, Bork stated that "There is no real use for them. For example, we hardly ever use multiple-choice in the computer based quizzes." And fifth, it is hard to write good multiple-choice questions. He based this on his years of experience working with the Educational Testing Service (ETS). Later, another letter writer added that the wording of the questions can be confusing, and he stated he had seen A-grade students do B-grade work on multiple-choice exams and vice versa (Varney, 1984). He attributed this to careless wording of stems and questions based on weak examples. Finally, another dissenter offered two more reasons why multiple-choice questions are not effective: "The right answer may have been arrived at by a fortuitous combination of errors," and test-takers with weak English skills or whose first language is not English may not perform as well because multiple-choice exams rely heavily on reading comprehension (Sandin, 1985).

6) AJP letters – In Defense of Multiple-Choice Exams

Defenders of the use of multiple-choice exams also made their case. In November 1985, one writer listed the reasons he felt some of the previous criticisms leveled against multiple-choice exams were unfounded and he offered some positive aspects of multiple-choice exams as well (Scott, 1985). Scott dismissed most of Bork's comments with one or two lines of rebuttal. For example, he considered Bork's third argument "hardly worth taking seriously" and felt that Bork's fourth argument was just plain false. He also offered that multiple-choice exams have the added benefit that they are easy to grade and remove "unfair subjective evaluations which plagues reader-graded exams." In a later letter, Iona noted that he is actually a supporter of multiple-choice exams (Iona, 1986). (He declared that it was unfair for Scott to label him a non-supporter of multiple-choice exams as he was simply criticizing one particular exam.) He outlined nine reasons why multiple-choice exams are an appropriate form of evaluation: (1) They allow for a greater variety of questions. (2) They can be “qualitative” questions regarding physics principles (i.e. Fermi-type questions). (3) Choosing between alternatives and having a general understanding is much more like real life. (4) Options act like hints. (5) The examiner can ask questions about subtle points. (6) Multiple-choice questions are the next best thing to essay questions. (7) The examiner can ask for quick numerical calculations and make them worth a point. (8) More material can be covered. (9) Multiple-choice exams are good for review. (Although the investigator agrees with most of these points, some, particularly 3 and 6, needed more than the two line comments to really explain what the author meant.) Iona then proceeded to describe the difficulties of writing multiple-choice items, but dismisses some of the points that are usually raised against multiple-choice exams. In particular, he notes that random guessing may not be so bad as it has little effect on the distribution of the scores.

7) Validity and Reliability of Multiple-Choice Exams

As the investigator looked into narrower and narrower topics surrounding the creation of a diagnostic exam, fewer and fewer relevant papers were to be found. Nonetheless, significant insight was gained from a few papers that dealt with issues surrounding multiple-choice exams. The first was a validity study of the multiple-choice section of the Advanced Placement Physics C exam (Pfeiffenberger and Modu, 1977). This study determined the relationship between scores on the multiple-choice section of the AP Physics C Exam and college grades. Among other things, this paper reported correlations ranging from 0.528 to 0.592, which the author states, "… are quite good when compared to validity coefficients usually obtained in similar studies."

A recent article in the AJP addressed the validity of the FMCE (Thorton and Sokoloff, 1998). Their analysis revealed that this particular multiple-choice exam is very well suited for determining what students understand and do not understand of a Newtonian view of the world. Comparisons between students' choice of options and free response items showed a strong similarity. The authors reported, "The agreement between the multiple-choice and open answer responses is almost 100%." This agreement was only reported for a subset of questions on the FMCE; the entire test was not the subject of their paper. A source of this exceptionally high correlation stems from the fact that many of the FMCE items present the examinee with up to seven options. These cover nearly every possible choice that a student might make. Another test of the validity of this exam was that multiple-choice questions on the FMCE correlated well with multiple-choice questions that probed the same concept, but were cast in a totally different format, on a different exam.

Another study, from the chemistry education research community, involved a comparison of student response patterns on multiple-choice questions to stem-equivalent constructed questions, in an attempt to evaluate differences in student problem solving techniques (Barnett-Foster and Nagy, 1995). The results revealed no significant difference in solution strategies and error patterns between multiple-choice questions and free-response questions. Oral interviews, however, were found to reveal some subtle differences in the frequencies of students' various problem-solving techniques. This study was also valuable because the authors noted some of the statistics typically associated with the multiple-choice exams employed in their study. In particular, the KR-20 (equivalent to coefficient alpha for dichotomously scored exams) ranged from 0.64 to 0.74. These values are equal to, or less than, the value of coefficient alpha for Version 3.0 of the DEEM. (A complete description of coefficient alpha will be reserved until Chapter 10.)

8) Dimensionality of the Force Concept Inventory

Another charged debate in the literature surrounds what the FCI measures. The authors of the FCI (Hestenes, Wells, and Swackhamer, 1992) put forth the notion that they "… have designed an instrument to probe student beliefs on this matter (force) and how these beliefs compare with the many dimensions of the Newtonian concept." The authors break the concept of force into six dimensions: kinematics; Newton's First, Second and Third Laws; superposition; and "types" of force. Huffman and Heller contended in their study that a factor analysis of students answer's to items in the exam showed that these six dimensions do not really exist (Huffman and Heller, 1995). They claimed that the FCI questions are only loosely connected and that the exam does not measure a single concept of force, nor does it measure the six conceptual dimensions as proposed by the authors of the exam. Consequently, Huffman and Heller suggested that the FCI not be used as a placement exam. They stated that it still could be used as a diagnostic tool and as a tool for evaluating instruction, as long as the instructor using the exam is aware that it does not really measure a single force concept. Their views were challenged.

Halloun and Hestens responded to these allegations by stating that the FCI is a good exam that measures a single concept – force (Hestenes and Halloun, 1995). They based their argument on their research into the validity of the exam, the research into student misconceptions that went into the options, and the knowledge that students' understanding of concepts are generally "vague and undifferentiated." The authors then gave suggestions about what the FCI scores mean and how the FCI can be used as a placement exam, when coupled with a simple math test, for accelerated or honors high school or college physics courses. In the appendix of their article, the authors also brought to light some of their feelings regarding the shortcomings of factor analysis.

Heller and Huffman responded to these statements by giving a detailed account of factor analysis techniques. They also adhered to their original criticisms and added that a high score on the FCI does not indicate that a student has created a universal notion of force (Heller and Huffman, 1995).

Pinkerton analyzed this debate and offered evidence that suggested factor analysis is not an appropriate way to analyze the FCI (Pinkerton, 1997). Instead, he conducted a Rasch analysis and concluded the FCI does measure one concept of force. However, he did offer five recommendations for improving the FCI, which included updating particular items and inserting more difficult questions to make the exam stronger.

9) Creating an Objective Test

After establishing that at least some portion of the physics education community considers multiple-choice exams an appropriate and effective means for measuring students' understanding of a topic, the investigator searched the literature for insights into the creation of such an exam. Needless to say there is a wealth of literature and textbooks in the field of psychology and education covering many aspects of test development. The subset of that knowledge covering multiple-choice exams is still vast and widely published. The investigator used many sources as references for creating this exam;

however, three were used most extensively: Crocker and Algina (1986), Mehrens and Lehmann (1984), and Gregory (1992). They were the primary sources because of their clear descriptions and extensive references to original works. The investigator used other sources, such as Berk (1982) and Horn (1970), for more specific tasks. When available, the investigator referenced original works.

Techniques of classical and modern test theory are firmly established in psychology and education research circles; they are not as entrenched in the physics education community. But, as the community's awareness grows, so does the number of publications devoted to issues surrounding test creation.

The first work directly describing the process of creating a multiple-choice exam to the physics education community was published in the AJP in 1981 (Hudson and Hudson, 1981). This work outlined the basic considerations of creating an exam including the length of the test and differential levels of problem difficulty. This paper failed to bring to light essential methods for evaluating test items statistically. Later, a much more thorough work was published (Aubrecht and Aubrecht, 1983). This piece delved into the topics of making a concept map, issues for item construction, basic test statistics, and a description of reliability and validity. An in-depth investigation, however, was left to the reader, and deeper issues like gender biases, global student response patterns, and dimensionality were not discussed. The reader should note that all of the material mentioned in both of the previous papers were already available in all but the most basic textbooks on test creation; the reference lists at the end of both these papers confirms this. Another more recent work was much less meticulous and, accordingly, was of much less use (Sobolewski, 1996).

Interestingly enough, the creators of two widely used diagnostic exams in introductory physics, the FCI (Hestenes, Wells, and Swackhamer, 1992) and the MB (Hestenes and Wells, 1992), did not report the methods they used to construct their exams. In the papers associated with each exam, the authors did not mention how they constructed the items, had them reviewed, created plausible distractors, or what the most basic statistics of the items were. Earlier papers, which were authored by a co-creator of the FCI and MB, provided some insight into the creation of these exams (Halloun and Hestenes, 1985a and 1985b). Other sources, however, did provide details into the process of test creation which related specifically to the DEEM.

A paper containing a test intended to examine students' interpretation of kinematics graphs also provided a description for fabricating a physics diagnostic exam (Beichner, 1994). Besides offering the test, the author laid out a flowchart for developing a diagnostic exam and described some of the important statistics used to evaluate the quality of an exam and its items. The author published his exam with the article and provided a sample of the statistics it generated: a coefficient alpha of 0.83, average point-Biserial coefficient of 0.74, and an average item discrimination index of 0.36. Since the exam covered a very narrow range of physics – kinematics graphs – it is not surprising it preformed so strongly. In addition, they provided a realistic benchmark to compare the statistics generated by the DEEM. Other works, outside of the physics education community, provided more insight into the creation of diagnostic exams.

A paper describing the development of a two-tier diagnostic exam to measure students' understanding of diffusion and osmosis contained detailed information regarding the creation of a multiple-choice exam (Odom and Barrow, 1995). The authors outlined their procedure as:

Phase 1

- The content boundaries were defined with a list of propositional knowledge statements.
- Content validity of the propositional knowledge statements were determined.

Phase 2

- Students' misconceptions were identified by interview.
- Multiple-choice questions with free response reasons were constructed and administered.

Phase 3

- Final test questions were constructed based on multiple-choice questions with free response reasons.
- The final test questions were revised and a pilot study was conducted.
- Final content and face validity of each item were determined with the assistance of a specification grid.
- The final version of the Diffusion and Osmosis Diagnostic Test was administered.

The authors described each of these steps in detail. They also reported a variety of psychometric properties of the exam: split-half reliability was 0.74, difficulty indices ranged from 0.23 to 0.95, and the discrimination indices ranged from 0.21 to 0.65. There statistics are very similar to the ones generated from Version 3.0 of the DEEM.

Another paper provided a view of a slightly different type of multiple-choice exam (Reif and Allen, 1992). The authors of this paper investigated students' understanding of acceleration, and they employed two types of open-ended tests to accomplish this. The first type of test contained questions about specific situations; the second test contained questions about groups of situations. The investigators observed the students while they attempted to answer questions from these two tests. Among many objectives of this work, the authors compared the problem solving processes of both novices and experts and then used those comparisons to create a multiple-choice exam about acceleration.

10) Grading an Exam

An issue that has received some attention in the previously mentioned works, and in two other recent papers, is flexibility in grading multiple-choice choice exams, or, more precisely, granting partial credit. One paper suggested permitting students to give more than one response (Hobson and Ghoshal, 1996); another suggested allowing students to explain their responses to receive some credit(Varlashkin, 1997). Although these are both fine suggestions, they have their shortcomings when applied to the DEEM. First, letting students write explanations to defend thier responses would increase the work of the grader. Since student explanations were optional, Varlashkin contended that the amount of extra work was not tremendous. Regardless, it still would constitute more work. In addition, if the instructor permits students to write explanations of their answers, then subjectivity is introduced into the grading process. Hobson and Ghoshal offered permitting students the option to write more than one answer. This would seemingly defeat the purpose of machine grading. The authors refuted this and offered a way to have exams machine graded. This procedure, however, will not work on all automated grading machines. In particular, it would not work on the equipment used to grade the DEEM.

A variety of other techniques have been attempted to give students partial credit, or a second chance, on multiple choice exams. The *answer-until-correct* (AUC) method is one of them (Gilman and Ferry, 1972). The way this method works is simple. An examinee answers a question and receives immediate feedback. If the student’s answer is wrong, then the student is permitted to try again. The exam is scored by subtracting the total number of answers the student gave (right and wrong) from the total number of possible responses which, in most cases, is equal to the sum of all of the options on the exam. Studies of this method suggested it does not appreciably improve the quality of the exam's psychometric properties (Hanna, 1975). And considering this type of test would need to be administered to a large body of students via a computer, the limitations are still too great to warrant this as a worthwhile means for grading a test. The security issues would be particularly difficult to overcome.

Another grading technique involves having the test creator weight certain answers. In other words, some options are not necessarily completely wrong and may be worth, say, half a point. The investigator considered this technique for the DEEM. For example, if a student chooses the wrong direction for the force on a stationary electron in a uniform, static electric field, then that student could be awarded some partial credit. This type of grading has been carried out in many forms (Patnaik and Traub, 1973; and Downey, 1979). Again, the returns of these types of grading procedures have been called into question, as they also do not show any added psychometric benefits (Hedrickson, 1971; and Downey, 1979).

It seems that the clearest way to grade a multiple-choice exam is to simply mark items right or wrong. Individual instructors may try some further interpretation; however, it is dubious that more insight can be gained.

11) Previously Published E&M Exams

Some of the authors to be discussed in section 9) published multiple-choice exams as examples of formats they used to probe students' understanding of E&M. All of these are of limited use to the research community because of their brevity and narrow focus. To add to this list, the investigator reviewed other sources of multiple-choice exams that focused on E&M.

The ETS administers Advanced Placement Exams in physics to high school students every year. The purpose of these exams is to provide high-scoring students the opportunity to "place-out" of a first and/or second semester of introductory physics instruction when they reach college. Half of the exam is multiple-choice and half is free response. Of the multiple-choice questions, many relate to E&M and some of those would surely be appropriate for the DEEM. The ETS, however, copyrights these exams and rarely publishes sample exam questions.

The only previously published work of a multiple-choice exam that covers nearly the same material as the DEEM was a book entitled "Electricity and Magnetism Diagnostic Tests," (W.A Rachier, 1973). This work was well done and covers topics such as the electric field, Gauss's Law, electric potential, the magnetic field, Ampere's Law and the magnetic fields of currents, and Faraday's Law and electromagnetic induction. Each section has about sixty multiple-choice questions, and the level of difficulty is high. Many of the questions explicitly involve integrals and complex algebraic manipulation. It is obvious, from its lack of presence in the literature, that this book's tests were never widely used. This is unfortunate, in the view of this physics education researcher, as many of the test questions are quite good. Again, these questions could not be used in the DEEM as they are copyrighted by the publisher.

The only other diagnostic exam investigating similar concepts as the DEEM is a work in progress. A consortium of four schools (Dave Maloney at Indiana-Purdue, Tom O'Kuma at Lee College, Alan Van Heuvelen at Ohio State University, and Curtis Hieggelke at Joliet Junior College) are in the process of creating a diagnostic exam for introductory electricity and magnetism. At the Summer AAPT meeting in 1997, they presented a status report of their work. The most interesting piece of information (besides the fact that they were having the same difficulties as this investigator!) was that they had decided to split their exam into two parts, an electricity part and a magnetism part. Also, they also do not define their universe of basic E&M topics in the same manner as this investigator. For example, they included questions on insulators and conductors, as well as diagrams of electric and magnetic fields for particular arrangements of charges and currents (Maloney, et. al., 1997).

12) Students' Misconceptions of E&M

The literature reveals a plethora of research that address students' understanding of objects such as bar magnets, circuits, and charges. However, very little research proved particularly relevant to the creation of the DEEM. Specifically, there is very little work addressing students' conceptions of the basic elements of electricity and magnetism, in particular field properties, electrostatic potential, and forces on charged particles. However, a few sources were available.

One of the most relevant works was presented at an AAPT Summer Meeting (Adrian and Fuller, 1997). These authors had been investigating college students' understanding of the electric and magnetic fields. They mentioned two results that were of particular interest: "flowing fields," and "field boundaries." The "flowing field concept," borne by many students, is in regards to the nature of the electrostatic field. Students think "something" actually flows through space, giving rise to a field. The other notion students carry is that an electrostatic field (and presumably a magnetostatic field), in otherwise empty space, has bounds. In other words, if one gets far enough away from a source of a field, say a proton, then the field will be equal to exactly zero. This investigator encountered both of these beliefs as well.

Many articles discuss students' conceptual understanding of how circuits work (Fredette and Lochhead, 1980; Cohen, Eylon, and Ganiel, 1982; and Eylon and Ganiel, 1989). However, even the most in-depth analysis that tries to link the so-called "macro-micro" relationships between electrostatics and electrodynamics, fails to uncover misconceptions at a fundamental level (Eylon and Ganiel, 1989; and Ganiel, 1988). It was not until much later that material was published concerning students' notions of the most basic concepts of E&M (Viennot and Rainson, 1992; and Rainson, Viennot, and Tranströmer, 1994). The authors of the first paper even state, "But the electric field itself, and more generally the notion of fields, has not been at the centre of any research on students' reasoning." The authors of this paper investigated students' misconceptions surrounding Gauss's Law and superposition of electric fields. Unfortunately, a major focus of the paper was the role of insulators, which is not a topic covered on the DEEM. One minor point they uncovered was that students would rely on the equation $\mathbf{F}=q_o\mathbf{E}$ to determine if there is a field at a particular location. Using this, students reasoned that if there was no test charge, q_o, then there could be no field. The investigator detected this line of reasoning in the interviews, as well.

A paper addressing students' misunderstandings of graphical representation of fields provided insight into possible pitfalls of representing fields on the DEEM (Törnkvist, Pettersson, and Tranströmer, 1993). Most of the problems students encountered were with somewhat complex field configurations. They also misunderstood the relationship between the diagram of the field and the path of a charged particle in that field. On the

DEEM, these problems were avoided because representations of fields were kept simple, most fields were homogeneous, and students were never asked to predict the path of a charged particle in a field.

Another study that focused on students' understanding of certain aspects of E&M lent insight into their problem-solving techniques when dealing with questions involving induction (Prosser, 1994). The study showed that students tended to use very superficial ideas and concepts to solve the problem of a magnet inducing a current in a loop of wire. Prosser generalized this finding to include students' problem-solving techniques in other areas of E&M as well.

One reason students may not correctly internalize concepts from E&M is they may not have properly internalized concepts from mechanics. Although questions on the DEEM were created with this in mind, some aspects of mechanics (forces and velocities, in particular) had to be used in order to make the test meaningful. Galili investigated the effect that the students' background in mechanics had on learning E&M (Galili, 1995). Although none of the specific examples discussed in his paper were relevant to this project, some of the conclusions were intriguing. Galili suggested that students may employ a two-stage process when conceptualizing fields: 1) *Creation*: first object $\rightarrow$ field; 2) *Action*: field $\rightarrow$ second object. This may lead students to physical contradictions, namely failing to apply Newton's Third law to what are truly action-reaction pairs. So, students' descriptions of the field itself may enhance their misconceptions.

A final paper, discovered by the investigator, was a case study of two students (Guth, 1995). This work revealed the students' ideas and misconceptions of the electric field and magnetic field. Guth determined that the two students who were chosen to represent the larger population did not have a clear and consistent field model. Some of the interesting results from this paper included the students' confusion between field strength and potential; ideas that field lines are "concrete entities;" difficulty in representing a field using vectors, not field lines; and lack of differentiation between electric and magnetic fields. This source proved very useful in the creation of the DEEM. It is unfortunate that it is the only one of its kind in the physics education literature.

13) Gender issues

This section will be divided into two parts. The first subsection will describe research on gender differences in science. The second subsection will present papers on gender differences on standardized multiple-choice exams. By reviewing the literature on gender differences, the investigator hoped to uncover underlying explanations for possible discrepancies between the genders on the exam

14) Gender Differences in Science Education

One study reported that males had a greater interest and experience with electricity (circuits) than females (Cambers and Andre, 1997). To determine how particular types of instruction affected achievement, the authors administered a thirty-question multiple-choice exam to a set of students. When factors of greater interest and pre-exposure were included as covariates in the analysis of exam scores, no gender differences in learning a physical science concept were evident. This suggests that classroom experience may play a major role in influencing achievement outcomes.

Two reports investigated gender differences in classroom discussions and how they related to student impressions of physics and learning (Alexopoulou and Driver, 1997;

and Guzzetti and Williams, 1996). Both studies concluded that males tended to conduct discussions in a confrontational manner, while females' discussion patterns centered on consensus. Alexopoulou and Driver concluded that these patterns had a noticeable effect on students' learning. Guzzetti and Williams' work was less quantified, but they concluded from their interviews with students that females in the class were "…restrained by their fear of the male students (not of the teacher) and of challenging the social norms that permeated the classroom interactions."

One factor that may be at least partially responsible for gender differences in science, in general, and physics, in particular, are the patterns of examples, questions, figures, and diagrams in textbooks. One study investigated the gender biases in physics textbooks (Walford, 1981). Walford found that a disproportionate amount of material in textbooks used males as the role models for pictures, questions, and in the text. Moreover, when males were depicted, they were often shown in more active roles – weightlifters, mechanics, and other workers – while females were shown in more passive roles – women looking in mirrors, a girl blowing a party whistle, and sunbathing. The male-chauvinistic trends in the photographs carried over to other aspects of the textbooks as well. Another author offered suggestions, in an opinion letter, about what types of examples could be avoided or rearranged to make them less "male" and more "female" (Lilley, 1986). She further suggested that textbook authors and instructors should include some problems that are strictly "female" to try to engage female students by making them feel that, "'Physics is part of your life, too.'"

15) Gender Differences on Multiple-Choice Exams

One report offered an overview of the research on the equality of multiple-choice for males and females (Hazel, Logan, and Gallagher, 1997). The authors found that many studies concluded multiple-choice exams favored male students, while exams that were more free form and involved human, social, or environmental emphasis, favored female students. They also found in their study that multiple-choice assessment techniques not only favored males, but the type of assessment had an impact on the students' desire to do well in the course and on the students' disposition to continue study in that field.

Two papers (Harding, 1979; and Thompson, 1979) offered a statistical analysis of the patterns of enrollment and test scores on public examinations in science (Harding) and physics (Thompson) in England. Both found a statistical difference between the results of male and female students on the multiple-choice sections of the exams, with males scoring higher.

To counter the notion that all standardized test are gender-biased, the ETS published a result that said, among other things, that the gap between males' and females' scores in science and mathematics has been narrowing over the past thirty years. That result was heavily criticized and many suggested that the large testing service had ulterior motives behind their study (Strosnider, 1997).

"The whole of science is nothing more than a refinement of everyday thinking."
– Albert Einstein

BASIC CONCEPTS OF ELECTRCITY AND MAGNETISM

An exam's concepts underlie the whole development process. A test designer must have a firm outline of what subject matter will be sampled. If one lacks a framework, the items may stray from their intended mark, the exam could become unnecessarily long, and the content validity and reliability may suffer, calling the entire test into question. For the DEEM, these concepts belong to the realm of electricity and magnetism as presented in widely used introductory, undergraduate physics textbooks (Halliday, Resnick, and Walker, 1997; Fishbane, 1996; Serway, 1994). The investigator excluded some topics because they have been used in different diagnostic exams or investigated previously in some other manner, and others because they were considered relatively unimportant or unsuitable for a multiple-choice exam format.

The intent of the DEEM is to probe students' understanding of the most basic and robust concepts of electricity and magnetism. Before one can begin to form a list of these concepts, a working definition of "basic" and "robust" must be formulated. The basic concepts of E&M are ones that form the foundation for other electromagnetic phenomena and are typically described, in introductory classes, as important unto themselves. An example of this is Gauss' Law. A robust E&M concept is one which is not strictly fundamental but which receives much attention in introductory E&M because of its usefulness. An example of this is induced currents.

A logical place to begin a discussion of the basic concepts of E&M might include the Lorentz Force Law, electrostatic potential and potential energy, and Maxwell's Equations. Given the exam's intended audience, the first two items in the list lend themselves well to questions that would not tax students' mathematical abilities; conceptually though, these can prove to be just as challenging as any concept in physics. The third item on the list, Maxwell's Equations, if employed in their integral or differential form, can prove challenging to introductory students. As a result, widely used textbooks approach Maxwell's Equations by exploring high-symmetry examples in which the integration is of a constant field along an uncomplicated path (or over an uncomplicated area) and the differentiation is of simple, time-dependent fields. Students taking the exam will not be required to explicitly integrate or differentiate. Implicit integration of Maxwell's Equations will be employed for closed surfaces where the direction of the field is provided. Students will also be expected to add vectors and scalar quantities.

Sections 6) through 10) will outline basic concepts of E&M covered on the DEEM. Furthermore, these sections will set boundaries. Section 11) will discuss topics omitted from the exam.

16) Lorentz Force Law

The general form of the Lorentz Force Law for a test charge, *q*, in an external, static electromagnetic field is given by:

$$\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}. \qquad (4:1)$$

where *F* is the force on a charge, *q*, moving with velocity, v, through a static electric and magnetic field, *E* and *B* respectively. Textbooks do not generally present this equation as a whole until each of its components (the electrostatic force and the magnetostatic force) are introduced separately. Even when the electric and magnetic forces have been addressed, little emphasis is placed on the general form that combines the two components. Consequently, the electrostatic force and magnetostatic force appear separately on the exam. Before delving into a description of forces on charged particles, the next paragraph will briefly describe field lines.

The general convention for field lines is used in the exam. Electric field lines begin at sources (positive charges) and end at sinks (negative charges). Arrows on the electric field lines point from positive charges to negative charges. Magnetic lines of force do not begin or end anywhere and the arrows along the lines point from the north pole of a magnet to the south pole of a magnet. Alternately, the direction of the magnetic field lines, from a wire carrying a current, is found from a “modified right-hand rule.” Application of this modified right-hand rule is as follows: point your right thumb in the direction of the current; your fingers curl around to describe the shape of the magnetic field lines (circles), and they point in the direction of the magnetic field vector. This second convention is the one that is explicitly employed in the exam. Properties of magnets, in the sense of a bar magnet, are not covered on the exam.

17) Electrostatic Forces

The general Lorentz Force Law for a test charge in an external, static electric field is given by:

$$\vec{F} = q\vec{E} \,, \qquad (4:2)$$

where q is the charge on the particle, and E is the external electric field in which the particle is immersed. This equation implies that the force on a particle due to an electric field is either parallel or anti-parallel (as determined by the sign on the charge) to the vector that describes the electric field at the point where the particle is located.

18) Magnetostatic Forces

The general expression for a particle in an external, static magnetic field is given by:

$$\vec{F} = q\vec{v} \times \vec{B}, \tag{4:3}$$

where *q* is the charge on the particle, *v* is the velocity of the particle, and *B* is an external, static magnetic field in which the particle is immersed. This equation implies that the force on a particle due to a magnetic field is perpendicular to both the particle's velocity vector and the magnetic field vector at the point where the particle is located. The conventional right-hand rule determines the direction of the force.

19) Properties of Electric and Magnetic Fields

The electric and magnetic fields both share the vector property that at any point in space the resultant field, due to multiple sources or an extended source, is the sum of all field vectors due to each source, or an integral over an extended source. These fields also share the property that as the distance between a source (say a point charge or a long straight wire with a constant current) and a point of interest increases, the magnitude of the field strength decreases. The details of the decrease in the magnitude are different for the sources of electric and magnetic fields employed on the DEEM. For a point charge, the magnitude of the electric field falls off as $1/r^2$; for a long, straight wire carrying a steady current, the magnitude of the magnetic field falls off as 1/r.

Origins of these two fields are quite different. The origin of an electrostatic field is a point charge. The direction of the field is radially outward or inward for positively charged or negatively charged particles, respectively. The origin of a magnetostatic field is moving charge. On the DEEM, all moving charge is confined to a wire (a.k.a. current). The direction of the field is determined by using the modified right-hand rule as discussed in section 2)

On the DEEM, students must relate how the magnitude of a field changes as one moves farther and farther from a point charge or a current carrying wire. They must also determine the direction and relative magnitudes of fields due to a number of sources.

20) Electrostatic Potential and Potential Energy

Potential and potential energy are not necessarily fundamental concepts of electricity and magnetism. The concept of electrostatic potential energy is drawn from the general definition of potential energy as it relates to force:

$$\Delta U(r) = -\int_{r_1}^{r_2} \vec{F}(r) \cdot d\vec{r} \ . \qquad (4:4)$$

where ΔU is the change of the potential energy of a particle moving along a path, under the influence of a conservative force, *F*. One could argue that the force should be derived from the potential, and not vice versa. Considering the level of this exam, that debate is somewhat academic and will not be pursued. The point is that, although potential and potential energy are important topics presented in undergraduate E&M, they do not solely "belong" to this realm of physics.

In electrostatics, there is a general assumption that the potential of an isolated point charge tends to zero as one moves out towards infinity. This assumption was mentioned on the exam before the items where it might be useful. The next two sub-sections will explore the ways in which electrostatic potential and potential energy are typically presented in an introductory class and how they were included in the exam.

21) Electrostatic Potential

The electrostatic potential is a scalar field defined at a point in space, given a distribution of charge at some other location(s) in space. The general expression for the potential due to a system of *N* point charges is:

$$V(r) = \sum_{i=1}^{N} \frac{1}{4\pi\varepsilon_0} \frac{q_i}{r_i}, \tag{4:5}$$

where r_i is the distance from point charge, q_i, to the location where one is calculating the potential. The options on the DEEM do not require students to know the value of the constant ($k_e=1/4\pi\varepsilon_o$) in front of the equation, nor are they expected to make an explicit calculation of the potential. The exam does demand students to determine the sign of the potential for a single point charge or a system of point charges at specific locations in space. It also expects students to determine where the potential for a given set of charges would be zero.

22) Electrostatic Potential Energy

The electrostatic potential energy is a scalar quantity defined for a system of two or more particles. The general expression for the electrostatic potential energy of a configuration of *N* point charges is:

$$U(r) = \sum_{j=2}^{N} \sum_{i=1}^{j-1} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}, \tag{4:6}$$

where q_i and q_j are two different point charges and r_{ij} is the distance between them. The DEEM asks students to infer the potential energy of systems containing two charges of equal magnitude. Again, the exam does not require students to know the value of the constant in front of the equation, nor are they required to make an explicit calculation of the potential.

23) Induced Currents and Lenz’s Law

One topic that is often covered in an introductory course, that is not a basic concept of electromagnetism, is induced current. A current is induced in a wire loop in response to the time rate of change of the flux of a magnetic field through a loop. Lenz’s Law adds that the direction of the induced current in the loop is such that it creates a magnetic field which opposes the change of the external magnetic flux. This is obviously an application of Faraday’s Law; however, it is a topic which many textbooks devote a great deal of attention. Therefore, items relating to this topic are included on the exam.

In an introductory course, students typically learn that the magnetic flux through a loop can increase or decrease with respect to time in many ways. The loop can move into or out of the field, the loop can rotate in the field, the area of the loop can change, and the magnetic field strength can change. Many of these methods for varying the magnetic flux through a loop are given in exam questions; however, only one method is present in any one question. Students are never asked to explicitly calculate a time derivative.

24) Maxwell’s Equations

A “true” diagnostic exam testing the basic concepts of E&M would focus heavily on Maxwell’s Equations, with the Lorentz Force Law included to describe the motion of particles. For pedagogical and historical reasons, this is not the way E&M is presented in introductory physics classes. Still, no diagnostic exam which tests electricity and magnetism would be complete without questions that probe students’ understanding of these basic equations. Figure 4–I sums-up Maxwell’s Equations in integral form.

Figure 4–I: Maxwell's Equations

$$\oint \vec{E} \cdot d\vec{a} = \frac{q_{enc}}{\varepsilon_o}$$

$$\oint \vec{B} \cdot d\vec{a} = 0$$

$$\oint \vec{E} \cdot ds = -\frac{d\Phi_B}{dt}$$

$$\oint \vec{B} \cdot ds = \mu_o i_{enc} + \mu_o \varepsilon_o \frac{d\Phi_E}{dt}$$

The first equation in the group, Gauss' Law, describes the relationship between electric field and charges. Section 5) already addressed this relationship. The second equation, sometimes referred to as Gauss' Law for Magnetism, essentially states that magnetic field lines must form closed loops – they have no beginning or end. In other words, magnetic monopoles do not exist. The third equation, Faraday's Law, states that a magnetic field that varies with time will produce an electric field. This equation finds two places in introductory E&M, first, as discussed in section 9), as an application in the form of induced currents, and second, as a description of how to produce an electric field explicitly in terms of a magnetic field and not explicitly in terms of sources. The final equation in the list, Ampere's Law, also has two places in an introductory class. First, it describes the origin of magnetic fields. Second, like Faraday's Law, it portrays how a magnetic field can be created via a time-varying electric field, independent of any explicitly mentioned currents.

25) Discussion of Omitted Concepts

It should be clear to any introductory physics instructor that many of the concepts normally discussed in an introductory electricity and magnetism class are not included on the DEEM. Some of the most glaring omissions are: circuits and circuit elements, properties of insulators and conductors, inductance, and electromagnetic waves. The investigator must supply some rationale as to why these standard topics are absent from the DEEM.

Circuits and circuit elements have been widely addressed in the literature and many diagnostic tools have been created to test students' knowledge of them. Also, the investigator does not consider circuits and circuit elements to be basic concepts of E&M. This is not to suggest these topics are unimportant. They simply are not appropriate in light of the purpose of this exam.

The investigator never considered including the topics of inductance and electromagnetic waves for several reasons. First, EM waves may not be covered in the same course as the topics outlined above. Second, few texts devote much attention to describing inductance and instructors may omit this topic from their class. Finally, the list of topics was already quite long, and, as was stated before, not everything can be included.

The properties of insulators and conductors were to be included in the exam originally. These two topics are a common opening for a typical E&M class. After the investigator conducted interviews with students, however, it became apparent these two topics were not appropriate for a multiple-choice exam. Students' ideas of the properties of insulators and conductors are hard to encapsulate in four or five options. Consequently, the investigator dropped these topics from inclusion on the exam.

The best interviews – like the best biographies – should sing the strangeness and variety of the human race.” – Lynn Barber

INTERVIEWS

The interview is a useful tool for uncovering the knowledge of another human being. In an ideal setting, when the interviewer poses a simple, straightforward question the interviewee candidly responds with what he or she knows. Unfortunately, few things in life are ideal. Questions can be loaded with leading cues, demand personal information, or cry out for a particular response. This chapter will attend to these and other issues.

The first section will detail the various purposes of the interviews. The next sections will describe the history of the interviews and the process by which volunteers were selected. The section following these will illustrate the process of creating the interview questions. From there, the next few sections will present interview protocols and details, and the chapter will close with sections devoted to transcription, revelations about exam material inspired by the interviews, and some final thoughts.

26) Purposes

The interviews served several purposes simultaneously. First, they trained the investigator in the art of writing effective interview questions, a skill later transferred to writing exam questions. Second, the interviews provided a testing ground for questions. Here, the test creator could detect if a question’s wording sent students down the wrong conceptual path, or led them directly to the answer. Third, the interviews provided a wealth of distractors to incorporate into the exam as attractive alternatives to the correct responses; many of these distractors were foreseeable, others were not. Fourth, the interviews established a baseline of physical constructs that students understood. This knowledge bestowed confidence on the investigator for making certain assumptions while writing exam stems and options. For example, the interviews enabled the investigator to explore the students' conception of the word "particle." A question like, "What is meant by the word *particle*?" would not find its way into the exam; however, the word "particle" appears in about half the exam's questions. It is important for the exam's designer to know what a typical test-taker is thinking when that person sees the word "particle." Finally, the interviews provided a means to narrow the number of topics to be covered on the exam.

27) Interview History

The first set of interviews was conducted durning the Summer semester of 1995. The investigator recruited six people from a small pool of students. Starting out with a small group proved useful. The investigator was able to hone interviewing skills and learn from mistakes on only a few people.

The second set of interviews occurred in the Fall semester of 1995. In contrast to the Summer semester, a large number of students were enrolled in Physics II. This provided an ample pool from which to recruit volunteers. Twenty-six people volunteered to be interviewed. The final set of interviews transpired in the Spring semester of 1996. Again, this semester provided a large pool of students, from which twenty-one people volunteered to be interviewed.

28) Obtaining Volunteers

For the investigator to recruit student volunteers, it was necessary to enter their recitation classes and request their help. The investigator's choice of which recitation sections would be to attended, to ask for volunteers, was dictated only by timing. The investigator strove to attend as many recitation classes as possible in the first days of the semester.

Once in a class, the investigator began with a short, prepared speech to inform students about the interviewing process. The investigator made it very clear from the start that their responses would not be reported to their Physics II instructors and their work would not count against, or for, their class grade. Details such as time commitments, the relationship between the lecture material and interview schedules, and pay were clearly presented. The investigator informed the students they must fill-out a W-2 form (if they had not already done so for work-study, an Undergraduate Research Project, etc.) as they would need to become employees of Rensselaer Polytechnic Institute in order to get paid $5/hour for their time. The investigator also remarked that their identity would remain anonymous and their responses would be transcribed from audio tape into an electronic format and saved in a password protected file under a coded filename. As a final annotation, the investigator added that although the interviews would be for the most part "one-way", volunteers could gain something (besides cash) by being confronted with the physics they had learned in class before they had to start studying for the next quiz or major exam.

After the information was delivered, the investigator answered any questions the students had. When their questions were satisfied, a sign-up sheet to be passed around the class was left behind. Students could only sign-up for available times on a first-come first-served basis. The sheet had a place for name, e-mail address, and telephone number. If there were no available times to fit a student's schedule, the student could simply write his or her name, phone number, and e-mail address on the back of the sign-up sheet. Attempts by the investigator would be made to try to accommodate them. After class, the investigator returned to pick up the sheet.

29) Creating Interview Questions and Diagrams

Before the interviews began, the topics that would be covered on the exam had been generally established. These were the starting points for the interview questions. Writing interview questions is not easy; it is a skill that must be honed. Predictably, since the interview is such a strong research tool there are many references for writing interview questions.

One source put forth seven questions to be considered when creating interview questions (Kerlinger, 1986). The investigator weighed this list when designing interview questions (as well as diagrams) and later when writing exam questions. The last two questions on the list (see below) are not relevant to this work, but for completeness they are all mentioned and, at least briefly, addressed.

1. Is the question related to the research problem and the research objective?
2. Is the type of question appropriate?
3. Is the item clear and unambiguous?
4. Is the question a leading question?
5. Does the question demand knowledge and information that the respondent does not have?
6. Does the question demand personal or delicate material the respondent may resist?

7. Is the question loaded with social desirability?

Question one requires little explanation. If a question was not directly, or indirectly, related to one of the concepts which were to be included in the exam, it was simply not posed.

Question two opens the debate of how to best obtain the desired information. During the first set of interviews (summer of 1995), all of the interview questions and answers were in the form of an oral exchange. Sometimes the interviewee answered oral questions, other times that person answered questions about specific diagrams. For the second and third round of interviews, some of the questions required students to draw arrows (indicating a direction) on a preprinted sheet. This method of data collection proved to be much more efficient on some types of questions, and it permitted students to express themselves in another manner. Occasionally, this procedure led to new insight about students' misconceptions.

Question three addresses the issue of creating clear and unambiguous interview questions. One way for a question writer to avoid writing ambiguous questions is to be concerned about so-called double-barreled questions – questions that ask two things at once (Kerlinger, 1986). Furthermore, a question's meaning remains apparent when it is not open to alternate interpretations. One faces a challenging task trying to avoid these two pitfalls. An advantage of writing questions concerning introductory physics is that once the parameters of the problems have been carefully laid out, there is little room for alternate interpretations. (The difficult part is not overlooking any these parameters.) A disadvantage of writing questions concerning introductory physics is the abundance of incorrectly used scientific terms in everyday language. Again, when one is aware of these problems, they can be circumvented with well crafted questions. Question five is closely related to this particular concern and some the ideas presented here will be discussed in more detail below.

Question four raises the matter about leading items. In the context of general interview questions, it is easy for an interviewer to mistakenly ask a leading question. “Have you read about the local school situation?” will draw many “Yes” responses because it is socially desirable to say you know or have read about particular issues (Kerlinger, 1986). An interviewer is somewhat relieved of worrying about this issue when writing questions for an interview that explores a person’s knowledge about scientific facts. The only type of question that could conceivably lead the interviewee to an answer is one so narrowly worded that only one response makes sense, or one containing the right answer (or some reasonable wrong answer) in the text of the question itself. Examples of the above two types are: “The force on a positively charged particle is perpendicular to the magnetic field and the velocity of the particle. In this diagram what is the direction of the force on the particle moving in this magnetic field?” (with an appropriate diagram shown); and “The force on a positively charged particle in an electric field is in the direction of the electric field, is it not?” In the first question, the number of choices for the force is reduced from infinitely many to two reasonable choices. In the second one, the student is prompted for the correct answer by the way the question is worded. In addition, neither question leaves open the possibility that there is no force on the particle.

To prevent these problems, the investigator wrote interview questions that were as open as possible. The two examples from the previous paragraph could be recast as “What is the direction of the force, if any, on the charged particle in the magnetic field in this diagram?” (with the appropriate diagram shown). And “What is the direction of the force, if any, on a positively charged particle in an electric field?" The problem with this type of question is that an interviewee can offer an infinite number of responses. This extreme possibility is, obviously, never realized; most students answer well within a range. However, the "openness" of these questions allows students' misconceptions to manifest themselves much more easily.

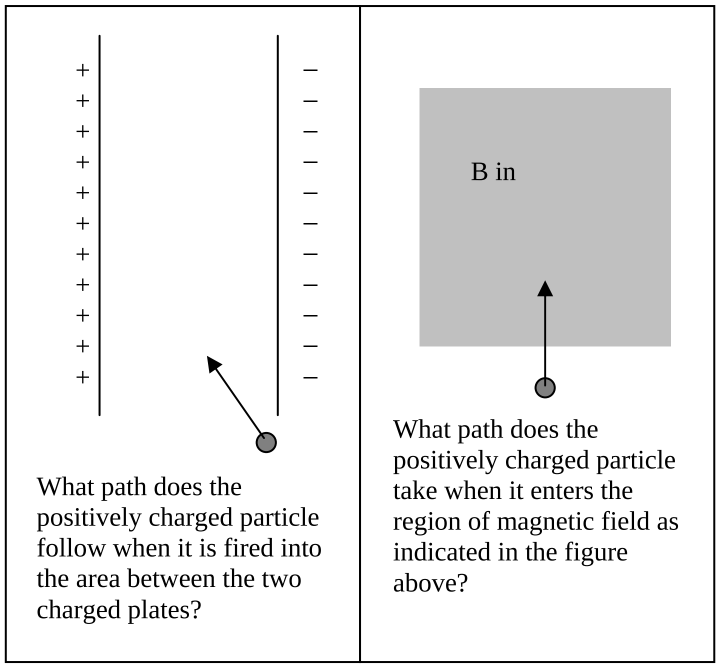
The investigator added another item to the list of questions to be considered when creating interview questions.

4a. Is the *set* of questions leading?

In other words, can the students deduce a particular pattern of response in a series of questions? For example, if the response "There is no force," is never a reasonable option, then students may avoid it, not because they know it is the wrong answer, but rather because it was never an answer in the past. Trying to guess students' patterns can be difficult, and dangerous. If one includes a small number of questions that have as a correct response "There is no force," then some of the students may be lead to some false negative responses. These students know the right answer but are set in a pattern. What is important is to alternate the arrangement of the keys to prevent students from falling into patterns and to keep them thinking about each question separately.

A major concern that arises from considering question five is that one does not want to create questions that require an understanding of E&M concepts, other than the one being tested, or any other concept of physics for that matter. Figure 5– illustrates two short examples of questions that require not only knowledge of a specific E&M concept but also of other concepts as well. The first question assumes the student understands that a charged particle in a uniform electric field (set-up by the charged plates), traveling in two-dimensions, will follow a parabolic path, or that the student can quickly use kinematics equations to derive that fact. The second question assumes that the student can not only grasp the direction of the force of a charged particle moving in a magnetic field, but also knows that a force perpendicular to the velocity will result in circular motion.

Figure 5–I: Examples of Poor Multiple-Choice Questions



It should be noted that these are not necessarily “bad” problems for teaching E&M. All of these problems can be found in standard elementary physics textbooks (see, for example Halliday, et. al., 1997 or Serway, 1994). When an instructor uses them properly, these types of question can provide substantial insight into students’ understanding of physics. They are, however, weak candidates for multiple-choice questions, and as such they, and other questions like them, were not included in the interviews. Beyond these considerations, a question writer must confront a deeper problem – how well does *any* one question probe a single concept?

An interview- or test-question writer who invokes words and phrases such as "positively charged particle," "force," or "uniform electric field" in a question such as, "What is the direction of the force on the positively charged particle that is at rest in an uniform electric field?" must consider if the students know the meaning of all of these words and phrases. Or, more precisely, does the students' understanding match that of an expert? Communication of any sort, whether it is a dialogue between colleagues or a question that will be read by students on a multiple-choice exam, assumes a basic knowledge set for the parties involved. The boundaries of this set are nearly static and well understood in a dialogue between colleagues. Therefore, there is less potential for miscommunication. Unfortunately, one realizes, while creating interview questions and conducting interviews, that the boundaries of this knowledge set for the question writer and the interviewees are dynamic and are not well understood.

The investigator established a baseline of students' E&M knowledge set by posing simple, straightforward interview questions. These were intended to probe a particular idea or construct. For example, questions such as, "What does the word 'particle' mean?", "What is 'neutral'?", or "What is a 'field'?" were asked prior to these words being invoked in other questions. These inquiries established what students had in mind when they heard particular words or phrases and how their ideas might shape their subsequent answers.

Students' responses to these questions were, for the most part, predictable. So much so, they seemed to have simply memorized definitions without incorporating limitations or assumptions. This is most unfortunate from a broader pedagogical point of view, but this "textbook terminology" did suffice for answering questions in the interviews, and on the exam; consequently, a deep investigation of these problems was not undertaken.

Questions six and seven (from the list at the beginning of this chapter) did not apply when considering questions for interviews of this sort. Electricity and magnetism is nearly featureless on the relief map of social norms. However, as a matter of common decency and at the expense of scientific objectivity, if a student seemed to get "tied-up" and frustrated with a question because of a complete lack of knowledge of the subject matter or an inability to express it, then that person was not forced to continue. That question was then dropped from that particular interview and/or the interviewer would call for a brief pause until the student was ready to proceed. The objective of these interviews was to provide data on students' understanding of electricity and magnetism, not to dehumanize the subjects.

30) General Interview Procedures

The first interview meeting was short. The student came in, introductions were made, and the interviewer explained again how the interviews would proceed and reassured the subject that what transpired would remain confidential. The investigator also explained to the student that specific forms of identification would be needed to become an employee of Rensselaer Polytechnic Institute. Since the monies were derived from a grant, this was the only way they would be paid. When they brought in the proper identification, they were directed to fill in the appropriate forms.

On the first day, the interviewer asked the students to call or e-mail in advance if they could not make their regularly scheduled thirty-minute session so a new time could be rescheduled. One possibility was for the student to stay for a long session the next week to catch up with the rest of the interviews.

The day-to-day details of the interview will be discussed in the proceeding section. As the semester progressed, student participation dwindled. If a student missed two interviews in a row, then the investigator called that person to find out if he or she wanted to continue with the interviews. Quite frequently, this prompt was sufficient to get some students back to the interviews. But more often than not, they never came back. To give students a break, there were no interviews the week of an exam. This went over well with the students, and it offered the investigator time to catch up.

As the semester ended, the students were reminded to get all of their paperwork in order so they would receive payment. They were also asked to sign a form on the last day that stated they were in agreement as to how many hours they had come in for interviews.

31) Interview Details

The concepts covered in the interviews were from material the students had been exposed to the previous week in class. The interviews could not cover the most recent topics because students who came early in the week would not have seen or read it. It was the interviewer's decision to conduct the same interviews in one workweek.

Each interview had a similar pattern. The investigator would read to the student a question, displayed on a computer monitor, and then wait for a response. The interviewee had an unlimited time to answer the question. If more information was desired, the investigator would ask the student, a simple, non-leading prod such as, "Could you explain in more detail what you just said?" to elicit more explanation. Sometimes the student would say that there was nothing more to add or could not explain what was just said. The investigator would simply move on to the next question without further comment.

This type of interviewing question is known as open-ended, and it is set in contrast to fixed-alternative. Open-ended items set a framework for respondents' answers and then let the interviewee respond at will. This format permits the researcher to obtain a clearer picture of what the interviewee understands, or thinks, concerning a particular concept. There is a lot of flexibility and latitude with this line of questioning; however, there are some disadvantages. Open-ended questions leave room for interpretation by the interviewer, and they present the possibility that the interviewer could lead the interviewee to the correct answer. Nevertheless, if one writes strong questions and is mindful of the interaction between the subject and the interviewer, these problems can be minimized. Fixed-alternative items are essentially multiple-choice questions designed for an interview. One of the shortcomings of this method is that the interviewer must try to guess what the most likely responses will be. This is particularly limiting when one is trying to obtain data to create distractors for a multiple-choice exam. On the other hand, fixed-alternatives remove subjectivity and are easier to use (Kerlinger, 1986).

As mentioned previously, the investigator read the questions from a computer monitor. The interviewer kept the current question at the bottom of the screen so, if possible, the student could not take cues from the next question. Most students just sat and listened to the questions as opposed to straining to read it off the monitor. Many of the questions had diagrams associated with them. If that was the case, then diagrams were printed on an 8.5” x 11” sheet of paper before the interview, or the investigator asked the student to look at the diagram on the monitor. In either case, the students gave comments on question(s) that related to the diagram.

The investigator avoided head nodding, or other body language cues to prevent the students from knowing they were “on the right track.” When the investigator felt that the student had submitted as much as that person had to offer, or if there was enough insight into what the student was thinking, the investigator would simply say, “O.K.” and move on. If the student indicated that more than a few seconds was needed to formulate an answer to the question, then the investigator would turn off the tape recorder; when the student was ready to answer, the investigator would turn the recorder back on.

32) Transcription

The investigator’s role as transcriber was to record, on paper, all that was said in the interview. This was the most time consuming part of the interview process; however, it was rather straightforward. The investigator would bring up the template of questions used in the interview and carefully transcribe all that was said (by both the investigator and the interviewee) into electronic format. The investigator even noted long pauses in the transcription.

The investigator would, from time-to-time, add questions to the interview or the student would ask questions in an attempt to elaborate on what was said. These were also included in the transcription. The student’s response to a newly created question was also transcribed. If a short, relevant dialogue ensued, then it was also added. Questions repeatedly added by the investigator to the interviews were simply added to the template, thus saving time on transcription.

As promised, once an interview's transcription was complete, the investigator erased the tape to protect the identity of the interviewee.

33) Topics Removed from the Exam as a Result of the Interviews

As a consequence of the interviews, the investigator removed two broad areas of introductory electricity and magnetism from the exam content: graphical representation of vector fields, and concepts dealing with conductors and insulators. Both of these areas proved to be extensive and the students' understanding too vague to merit including them on the DEEM.

Graphical representation of fields provided too many problems for the students. Particularly, the students had difficulty understanding what was asked and what they were expected to draw. The students' most common problem was drawing a representation of the vector field of a single point charge or a current-carrying wire. Instead of drawing sets of vectors at specific locations in space, the students inevitably fell back on drawing lines heading out from positive charges and into negative charges. There was no difference in their minds between a diagram of a vector field and so-called "lines of force." When a question did ask students to draw lines of force, they most often fell back on the simplest rules and diagrams. Some students remembered a few steps that should be employed when drawing these diagrams, and others just drew whatever seemed correct. Very few diagrams had all of the aspects of properly drawn field lines or vectors. It is interesting to note that one of the students' biggest problems arose when asked to draw electric field lines for two particles – one charged and the other neutral. Only a small number students knew how to approach drawing a diagram for this scenario.

Interview subjects drawing the scalar electrostatic potential lines were not quite as challenged. This stemmed from the fact that there can be no misunderstanding on the students' behalf whether a question actually means potential lines or perhaps "vector lines" (which are meaningless when one is talking about a scalar potential). Once it was clear to the investigator that including four or five diagrams of field lines on the exam and having the students choose the one that corresponded to the description was no longer viable, using problems with electrostatic potential lines seemed pointless.

A consequence of the investigator pondering the use of complex diagrams in the interviews, and on the test, was the triggering of a small epiphany, of sorts, regarding the type of questions that should be considered for the exam. Any question best answered by having students draw diagrams that are even just one-step beyond a simple arrow are best left off a multiple-choice diagnostic exam. This is not to say these questions are not valuable diagnostic tools, it is simply that there are too many outcomes and too much insight into the students' understanding is lost if they are given only a small handful of choices from which to make a selection.

The other area of unacceptable exam topics was conductors and insulators. Again, this domain proved too vast and the students' misconceptions just as wide. In particular, interview questions could not overcome two major obstacles. First, the students did not know exactly what the words "conductor" and "insulator" meant. Most students gave their short, "textbook" definition that conductors conduct electricity and insulators do not. With this description in mind, they were unable to explain how one could get a net charge on an insulator or how an insulator with no net charge would be effected in an external electric field. Conductors proved to be less challenging, but students still had problems. The most common barriers included not being able to properly describe how charges distributed themselves on conductors and how the charges on a conductor rearrange themselves in response to an external electric field, particularly in a neutral conductor.

Enough evidence had accumulated to warrant the omission of these topics from the exam. As was stated before, not all concepts can be covered on a single exam. Other investigators have previously covered some of these topics before; this investigator covered some different ones here; and there will still be some topics left for investigators to cover in future diagnostics.

34) Final Notes Regarding the Interviews

Interviewing is a very powerful technique for gleaning students' understandings and misunderstandings of electricity and magnetism. And although this was very demanding, it was also time well spent. Interestingly enough, after the semester many of the students involved were interested in what became of their responses and "How well they did."

Armed with all of the data acquired from this stage, a test creator is prepared to move onto the next stage of creation of a diagnostic exam – writing the test stems and options. But, before discussing those, the next chapter will delve into issues of reliability and validity. One can write strong items only with a firm understanding of these two, important topics.

RELIABILTY AND VALIDITY

Reliability is a quantitative measure of a test's ability to duplicate its own assessment of a particular subject. Validity has many forms. In relation to the development of the DEEM, validity implies a subjective measure of the content of the diagnostic tool. This chapter will discuss reliability and techniques for improving test reliability through better item construction. Chapters 10 and 11 will present statistical tests for reliability. This chapter will also discuss validity theory and approaches for determining validity. Chapter 9 will review expert analysis of the exam's content as a way of assessing the exam's validity.

35) Reliability

For an exam to be a meaningful instrument it must be consistent, or reliable (Baron, 1998). In other words, the score one receives on one trial of the exam should closely match a re-test score on the same test, assuming no knowledge could be gained simply by taking the exam. The theory of reliability assumes a person's measured score (or observed score) is the sum of that person's *true score* and the error in measurement. The true score is the score a test would return if it were possible for the instrument to absolutely assess a person's knowledge regarding a particular subject. In classical test theory, the variance (the standard deviation squared) in observed scores is the sum of the variance of the true scores and the variance of unsystematic errors:

$$\sigma_{obs}^2 = \sigma_{true}^2 + \sigma_{err}^2. \qquad (6:1)$$

The reliability, *r*, of the test is defined as:

$$r=\sigma_{true}^2/\sigma_{obs}^2 \qquad (6:2)$$

(Aiken, 1976). Although one can never measure the reliability directly; there are many techniques to estimate the reliability of a test, or test items. Also, a test designer's inability to directly measure reliability does not prevent that person from knowingly increasing it. By combining equation 6:1 and equation 6:2, one arrives at an expression for reliability in terms of the variance in the error scores and the observed scores:

$$r=1-\sigma_{err}^2/\sigma_{obs}^2. \qquad (6:3)$$

Equation 6:3 suggests that if the observed scores' variance increases, then the reliability increases, assuming the error scores' variance does not change (Aiken, 1976). For a dichotomously scored exam, the observed score's variance can be as large as 0.25. (Recall, the variance is the square of the standard deviation, and the maximum standard deviation for a question graded right or wrong is 0.5.)

There are a number of ways for a test designer to estimate the reliability of an exam. Coefficient alpha is one of the most widely used tools. Coefficient alpha is generally considered a measure of the internal consistency of an exam. The following paragraphs will define internal consistency and how it relates to reliability and Chapter 10 will contain a detailed description of how to calculate coefficient alpha.

A test designer can estimate the reliability of an exam using a number of techniques. Two of the most common techniques are a test-retest of the exam and a split-half test. A test-retest requires the exam to be administered to a pool of examinees and then the same test is readministered to the same group a short time later. The average correlation between scores of examinees is a most basic measure of reliability. If re-testing is not possible, or desirable, the designer can evaluate the reliability using a technique known as the *split-half test*. The split-half test is the correlation between two equivalent halves of an exam given to the same test pool during the same test administration. A potential problem a test designer may encounter when performing a split-half evaluation is there are many ways to divide an exam in half. Cronbach solved this problem by introducing coefficient alpha, which is the average of all possible spit-halves (Cronbach, 1951). This feature earns coefficient alpha the claim of being a measure of the internal consistency of an exam. But, coefficient alpha can not determine the reliability of all types of exams. The following paragraph will describe what type of exam is an appropriate subject for a calculation of internal consistency using coefficient alpha.

Test reliability can be affected by the speed at which an examinee is intended to take the test. A *speed* test is one where people are expected to get every item they encounter correct, but, because of a limited time most people will not finish. A *power* test is one that everyone is intended to finish, but, because of the difficulty of the items many people will get a number of problems incorrect. Most tests are neither power tests nor speed tests (Traub and Hambleton, 1972). If a test is predominately a test of speed, then it is inappropriate to use methods of internal consistency to measure the reliability of the exam (Gregory, 1992). If a test is judged to be a power test, then measuring the internal consistency can determine the test’s reliabilty. The investigator constructed the DEEM to be a power exam and Chapter 9 will provide evidence that it is. Therefore, it is appropriate to use the measure of internal consistency to determine the reliability of the exam.

One way a test designer can increase the reliability is to increase the test length. This is simply because, in general, the longer the exam, the closer the observed score is to the true score. To estimate the reliability of a longer exam, one can invoke the Spearman-Brown Prophecy formula (Merhrens and Lehmann, 1984):

$$r_{new} = \frac{Kr}{1+(K-1)r}, \tag{6:4}$$

where K is the ratio of the number of exam items on the new exam to the number of items on the old exam. This equation suggests that very short tests, or sub-tests derived from a longer test, are less reliable than longer tests.

Another way a test designer can affect the overall reliability of the exam is by maximizing the true score variance through the appropriate choice of item difficulty. To maximize variance, the item difficulties should be nearly equal and of medium value (Gulliksen, 1945). For a test creator to calculate what the most appropriate value is, one needs to remember that students can get the correct answer by guessing. Chapter 11 will provide more details on determining appropriate level of item difficulty.

The number of options offered on an item can also influence reliability. In one study it was determined that three options were found to produce the most reliable results as long as the total number of options was the same across the whole exam (Lord, 1977). A test creator's decision to determine the optimal number of options is settled by the appropriate number of distractors. If extra, non-plausible, distractors are included in an item's options, then they may lead to the ambiguity of the item, decreasing its reliability (Oosterhof, 1996).

The use of the phrases "all of the above" or "none of the above" are often used in creating multiple-choice options to increase the number of options. These will often lower the reliability of the exam. "None of the above" is used to prevent students from realizing there is an inconsistency in their logic. Students usually have trouble, however, determining if a wrong answer is always and absolutely wrong, so “none of the above” could lead to inconsistencies in the students’ mind, driving even the weakest students away from that choice. One way to avoid this is to design test questions that use options that always and absolutely eliminate any of the other options from being correct (Oosterhof, 1996). A list of options that offer the directions of a force or a field is a case where it is clear there is only one choice. For example, for a given situation, either the force is right or it is left, certainly not both.

The option “all of the above” also can reduce the reliability of a test. It tends to have the effect of reducing the number of options down to two, thereby removing other plausible distractors. Similarly, if one of the options is clearly incorrect then the test taker immediately removes “all of the above” as a possible answer (Oosterhof, 1996).

There were items on the DEEM that utilized options such as “all of the above” and “none of these.” But, these options proved to upset the patterns of responses and did not provide any additional insight. Consequently, they were removed from the latest version of the DEEM.

The subjectivity of a score will also lower the reliability. Since multiple-choice exams are graded electronically, one would expect the highest level of objectivity in grading. So, grading will have no effect on the reliability of the DEEM.

Group homogeneity will also influence the reliability of the exam. The more diverse the group, the more reliable the exam. This is simply a result of the increase in the variance of the observed scores. This issue is only a concern when the test is reported to have a particular reliability and then is administered to a group that is more homogenous than the intended audience.

Following all of this discussion of reliability, one must ask, “How reliable should the test be in order to be useful?” The answer depends on how the test is going to be used. If the exam will be used to make critical decisions, then one would require a very high reliability. If the decisions are going to be made about an individual as opposed to a group, then, again, the reliability should be high. Since the DEEM is not intended to be used to make critical decisions about individuals and will most likely be used to track changes in large groups of students, its reliability is not a critical matter. However, improving the reliability only strengthens the test and makes it more attractive to potential users.

36) Validity

Validity is a measure of how well the test results reflect what the examinees know about the subject the test was intended for (American Psychological Association, et. al., 1985). There are two types of validity: criterion-related validity and representative validity. Criterion-related validity is the predictive power of a measure. In other words, a test with a strong criterion-related validity precisely forecasts how well the subject will perform in a particular setting. Representative validity is the ability of the exam to make inferences about the behavioral domain of a person being measured. This type of validity can be further divided into *signs* and *samples*. Signs explain the domain; samples describe the domain. If an exam is going to serve as a sign it must have high construct validity; if it is to be treated as a sample it must have high content validity (Mehrens and Lehmann, 1984). Construct validity is the degree to which an exam measures an intangible quantity such as depression, happiness, leadership, or mental retardation. Content validity is determined by how well the exam portrays the domain of concepts it is intended to represent. The better the sampling of ideas from the domain, the better the exam (Messick, 1980). The DEEM's purpose is to serve as a sample, so its content validity must be conclusive and its criterion-related validity and construct-validity need not be as strong.

Content validity is particularly important for achievement tests, and a diagnostic exam is certainly is one form of achievement test. From the results of a diagnostic exam, an instructor wants to make a judgement about how well a student or class has learned a particular set of concepts, in this case, the basic concepts of electricity and magnetism. To ensure the judgement is sound, the content of the exam must truly represent the subject matter (Mehrens and Lehman, 1984).

Criterion-related validity can be measured by correlating the outcome of the exam with outcomes from the setting the exam is intended to predict. Likewise, construct validity can be expressed as the relationship between the outcome of an exam and some other standard measure of a particular trait. A test designer creating a content valid exam does not enjoy the luxury of being able to calculate simple correlations. However, the test designer can overcome these shortcomings by assessing the content validity of the exam in other ways. One of the more obvious means to gauge the content validity of an assessment tool is to have it examined by experts in the field and have them criticize its content. This sort of evaluation can be quantified, in a rudimentary sense; however, it seems the most important aspect is to have different perspectives on the subject domain and the manner in which the items address that domain.

Face validity is another way to address content validity, however simple-minded it may seem. Basically, the exam must “look good.” This is strictly a method of increasing the acceptance of the exam by the test administrators and examinees (Nevo, 1985).

Similarly to reliability, when one address validity, one needs to ask, “How valid should a test be?” And again, the answer depends on what it is to be used for. Obviously, one wants to create a test that is of the highest level of validity. In the case of the DEEM the only sure way to do that is to have it reviewed by pedagogical experts in the field of introductory electromagnetism and take their input accordingly. Since the DEEM is not intended for placement or prediction, a strong validity is not crucial in that respect. If the exam is determined to be invalid in any respect, it will not be employed. Therein lies the motivation to make this a valid exam; the more valid the content, the wider the acceptance.

37) CONSTRUCTION OF ITEMS

After the test designer determines what concepts the exam will cover, he or she can begin writing items that address those concepts. Generally, an item is composed of two parts: the stems (the questions or statements of the problem) and the options (answers).

An item writer must consider a host of issues when creating an item. The investigator found two sources particularly helpful in supplying a list of considerations (Horn, 1970; and Oosterhof, 1996). Their recommendations fall into two categories: suggestions dealing with writing stems and suggestions dealing with writing options. Sections 24) and 25) are devoted to the process of writing stems and writing options, respectively.

The investigator created two additional item components: pre-stems and extended stems. A pre-stem is a statement preceding the stem that gives instructions to the examinee based on choices from the previous question. An extended stem is a broad set of statements that apply to a specific set of items. Section 26) will discuss details and issues concerning the writing of pre-stems and extended stems. The last section of this chapter presents arguments regarding the level of difficulty for the items.

38) Stems

When one is writing a stem, its purpose should be considered first. In other words, one should consider if a stem probes an idea or concept that is part of the behavior to be investigated. Furthermore, the question should test only a single concept. If an item tests more then one concept simultaneously, the interpretation of the results becomes difficult, if not impossible. Other points one should be aware of when writing a stem are that the language should be clear, the reading level appropriate, and statements regarding figures and diagrams must be accurate. Another, perhaps less obvious, point an item writer should consider is the emphasis of adjectives or adverbs that reverse or significantly alter the meaning of the stem (Oosterhof, 1996). For example, words such as “not”, “least”, or “except” should be emphasized. An item writer accomplishes this by boldfacing or italicizing these words or phases.

One author (Oosterhof, 1996) suggested that the use of extraneous information in a stem can be misleading and confusing, while another (Horn, 1970) advocated the use of extra information. The DEEM’s stems do not contain extraneous information. The exam is difficult enough as it stands; further complications would only serve to frustrate and disorient students.

Stem writers also need to avoid creating stems that force test takers to work out individual problems for each option. An short example will help illustrate this point (Horn, 1970):

> Which of the following has a volume of 66 cu. in.?
>
> a) Cone: Radius of 3”, height of 7”
> b) Prism: Rectangular base 3” x 4”, height 6”
> c) Pyramid: Square Base with sides 3”, height 6”
> d) Cube: Sides 4”
> e) Cylinder: Radius of base 3.5” height 8”

A student only receives credit on a typical multiple-choice problem if that person gets it correct, and no partial credit is given. This is problematic for the examinee because if there is a simple mistake in the logic or mathematics, then no credit is given. In the above example, there is potential for this problem in each of the five options.

Finally, one should not create stems that depend on knowledge outside the realm of the exam. This was a particular challenge for the investigator while creating the DEEM. There are many strong questions that probe students' understanding of physics using electromagnetic concepts; however, many of these same questions also rely on concepts from mechanics. Chapter 5 has examples of these types of questions, as they were avoided in the interviews as well.

39) Options

Once one has created an effective stem, then one must create equally effective options. An item writer faces no less of a task creating options than when creating stems. One must consider many points when composing options. First, all options must be parallel in context and arranged logically (increasing or decreasing magnitude, alphabetically, etc.). Second, repetitive words must be eliminated. Third, only plausible distractors (incorrect options) alongside an undeniably correct key (correct option) should be included. Fourth, the grammar of the options should be consistent with the stem. Fifth, it is important to shun options similar to "none of the above" and "all of the above." (Chapter 6 provided insight into why these two types of options should not be employed.) Finally, it is essential to avoid including options based on knowledge outside the realm of the exam (Oosterhof, 1996).

40) Pre-Stems and Extended Stems

The construction of the DEEM introduced two more parts to an item, or set of items, namely extended stems and pre-stems. An item writer should employ the same standards when constructing extended stems and pre-stems as when constructing stems. Extended stems introduce other considerations as well. First, the information given in an extended stem must not exclude options that are reasonable distractors. Second, the items the extended stem applies to must be referred to explicitly. Third, a nested extended stem must not contradict other nested extended stems or the main extended stem of which it is a part.

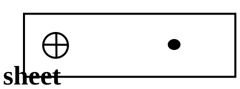
Pre-stems are not as troublesome. Their function is to keep all the distractors in the previous item plausible by not rendering the current item meaningless based on the response of the previous question. Figure 7– presents an example of an extended stem and a pre-stem.

Figure 23–I: Examples of an Extended Stem and a Pre-Stem

For questions 10-33 the dots, • , represent points in space (they are not particles).
The circles represent particles with a net negative charge, ⊖, or a net positive charge, ⊕. The magnitudes of the net charges are all equal.
The thick, black lines, ━━━, represent side-views of long, conducting wires. An "**I**" with a solid arrow next to it (e.g. **I**→) indicates the direction of current along the wire in the plane of the page. No "**I**" in the figure indicates that there is no current along the wire. The magnitudes of the currents are all equal.
Any and all fields which may be present in each figure are due only to charges or currents in the figure.
Assume that the value of the electrostatic potential equals zero an infinite distance away from a point charge.

10) For the figure in the box to the right, the net electric field at the dot points...

a) → b) ← c) ↑ d) ↓ e) There is no net electric field.



11) **If you chose answer "*e*" in question 10, then leave line #11 on your answer sheet blank, skip this question, and go directly to question 12**. Relative to the magnitude of electric field at the dot in the figure for question 10, what would happen to the magnitude of the electric field at the dot, if the dot is moved to a new position farther to the right?

a) The magnitude of the electric field at the dot would be greater.
b) The magnitude of the electric field at the dot would be less.
c) The magnitude of the electric field at the dot would not change.
d) There is not enough information.
e) None of these.

Before the examinee reaches item 10, that person must read the extended stem that sets parameters and states assumptions for items 10-33. Once that is done, then that person moves on to item #10, answers it, and then moves to item #11. Here if the examinee were to have chosen option *E* in item #10, then item #11 would become meaningless. To prevent this logical inconsistency, a pre-stem was added to item #11 indicating that if one choose option *E*, then one should skip item #11 and proceed directly to item #12.

41) Item Difficulty

An item writer faces one last concern: How difficult should the items be? One approach is to create some problems that are easy, some that are difficult, and most that are somewhere in-between. This prevents the examinees' scores from piling-up at the top-end or bottom-end of the scale. This approach, however, contradicts creating a more reliable exam (Glass and Hopkins, 1970). Ultimately for this type of exam, the investigator's concern to increase reliability dominated decisions regarding item difficulty. Therefore, when possible, the investigator created items to be of medium difficulty.

"Do not on any account attempt to write on both sides of the paper at once."
– Sellar & Yeatman

GUIDELINES FOR ADMINISTRATION

The test creator must lay out administration guidelines to ensure the instructional community administers the exam as intended by the creator. This chapter will describe the three basic administration obligations that pertain to the DEEM. The administrator's first responsibility is to limit the time students have to finish the exam. The second responsibility is to administer the exam at prescribed points during the semester. The administrators' final responsibility is to motivate the students. When the exam is disseminated to the physics education community, the investigator will include a description of these three obligations to help establish a uniform testing environment.

42) Time Limit

The instructor should permit the students fifty minutes to complete the DEEM. Although this exam is not a speed test, allowing the students more time would surely boost their scores, while giving them less time would certainly reduce their scores.

If a student has a learning disability, requires extra time, or needs to take the exam under special circumstances, interpretation of the pre-test and post-test scores is left up to the individual instructor.

43) Administration Dates

The instructor must administer the exam at prescribed points in the course. The appropriate date for the pre-test is before any exam material has been covered in the course. If one gives the pre-test too late in the semester, then it becomes impossible to discriminate between what the students knew before they entered the class and how much they had gained from instruction. The appropriate date for the instructor to administer the post-test is after all the exam material has been covered and the students have had a short time to internalize the concepts. It would be appropriate for the instructor to administer the exam a few days to a week after all the concepts on the test have been covered in the class. Moreover, one should not wait too long to administer the post-test, as this would only give students the opportunity to forget fine details that could affect their scores.

44) Student Motivation

The instructor must provide some sort of motivation for the students to perform well on this exam. This is an essential administrative obligation. If students do not try, then the data have no meaning. Depending on how the instructor intends to use the data, student motivation on the post-test is particularly critical. One can unintentionally lower student motivation by delaying the post-test date. When the students feel the exam has no meaning for their education (or grade) they will not take it seriously. Even if the instructor plans an appropriate class for the post-test, it is still necessary to provide some incentive to the students. The proper level of motivation is a difficult parameter to define. Ethics, preferences, and standards all play a role in how far one should go to motivate the students to perform their best on an exam. An effective tactic used by the instructional staff at Rensselaer Polytechnic Institute was to make the post-tests count as either extra credit or an in-class grade for that day.

An instructor also faces a difficult task when trying to provide motivation to the students on the pre-test. It would be unfair for an instructor to assign a grade to a student's performance on the pre-test, as different students will have different backgrounds. However, the instructor should provide some motivation on the pre-test. One should be particularly wary of notifying students that they are taking a pre-test. If students realize the pre-test is only to establish a baseline, from which they will be compared to later, then it is in their best interest to do poorly. For the DEEM the pre-test motivation was provided in a prepared statement read to the students before the exam began. The statement explained to the students that the exam was "…to test their basic understanding of the concepts of electricity and magnetism…to improve the educational environment, so it is in your best interest to do well." Furthermore, it stated that "the exam will be graded." Although this is not a strong motivating factor, it seemed to work. Also, it was noted by the investigator and others that when students are given a test, particularly in the first days of class, they tend to go into "test-mode" and try hard anyway. This effect is mostly likely not universal, but it seemed to be prevalent in the pool of examinees used to generate data to help refine the exam.

TEST ADMINISTRATION, GRADING, AND REVIEW

For a test creator to originate a successful and widely accepted exam, many steps must be followed (Crocker and Algina, 1986). Two steps in particular demand so much time and effort and are so critical to the development of the exam they warrant special attention. The first is field-testing the exam; the second is having experts in the community review and comment on the exam's content. In general, field-testing (and grading) provides large quantities of data to be analyzed and evaluated, establishing the effectiveness of the exam from a statistical point of view. Review by the education community provides opinions regarding the exam's content, the relevancy of the questions, possible alternate interpretations of items and responses, and misleading wording. In other words, field-testing and review provides a wealth of information that improves the exam's validity and reliability.

The first section of this chapter will explain in detail the process involved in the administration of the first three versions of the DEEM. The second section will describe the grading of the exam. The final section will detail the means by which the exam's existence was announced to the physics education research community and how feedback was obtained.

45) Test Administration

Administering of the DEEM followed strict procedures. The exam was to be given on or around the first day of classes and then re-administered after all the material on the exam had been covered. The post-test date did not necessarily coincide with the last day of class. All of the material on the exam was covered well before the end of the semester. It should be noted that two copies of the exam (pre and post) were not created for every student. Because the classes were staggered in small sections or labs (with about thirty students in each), only about one hundred and fifty copies were needed for everyone for both exams. This means that many students would use an exam that had been used previously by someone else. Since the students were asked not to write on the exams, it is doubtful that this procedure helped anyone’s scores. A brief scan of the exams after they were used did reveal that very few of them had markings on them.

On the day of the exam, students were provided with bubble sheets to fill in their name, gender, social security number and answers to the exam. The students had fifty-minutes to complete the test. Before the exam began, the test administrator read a statement to the students telling them what the exam was about and what to do. The purpose of the statement was to establish uniformity in the administration of the exam from instructor to instructor, to inform the students as to what they were undertaking, to set up the procedures for the students, and to provide some motivation as well.

While students were taking the exam, the administrator was free to answer questions about the exam, but not about details of electricity and magnetism. For example, if someone did not understand the way an item was worded, they were encouraged to request clarification. However, test-takers were discouraged from asking specific questions about EM. Say, for example, asking a question about the force on a positively charged particle in a magnetic field.

After fifty minutes had passed, the test administrator collected the exams and bubble sheets. Most students did not take the full fifty minutes to complete the exam. As a rule, the majority of the papers were handed in after thirty to thirty-five minutes. This was true for both the pre-test and post-test. The investigator's analysis of the number of people who finished the pre-test (Version 3.0) revealed about 95% of the students answered at least items 1 through 52 (out of 70) and about 90% finished the exam. On the post-test, about 97% of the students reached question 62 and about 96% of the students finished the exam. On Version 1.0, a longer, but somewhat easier exam, nearly all the students (96%) finished the entire exam, pre and post.

The following three subsections will be an overview of each administration of the DEEM to provide details regarding the testing environment.

46) Version 1.0

The investigator administered the first version of the DEEM to the Physics II class at Rensselaer Polytechnic Institute in the Fall Semester of 1996. This class was taught in the traditional style of lecture/laboratory/recitation. The DEEM was administered to the students in their first laboratory meeting, which was held the second week of classes. This was not an ideal setting to administer the pre-test, because by the second week students had been exposed to some of the material on the exam. This undoubtedly raised pre-test scores somewhat. The administration of the post-test was done at a reasonable time during the semester. One problem with the administration of Version 1.0 was that the number of students who took the post-test, 273, was quite small compared to the number of people who took the pre-test, 437. This was a result of the normal decrease in attendance in a class offered in a traditional style of instruction at Rensselaer Polytechnic

Institute. Out of the 273 people who took the post-test, the investigator was able to match 213 students. This was a reasonable number to conduct most of the appropriate statistical analyses.

47) Version 2.0

Version 2.0 was administered in the Spring Semester of 1997. By this time Rensselaer Polytechnic Institute had moved to implementation of the studio version of Physics II (Wilson, 1994). The administration of this version was on the first day; however, because of time constraints the students were only permitted approximately thirty minutes to take the pre-test. Needless to say, students generally were unable to finish the pre-test. Also, many students did not put their names or social security numbers on the exam, which effectively removed them from subsequent participation in any in-depth analyses. The post-test was offered on the very last day of class. The students had already received their grades for the class, and many were not even present. Student motivation was very low. As the investigator observed students taking the exam, and from a cursory review of returned exams, it became readily apparent that the whole semester's batch of data was worthless. The investigator conducted informal exit polls with students after they took the exam. These only served to confirm this suspicion.

48) Version 3.0

The exam was again delivered to the students in the Studio format in the Fall semester of 1997. The instructors administered the pre-test on the first day of class. The post-test was given to the students approximately two-thirds of the way through the semester and to an extremely large group of motivated students. The course supervisor, Dr. Casabella, recognized that if the exam was offered as extra-credit to be included in their in-class grade, the students would try to do well on the exam. This certainly seemed to be a correct assumption. Three hundred and ninety-five students took the exam in the beginning of the semester; 379 students took it at the end of the term. The union of those two sets yielded 347 matched students. (This number would have been even higher, except there was confusion in one section as to whether or not students should put their names and social security numbers on the sheets.)

49) Grading the Exams

For each of the three administrations of the exam, the students filled in their responses to the multiple-choice questions on a *General Purpose Answer Sheet: form no. 16482* with a #2 pencil. Students also filled in name, social security number, and gender as well as the corresponding bubbles. Students were not instructed to fill in other fields such as grade and birth year, on the information sheet.

The investigator graded the exams using Rensselaer Polytechnic Institute registrar's program *TEST120*. An optical scanner read the information off the Scantron sheets and fed this data into the program for grading. The investigator prepared the program to scan by first scanning a key sheet that had all of the correct responses filled in on it. Once the machine was ready, the students' papers were then fed into the optical scanner one sheet at a time. The optical scanner could detect any extraneous marks or errors in a darkened bubble on the sheet. If the scanner detected an error, it would indicate so and not allow the computer program to grade the sheet. When this occurred, the investigator would correct or erase the stray marks and feed the Scantron sheet through again. The investigator would also correct any information that was not properly filled in. For example, a student may have only entered his or here name in the appropriate field and not filled in the bubbles underneath. In that case, the investigator would carefully darken the appropriate bubbles below the letters of the student's name.

When all of the sheets were graded, the program would save the results on a double-density 3 1/2" floppy disk with a *.dat* extension. Unfortunately, this format was not compatible with Excel, so modifications had to be made. First, the data had to be melded from several small sets of data with paragraph marks at the end, to one long set of data with a single, terminating carriage return. This was done using an author-created macro in MS Word. Next, the data had to be opened in Excel and partitioned properly. The grading program returned four sets of data for each student – biographical, answers to each question (0-4 and blank), each question graded correct (1) or incorrect (0), and raw score. The investigator separated the biographical data by last name, first name, middle initial, gender, and social security number. Any other information that was provided by the students and read by the Scantron was eliminated at this point. The answers to all questions comprised one, long field, so the investigator had to divide that string making each integer represent one question. The same thing was done for each item's score.

From there, the investigator had to alter the data to account for the questions that had *omit* as the correct answer. The grading program did not recognize *omit* on the answer key as a valid response. Therefore, everyone, regardless of whether they answered the item right or wrong, was marked incorrect. The investigator fixed this by updating the column associated with the items that had *omit* as the correct grade. One point was awarded for a blank response; zero was awarded for anything else. This procedure rendered the raw score, as returned by the grading program, invalid. The investigator obtained the actual raw score by commanding the spreadsheet to add up 1's in a row that were located in cells indicating whether or not a item was correct. (One row in the spreadsheet was devoted to each student.) At this point, the students' responses were properly graded and in a format that made them ready to be analyzed.

50) Review of the DEEM

The investigator announced the availability of the first version of the exam to the physics education community via the *phys-l* and *physlrnr* listserves in August of 1996. The announcement included the web site address (currently– http://ciue.rpi.edu/jeff) where information regarding the exam could be found, as well as a form that interested physics instructors could fill out to obtain a copy of the exam. The site contained a brief history of the creation of the exam, as well as its length, intended audience, and examples of interview questions and exam items. At that time there were two versions of the exam, a 76-question version (referred to as Version 1.0) and a 30-question version. Since, very few physics educators requested the short version, and because thirty questions is insufficient to properly cover all the material, the investigator discontinued offering that exam.

The form interested instructors were asked to fill out posed a few simple questions: "Name?", "Name of institution?", "How many students will the exam be administered to?", "Desired exam format? (MSWord or .pdf)", and "Comments?" By filling out the form, the instructors agreed not to use any part of the exam for their own work without permission of the investigator, and they would not publish or claim any of the questions as their own. The instructors did not have to administer the exam if they did not want to, but if they did, they were not permitted to publish any data from the exam without permission of the investigator. To ensure the investigator had a record of everyone that had a copy of the exam they were also not permitted to share the exam with anyone.

Once a request for the diagnostic exam was received, the investigator would electronically mail the exam in the requested format. A small note was included, thanking the instructor for showing interest in the exam and stating that the investigator looked forward to getting results from them and/or receiving comments or criticisms. When the investigator finished Versions 2.0 and 3.0, he announced them on the aforementioned listserves, as well as the *physhare* list. To date over one-hundred and fifty physics educators from around the world have requested a copy of the exam.

The majority of the instructors who requested the exam seemed to be only curious about it. Most, although not all, of the instructors' comments were complimentary. The complimentary comments fell into categories such as, "I'm glad that someone is finally doing this," and "It looks like you've done a nice job." The less praiseworthy comments were along the lines of, "You did not cover a lot of topics," and "The test is too repetitive." Many of the instructors offered suggestions, most of which were very helpful. Details of specific comments and criticism on particular items will be reserved for Chapter 13, when an in-depth look at each question will be undertaken.

"The ultimate umpire of things in life is – Fact." – Agnus C. Laut

DISCUSSION OF STATISTICAL TERMS AND TECHNIQUES FOR TEST ANALYSIS

51) Introduction to Chapters 10, 11, and 12

This and the following two chapters encompass the statistical analysis of the exam. This chapter focuses on the statistics of the exam as a whole. The following chapter focuses on the items individually, and Chapter 12 deals with gender biases in the items.

The investigator procured data for the statistical analyses using a matched set of students (from the pre-test to the post-test) as the population of examinees. In other words, the investigator used only scores from students that took both the pre-test and the post-test. Although this tactic can limit the size of the examinee pool, benefits far outweigh losses. By matching students, one can easily search with confidence for changes in scores and response patterns from the pre-test to the post-test. This strategy removes doubt that perceived variations are real, and not simply an artifact of the shifting body of students. As it turns out, not many of the students' exams were dropped as a result of missing one of the two rounds. There were 395 examinees that took the pre-test and 379 that took the post-test of Version 3.0 of the DEEM. This means there was a maximum of 379 students participating in both exams whose scores could be matched. Amazingly, the investigator was able to track 348 students from pre-test to post-test. When one considers that 100 examinees is the general rule-of-thumb for an adequately sized group, 348 certainly provided a large pool of grades from which the analysis of the exam could take place.

52) Introduction to Chapter 10

This chapter describes the statistical tools used to gauge the effectiveness of the entire exam. Generally, each section of this chapter is divided into three parts. First, the section will define the technique at hand and explain, when necessary, the variables of any equation. Next, each section will offer general assumptions, more in-depth details, and concerns (if warranted) about the particular tool or technique. Finally, each section will present results, as they pertain to Version 3.0 of the DEEM, with comments regarding their meaning for the exam as a whole.

53) Score

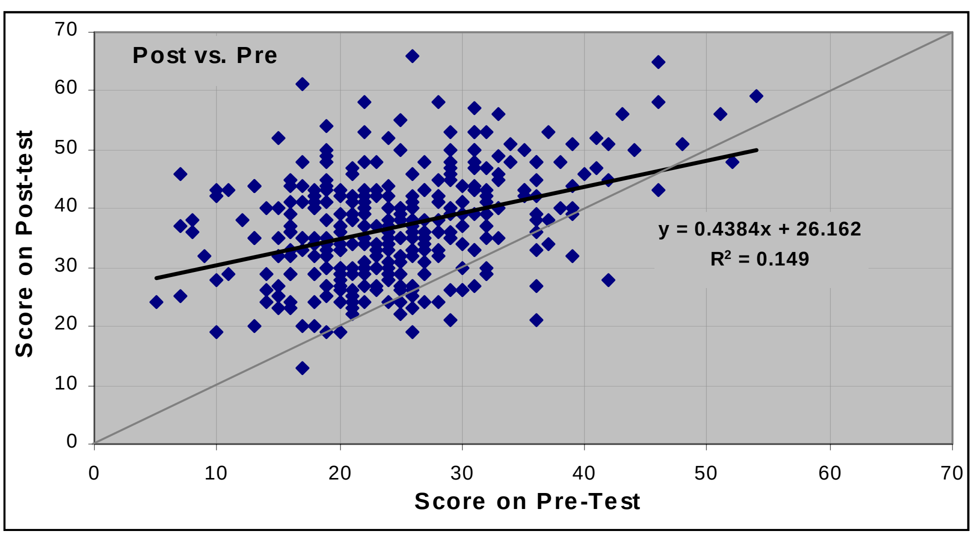
A student's score is the number of correct answers on the exam. There was no penalty for guessing; therefore, no formula was needed to correct for random answering. Scores are *not* listed as percentages (i.e. 100*number correct/total number of test questions) unless stated otherwise. The average score on Version 3.0 of the DEEM was 24 (±8.0) on the pre-test and 37 (±9.1) on the post-test out of 70.

One small note about the score: if a student turned in a blank paper, or did not reach item 33, or for whatever reason left item 33 blank, that student still received credit for that question, because the correct response to item 33 was to omit it. In other words, to achieve a score of 0, an examinee would have had to answer every question wrong (one way to do this would have been to simply leave all the items blank) and also filled in item 33 with a response of *A*, *B*, *C*, *D*, or *E*.

54) Correlation Between Post-test and Pre-Test Scores

One would expect there to be some positive correlation between pre-test and post-test scores. Presumably everyone, or nearly everyone, should increase their score. Figure 9–I is a graph of post-test scores versus pre-test scores. Data points represent the individual scores. The thin, diagonal line splits the area of the graph in two. Examinees in the upper-half increased their score from pre- to post-test; examinees in the lower-half decreased their score from pre- to post-test. It should be obvious to the casual observer (and admittedly reassuring) that most students increased their score. MS Excel included the thick, black line as a simple linear regression line and the equation of this line and R^2 as an indication of the fit of that line. It is not surprising the regression line has a positive slope, but a value that is less than one. It simply reinforces the fact that those who scored well on the pre-test were "forced" to achieve a score on the post-test approaching perfect to earn an increase. Later in the chapter, the relationship between scores and gains will be addressed.

Figure 9–I: Post-Test Scores vs. Pre-Test Scores on Version 3.0 of the DEEM



9.1.1 t-Test

The investigator preformed a dependent group t-test on the difference scores to determine if the increase from pre-test to post-test was significant. The equation for calculating the dependent group t-test is:

$$t_{\overline{D}} = \frac{D_{ave}}{\sqrt{\dfrac{\sum_{i=1}^{N} D_i^{\,2} - \dfrac{\left(\sum_{i=1}^{N} D_i\right)^2}{N}}{N(N-1)}}}, \tag{9:1}$$

where D_{ave} is the average difference between post-test and pre-test scores, D_i is an individual difference score, and N is the number of students (McGrath, 1996). The post-test of Version 3.0 yielded a t-test value of 24.6. The critical value for a two-tailed t-test with a 95% confidence level is 1.97. Version 3.0 clearly passed this critical value, so the increase from pre-test to post-test can be considered significant.

55) Measures of Central Tendency and Variability

The following is a brief overview of commonly used statistical terms to describe properties of exam scores. The underlying assumption is that the exam scores closely followed a normal distribution. The validity of this assumption will be put to trial in the second half of this section.

56) General Statistics

- MODE –The mode is the most frequently occurring value in a set.
- MEDIAN –The median is the mid-point value for a set.
- MEAN – The mean is the average value in a set.
- MINIMUM – The minimum is the lowest value in a set.
- MAXIMUM – The maximum is the highest value in a set.

- STANDARD DEVIATION – The standard deviation is the average of the differences between the scores and the mean. The standard deviation of a set is given by the equation:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}, \qquad (9:2)$$

  where X_i is an individual value in the set, μ is the mean of the set, and N is the total number of items in the set.

Table 9–I summarizes the results of these general statistics for Version 3.0 of the DEEM. (Remember these values are in terms of raw scores out of 70, not percentages.)

Table 9–I: Results for Version 3.0 of the DEEM

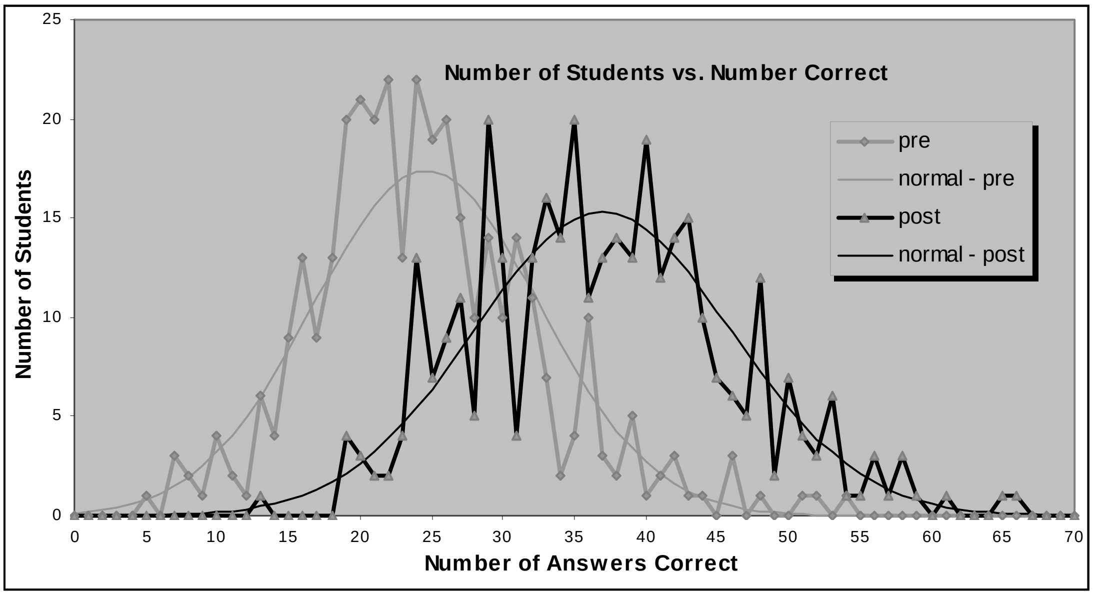
N= 347	Pre	Post
Maximum	54	66
Median	24	37
Mean	24.64	36.97
Mode	24	35
Minimum	5	13
Standard Deviation	7.97	9.05

This table shows no remarkable features. The median, mode, and mean are all very close in both administrations, and there is a clear shift toward higher scores from pre-test to post-test. This shift is explicitly indicated by the increase in the *maximum* and the *mean* from pre-test to post-test. The standard deviation widens somewhat, but this is indicative of the more motivated students increasing their scores more than the less motivated students.

57) Tests of Normal Distribution

The DEEM's exam scores, both the pre-test and the post-test, closely followed a normal distribution. One way to see this is from a graph (Figure 9–II) of the frequency of the number of students versus score.

Figure 9–II: Frequency Distribution of Students' Scores Compared to Normal Curves



By inspection of Figure 9–II, one can see that the two distributions of grades appear normally distributed. The thin-lined, smooth curves represent a normally distributed population, normalized to the same number of people and with the same mean and standard deviations as the set of scores to which they are being compared. To move beyond a simple visual inspection, one should calculate the skew (S) and the kurtosis (K) of these distributions.

The skew of a distribution indicates how much a set "leans" toward low scores or high scores, relative to the mean. A positive skew means there are more scores at the low end, while a negative skew means there are more scores at the high end. The equation to calculate the skew (*S*) is:

$$S = \left(\frac{n}{(n-1)(n-2)}\right)\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^3, \qquad (9:3)$$

where n is the number of students, x_i is an individual score, μ is the average, and σ is the standard deviation of the set (Glass and Hopkins, 1984).

The kurtosis (K) measures how much the distribution is peaked or flattened as compared to the normal distribution. A positive kurtosis corresponds to a peaked distribution, while a negative kurtosis corresponds to a flattened distribution. The equation for kurtosis (K) is:

$$K = \left\{\left(\frac{n(n+1)}{(n-1)(n-2)(n-3)}\right)\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^4\right\} - \frac{3(n-1)^2}{(n-2)(n-3)}, \qquad (9:4)$$

(Glass and Hopkins, 1984).

The skew can have values from $-\sqrt{t}$ to $\sqrt{t}$ and the kurtosis can have values from -t to t, where t is the number of test questions. For the DEEM (Version 3.0), the limits of the skew and kurtosis are ±8.4 and ±70 respectively. The DEEM yielded the following values for the skew and the kurtosis: $S_{pre} = 1.34$ and $S_{post} = 1.04$, and $K_{pre} = .54$ and $K_{post} = -0.18$. These results should be of no surprise in light of Figure 9–II. The slightly positive skewing of the results suggests the exam is more on the “difficult side” than on the “easy side”; a fact supported by the mean. Although both of these calculations suggest the data set is only slightly off from a normal distribuation, they also point to the fact that the group of students moved toward a more normal distribuation in the post-test. Also, when one considers the range of values these two indices can have, neither one is particularly worrisome.

58) Percent Gain

The percent gain, as it is used here, was put forth to describe the increase in a student's score from pre-test to post-test on the *Force Concept Inventory* (FCI) (Hake, 1998). The equation for percent gain is:

$$\%\text{ gain} = 100\frac{\#\text{correct}_{\text{post}} - \#\text{correct}_{\text{pre}}}{\#\text{of test questions} - \#\text{correct}_{\text{pre}}}, \tag{9:5}$$

where $\#\ correct_{post}$ and $\#\ correct_{pre}$ are a student's scores on the post-test and pre-test respectively, and *# of test questions* is the total number of questions on the exam, or restated, the highest possible score. Percent gain is a quasi-normalized measure of how much a student's score changes.

There are two ways for one to report the percent gain for an exam administered to a group of students, either as the *average of the gains* or as the *gain of the averages*. The first method looks at each examinee's gain and then averages all of the gains together. The second method looks at the average pre-test and post-test scores of all the examinees and then finds the pecent gain. Although this may seem like a trivial point, the two methods yield different results. For example, the first method could easily become discontinuous if just one person achieves a perfect score on the pre-test (see equation 9:5). For this to happen with the second method, everyone would have to get a perfect score on the pre-test. For the FCI, it is presumed that the second method is the one that is commonly reported because of its ease of calculation and because it somewhat side-steps the issue of discontinuities. It is not stated in the more important pieces of literature regarding the FCI, however, which one is actually used.

The careful reader will have noted that the percent gain can be as high as 100%. This is achieved by anyone who seizes a perfect score on the post-test, regardless of the pre-test score. But percent gain can be as low as -100*T %, where *T* is one less then the number of test questions. (Actually, it can be an infinitely negative number if, as previously stated, a person scores perfectly on the pre-test.) So, barring a perfect pre-test score, the range for the percent gain on the DEEM is $-6900\% \leq \%gain \leq 100\%$. Given the wild theoretical bounds, it seems odd it would have such appeal as a publishable test statistic. Closer inspection, however, reveals a much more docile side. As long as the examinees' pre-test scores avoid perfection and their scores tend to increase on the post-test, this analysis tool is perfectly acceptable. This pattern of scores forces the grades to fall in a region where they are not going to cause the gains to spiral down towards negative infinity. However, there still remains a case that one must consider, that is, when a student achieves a perfect score on the post-test and receives a gain of 100%. When this happens, there is no distinguishing feature to resolve between the student who knew nearly everything on the pre-test and achieved, say, 99.0% and then went on to get a 100% on the post-test, from the student who only could achieve a 20% on the pre-test, but learned all of the material (or at least a good piece of it) and also got a 100% on the post-test. Here, the only solution is for the instructor to pay close attention to the details and ensure that the examinee's scores providing ambiguous data, by lingering in the extrema, are dealt with carefully.

The creation of a diagnostic exam is, in itself, a means for preventing such possibilities from occurring in the first place. The creator designs an exam with plenty of distractors. This keeps the pre-instruction scores low, while providing room for scores to increase, but preventing many post-instruction scores from being perfect, or nearly perfect. After all, one would expect students entering a class to have little or no knowledge about the subject. And one would expect students leaving the class to have learned something about the material. So, it appears that percent gain is an appropriate, although not perfect, way to report relative changes in students' scores.

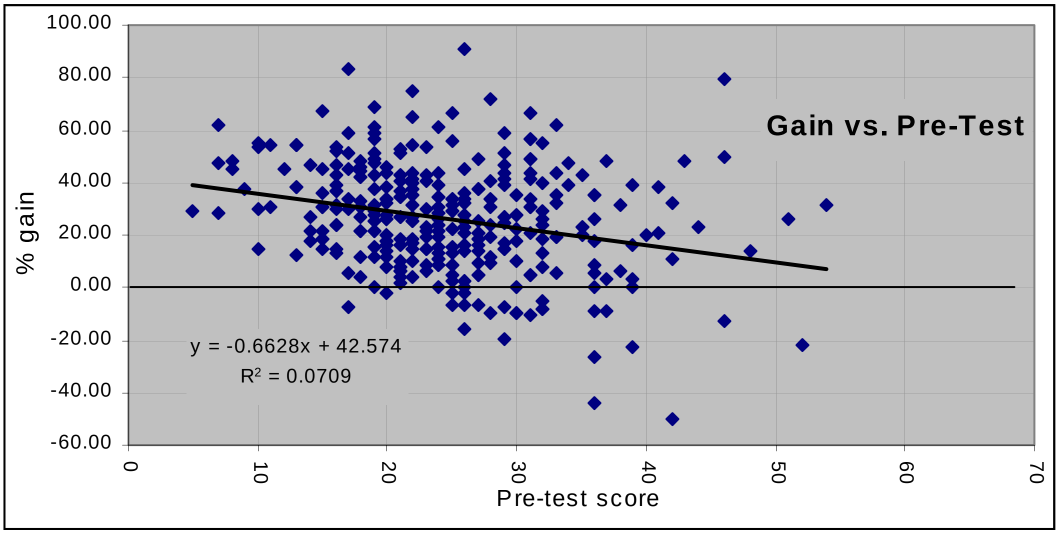
For the DEEM, the gain of the averages was 27.2%. One can not definitely say that this gain is "strong" or "weak" as there is nothing to compare it to. However, in light of the pre-test and post-test scores and standard deviations, this gain is reasonable and significant.

59) Relationship Between Percent Gain and Score

Ideally, one would want a class where everyone, regardless of what they knew coming in, left understanding all that was taught and as a result scored very high on the post-test. Unfortunately, this is rarely the case. But it is interesting for one to note how students' gains were dependent on their pre-test scores.

Figure 9–III is a graph of students' percent gains versus pre-test scores. MS Excel included the thick, black line as a simple linear regression line and the equation of that line in the graph. This figure indicates that students who scored low on the pre-test tended to raise their score by a higher percentage as compared to those who scored well on the pre-test. This result is not surprising. A student who enters a class knowing most of the material will only improve slightly, or get worse. A student who knows very little can only go up. Any data point below the x-axis indicates an examinee whose score on the post-test was lower than that on the pre-test.

Figure 9–III: Relationship Between %Gain and Pre-Test Scores



60) Average and Standard Deviations of Item Analysis Techniques

Another way for a designer to glean information about an exam is to calculate the average and standard deviations of item analysis techniques. For example, item discrimination is the difference between the percentage of students in the upper 27% of the class that answered the question correctly and the percentage of students in the lower 27% of the class who answered the question correctly. (Details of rationale behind using the upper and lower 27% as well as results of the item discrimination will be left until the next chapter.) A test designer is provided with quick insight into the exam by looking at the average and the standard deviation of all of the item discriminations. For example, if the average of the item discriminations is very low, the designer is sent a clear signal that many of the items need to be revised. The next chapter discusses individual item analysis techniques, and some of their averages and standard deviations will be reported when it provides useful information about the exam.

61) Coefficient alpha

Coefficient alpha is an estimate of the internal consistency of test scores from a single set of examinees (Crocker and Algina, 1986). The equation for α is given by:

$$\alpha = \left(\frac{N}{N-1}\right)\left(1 - \frac{\sum_{j=1}^{N} \sigma_j^{\,2}}{\sigma^2}\right). \qquad (9:6)$$

Here N is the number of items, σ^2 is the variance in the test scores, and σ_j^2 is the variance of each test item. It should be clear that α takes on values $0 \leq \alpha \approx 1$. Coefficient alpha is a strong measure of the internal consistency of the exam. One way for a test designer to measure exam reliability is to divide the exam into two parts and administer them to two separate, yet similar, pools of examinees, then measure the correlation between the scores of the two groups. There are many possible ways, however, to split an exam in two; trying them all would be impossible. Coefficient alpha is the mean of the correlation between all the possible halves of the exam (Aiken, 1967). So, the larger the value of α the greater the reliability of the exam, or, as stated earlier, the greater the internal consistency of the exam.

One must cautiously interpret the DEEM’s coefficient alpha. Since the exam covers many different topics in electricity and magnetism, one would expect the internal consistency of the exam to be low. For example, many more students grasp the direction of the force on a charged particle in an electric field as compared to the idea of potential energy. Moreover, one would expect for introductory students that there is a weak, or completely lacking, connection between those two concepts. Therefore, one would surmise the internal consistency of the exam would be low. The value of α for the DEEM was α=0.60 (pre-test) and α=0.74 (post-test). Certainly this is not a robust score of 0.90, but it is the same as a reported value for the FCI a test designed to measure only one concept, force (Pinkerton, 1997).

62) STATISTICAL TERMS AND TECHNIQUES FOR ITEM ANALYSIS

This chapter will introduce the statistical techniques used to evaluate the effectiveness of the exam's items. For the most part, the methods in this chapter treat items individually; however, some item analysis tools should be thought of not only as a means unto themselves, but also in the context of the entire exam. For example, patterns of student responses on a particular item can suggest something about how effective that question is. But, a test designer may gain better insight by weaving those patterns into the larger fabric of similar questions.

Generally, each section of this chapter is divided into three parts. First, the section will present simple definitions and descriptions of the equations and variables used to familiarize the reader with terms that may not be commonplace or which may be interpreted in more than one way. Next, each section will deliver arguments supporting particular choices of parameters in cases where judgment is permitted. For example, one's interpretation of an item difficulty index is not subjective; however, one can, and should, adjust the parameters in determining the index to appropriately fit the pool of examinees. Finally, each section will examine cases of items, or sets of items, to provide concrete examples of the item analysis technique and a glimpse of what is coming in Chapter 13, when a full-scale investigation of each question will be undertaken.

9.2 Frequency Table

A frequency table depicts the percentage of students that chose a response for a particular question. For example, on Version 3.0 of the DEEM there are five or six possible responses for each question. They are *A*, *B*, *C*, *D*, *E*, and "*omit*". Obviously, a question with four given answers would use *A-D*, and a question with five given answers would use *A-E*. *Omit* indicates that a student did not answer that question and moved on to the next one or, in the case of the last few questions, simply did not finish the exam. Remember, *omit* is not necessarily a wrong response or an admission by the test taker of a complete lack of knowledge on a particular question. A number of stems direct the student to skip that question if that person gave a particular response on the previous question.

The value of the frequency table is that it provides the test creator a means to extract information regarding certain trends and patterns in the way students answer a particular question, or set of questions. The entire frequency table is located in Appendix A (Table A–I). As an example, in Version 3.0 of the DEEM the frequency table for items 40-43 is shown in Table 48–I. The number in each cell is the percentage of students who chose that option. The highlighted cells are the correct responses (also noted in the far, left column). The boldfaced numbers indicate the option that drew the most responses.

Table 48–I: Frequency Table for Questions 40-43 of Version 3.0 of the DEEM

Key	Item	Frequency Table											
		A		B		C		D		E		Omit	
		pre	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*
D	*40*	**37.97**	10.82	2.28	0.26	9.11	1.58	33.42	**76.25**	14.18	10.55	3.04	0.53
E	*41*	17.97	17.41	4.05	0.00	6.84	0.79	15.70	14.25	**52.41**	**67.55**	3.04	0.00
E	*42*	**44.30**	**42.74**	3.80	1.32	8.10	3.17	21.01	26.65	18.48	25.59	4.30	0.53
E	*43*	12.15	10.55	4.30	0.26	8.61	2.37	14.94	11.35	**55.70**	**75.20**	4.30	0.26

Again, in Chapter 13 the investigator will describe how this table and other data were used in the creation of the final version of the exam. However, a quick look at this chart is warranted so one can see how it may be employed. (The reader may refer to Appendix C for a complete copy of Version 3.0 of the DEEM.) It should be obvious that students have a firm grip on questions 41 and 43, both in the pre-test and in the post-test. In the pre-test, question 40 addresses their misconceptions about the role the direction of the magnetic field plays in the force on a moving charged particle. Most students resolved those issues as indicated by the post-test frequencies. Question 42, however, clearly indicates students did not learn to associate the role the direction of the electric field plays on the magnitude of the force on a charged particle. There is some latent belief that the maximum force on a charged particle is in the direction of the field.

Another application of the frequency table is to view how upper and lower scorers on the exam answered a particular question. The upper and lower groups are the top twenty-seven percent and the lower twenty-seven percent, respectively. The investigator chose the upper and lower twenty-seven percent because they are associated with the item discrimination index, *D*. Section 9.4 will discuss the details of this index. The entire upper/lower frequency table (post-test only) is located in Appendix A (Table A-II). As an example, the post-test's upper/lower frequency table for items 40-43 are in Table 48–II.

Table 48–II: Frequency Table for Upper and Lower Scoring Students on Version 3.0 of the DEEM

		Frequency Table (post)											
		A		B		C		D		E		Omit	
		upper	*lower*	*upper*	*lower*	*upper*	*lower*	*upper*	*lower*	*upper*	*lower*	*upper*	*lower*
D	*40*	4.26	17.02	0.00	1.06	1.06	2.13	**92.55**	**64.89**	2.13	13.83	0.00	1.06
E	*41*	4.26	27.66	0.00	0.00	0.00	1.06	6.38	29.79	**89.36**	**41.49**	0.00	0.00
E	*42*	29.79	**40.43**	2.13	3.19	0.00	8.51	19.15	31.91	**46.81**	15.96	2.13	0.00
E	*43*	1.06	22.34	0.00	1.06	0.00	6.38	4.26	18.09	**94.68**	**51.06**	0.00	1.06

One can see, in light of Table 48–II, that all four items act as good test questions in the post-test. One thing to note is that on question 42 a small percentage of high scorers on the exam skipped this item. This is an unexpected result, and it shows the power of this simple analysis tool. An instructor may want to follow up on this to determine why some of the better students skipped this item.

9.3 Item Difficulty

Item difficulty, *p*, is essentially the average score for a particular exam question. On a dichotomously scored exam, this is the total number of examinees who answered the item correctly divided by the total number of examinees. Item difficulty ranges from 0.00 to 1.00. A high item difficulty – above 0.5 – indicates that a majority of the students answered the question correctly. In other words, the higher the value of the item difficulty the *easier* the test question.

The target item difficulty for item *i* is computed from the formula:

$$p_i=(1+g_i)/2, \qquad (48:1)$$

where g_i is the *chance of success level* (csl) for that item (Gregory, 1992). This equation states, ideally fifty percent of the students (1/2) know the answer, however there is also some percentage (g/2) that will guess the answer correctly. This target must be kept in mind when creating a test to help maximize the variance, which, as discussed in Chapter 6, will increase the reliability. The variance, σ_i^2, of each item is related to the item difficulty by:

$$\sigma_i^2=p_iq_i, \qquad (48:2)$$

where $q_i=1-p_i$.

Most of the questions on the DEEM are five-choice questions, so the csl is 1/5 = 0.20; therefore, the target item difficulty is (1.0+0.2)/2=0.60. For a four-choice question the target item difficulty is 0.63, and for a six-choice question it is 0.58. Version 3.0 of the DEEM has an average item difficulty of 0.53 (post-test). Therefore, the average variance is $\sigma_{ave}^2 = 0.53\ (1-0.53) = 0.25$. Item difficulties ranged from 0.07 to 0.99. To improve reliability, a test designer must consider improving or eliminating those items lying near the extrema.

9.4 Item Discrimination

The item discrimination index indicates how a particular test question discriminates between high scorers and low scorers. Ideally, one desires highly discriminatory exam questions. That implies that high scorers on the exam tend to answer the question correctly, while low scorers tend to answer the question incorrectly. The item discrimination index is given by the equation:

$$D=p_u-p_l, \qquad (48:3)$$

(Gergory, 1992). The variable p_u is the percentage of high exam scorers who answered the item correctly; p_l is the percentage of low exam scorers who answered the item correctly. The item discrimination index has a range of $-1 \leq D \leq 1$. An item discrimination of 1 indicates that all the top scores answered the question correctly, while all of the low scorers answered it incorrectly; an index of -1 signifies the opposite scenario. For example, if 80% of top scorers on an exam answered an item correctly and only 30% of low scorers answered that item correctly, then the item discrimination for that question is D = 0.80-0.30 = 0.50.

What constitutes the upper portion and lower portion of the group depends on the sample size and distribution. A straightforward approach might simply be to compare the upper fifty-percent to the lower fifty-percent. The literature suggests other alternatives that may provide item discriminations that indicative of the actual discriminating power of an item. Gregory puts forth that the upper and lower ranges should be somewhere between 10 and 30 percent of the total number of examinees and, “If the exam scores are normally distributed, the optimal comparison is the highest scoring twenty-seven percent versus the lowest scoring twenty-seven percent of the examinees.” He further suggests that “if the distribution of total test scores is flatter than the normal curve, the optimal percentage is larger, approaching thirty-three percent” (Gregory, 1992). Other texts follow similar lines (Crocker and Algina, 1986). Based on results of the last chapter, the investigator chose to use the upper twenty-seven percent and the lower twenty-seven percent to calculate the item discrimination index.

This raises the issue of what value of *D* the designer should consider strong. Clearly, a negative value for an item is highly undesirable. Any question that selects favorably for examinees that are otherwise doing poorly on the test should be examined very carefully. What is a positive *enough* value of *D* to make an item obviously worth keeping? The literature varies somewhat on this issue. Summaries of two opinions are provided in Table 48–III.

Table 48–III: Summary of Item Discrimination Indices

Gregory		Crocker and Algina	
1.00	Ideal item but never achieved	0.40 ≤ D	The item is functioning quite satisfactorily
0.60	Excellent item but rarely achieved	0.3 ≤ D ≤ 0.39	Little or no revision is required
0.26	Very good item with high difficulty	0.20 ≤ D ≤ 0.29	The item is marginal and needs revision
0.00	Poor item that should be revised	D ≤ 0.19	The item should be eliminated or completely revised

Clearly, Gregory has laid-out a much coarser and liberal scale, while Crocker and Algina have demanded higher standards. Both texts, however, state these parameters are merely suggestions that must be taken in the context of other item analysis techniques and of the exam as a whole.

The average value of the item discrimination on Version 3.0 of the DEEM was 0.32 (post), with a maximum value of 0.62 and a minimum value of -0.09. According to Gregory's criteria, the average item is better than "Very good;" according to Crocker and Algina, "Little or no revision is required." In either case, the average item on the exam is discriminatory. As a more concrete example and one that is to the extreme, attention can be directed towards item 57. In this question, the pre-test item discrimination index was 0.21; however in post-test, this item's discrimination index dropped to -0.09. This is a most interesting result, particular in light of the fact that a similar item (#54) did not suffer the same fate. Chapter 13 will describe an item-by-item analysis; this and other weak questions will be evaluated and updated.

9.5 Correlation Coefficients

Correlation coefficients describe the linear relationship between two sets of data. These coefficients have values between 1 and -1. A correlation coefficient near 1 indicates a strong, positive relation between two sets of data, a coefficient near 0 indicates no relationship, and a value near -1 indicates a strong, negative relationship. Since test creators frequently employ correlation coefficients in a variety of item analysis tasks, they are cast into many different, yet equal, forms. The investigator's item analysis of the DEEM's employed the Pearson Product Moment and Point Biserial (or more appropriately, the modified Point Biserial).

9.5.1 Pearson Product Moment

The Pearson Product Moment Correlation Coefficient, or simply the *correlation coefficient*, describes the linear relationship between two sets of data in a most obvious form. The equation to find the relationship between data set *X* and data set *Y* is

$$\rho_{xy} = \frac{\sum_{i=1}^{N} x_i y_i}{N\sigma_x \sigma_y} . \tag{48:4}$$

Here x_i and y_i are deviation scores, *N* is the number of examinees, and σ_x and σ_y are the standard deviations of X and Y, where X and Y could be a test item, test score, etc. The deviation score is defined as $x_i = X_i - \mu_X$ where X_i is the score of an individual test taker on item X and μ_X is the average score of all the test takers on item X.

One way for a test creator to explore items is to examine the correlation among particular questions that deal with the same topic. To do this, the investigator grouped items into three levels: "primary", "secondary" and "tertiary." Items grouped together as primary were ones that not only dealt with the same topic but also were located adjacent to one another on the exam. For example, items 62-65 all dealt with induction of a current in a wire loop, so all possible pairs of items in that group were labeled as having a primary-level relationship. Item sets labeled secondary dealt with the same topic, but did not probe the exact same aspects of the concept as did the primary items. Furthermore, sets of items labeled secondary were not located adjacent to one another on the exam. As an example, questions 69 and 70 investigate the relationship between relative motion and induction. The correlation between either of those two items with an item from numbers 62-65 are considered secondary-level. Tertiary-level pairs are similar to secondary-level but they focus on yet another aspect of a particular concept. For example, questions 38 and 39 also dealt with induction in wire loops; however, they are considered to have a tertiary-level relationship when paired with items from numbers 62-65.

The investigator's analysis of the DEEM employed the Pearson Product to find correlations between pairs of questions determined to have primary, secondary or tertiary level relationships. The average correlation, ρ_{ave}, between items having a primary relationship on the pre-test was ρ_{ave}=0.23 and on the post-test was ρ_{ave}=0.28. The average correlation for *any* two items (this includes obviously related items, but excludes items correlated to themselves) on the exam was ρ_{ave}=0.05 (σ=0.10) on the pre-test, and ρ_{ave}=0.07 (σ=0.11) on the post-test. Items considered having secondary and tertiary relationships had average correlations of approximately 0.06 (pre) and 0.11 (post). When one calculates the average and standard deviation of correlations between pairs of items in a set deemed to be primarily related, the correlation between an item and itself must be omitted. Failure to do so would serve only to create a false sense that a group of items is more positively correlated than they are, and perhaps artificially shrink the size of the standard deviation.

As a concrete example, questions 62-65 have an average correlation with each other of 0.04 (pre) and 0.21 (post). As was previously mentioned, the investigator considers these items to have a secondary-level relationship with questions 69 and 70. The average correlation among the pairs of questions between items 62-65 and 69 and 70 is 0.03 (pre) and 0.11 (post). The entire table of correlation coefficients can be found in Appendix A, Table A–IV.

9.5.2 Point Biserial Correlation

The Point Biserial Correlation (PBC) is a correlation between an item scored as 0 or 1 (wrong or right) and a continuous variable, such as total test score. The investigator used it here to evaluate the correlation between student responses on an item and the score the student received on the exam. In its most basic form the equation for the PBC is:

$$\rho_{pbis} = \frac{\mu_+ - \mu_x}{\sigma_x}\sqrt{p/q}\,, \qquad (48:5)$$

where μ_+ is the mean exam score for those that answered the item correctly, μ_x is the item's mean score, σ_x is the standard deviation of the exam scores, p is the item difficulty, and $q=1-p$. This equation must be used with caution as the score for the item is included in the total score. Crocker and Algina suggest a modified point biserial correlation equation to remedy this problem:

$$\rho_{i(x-i)} = \frac{\rho_{xi}\sigma_x - \sigma_i}{\sqrt{\sigma_i^{\,2} + \sigma_x^{\,2} - 2\rho_{xi}\sigma_x\sigma_i}}\,. \qquad (48:6)$$

Here ρ_{xi} is the point biserial correlation and σ_i is the standard deviation of the item. Crocker and Algina also indicate that once the number of items is greater than 25, this form does not necessarily need to be employed. On the DEEM this was not found to be the case. The average difference between the results from application of equation 48:5 and equation 48:6 was -0.04. In other words, the average modified point-Biserial correlation (MBP) was lower then PBC, as expected. Furthermore, on average it was 0.04

lower. Since the authors did not offer a suggestion as to when the difference should be considered significant, it would be wise for the investigator to err on the side of caution and report the MPB. After all, the MPB is a closer representation of the actual correlation between an item and the test score. (The full table of MBPs can be found in Appendix A, Table A-II.)

9.6 Item Reliability Index

Another way for a test designer to examine, and improve, the internal consistency of a test is to use the item reliability index, *IRI*. This parameter is simply the product of the Point Biserial Coefficient of a particular item and that item's standard deviation:

$$\text{IRI}= \sigma\rho_{pbis} \tag{48:7}$$

(Crocker and Algina, 1986).

Since this is the product of a correlation coefficient and a standard deviation of a dichotomously scored exam, the IRI can fall between ±0.5. The IRI and the Point Biserial Coefficient must always have the same sign since the standard deviation can not be a negative number.

As a hypothetical example, if an item correlates well with the overall score, ρ_{pbis}=.8, and nearly everyone answered it correctly, σ=0.05, then that item will have a low IRI (0.04). But, if half the examinees answered that item correctly, σ=0.5, then the item has a high IRI (0.4).

On Version 3.0 of the DEEM, the values of the item reliability index run from 0.23 down to -.05 on the post-test. Similar results were obtained for the pre-test (0.22 down to -.02). This suggests that many of the items either did not have large variances and/or items did not correlate well with exam scores. This should not be alarming. The DEEM covered a wide range of topics, so it should not be surprising that individual items did not correlate well with overall scores.

9.7 Item Tables

The investigator created the final two techniques independent of any other work; however, it is doubtful they are novel methods for obtaining information about exams. The first technique is the Single-Item Movement Table (SIMT). It tracks how student response patterns changed on a single item from pre-test to post-test. The second technique is the Dual-Item Frequency Table (DIFT). It tracks student response patterns on two separate questions either on the pre-test or the post-test.

The value of these two techniques lies in their ability to help one evaluate the effectiveness of distractors by investigating which ones work in the pre-test and the post-test, and how students move among those distractors from pre-test to post-test.

9.7.1 Single-Item Movement Table

The Single-Item Movement Table (SIMT) compares how students answered a particular question in the pre-test and compares it to the post-test. Specifically, it breaks down the number of people who answered *A*, *B*, *C*, etc. on a particular item in the pre-test and puts those numbers into “bins” associated with how students answered the question in the post-test. An example of the SIMT for item 40 is provided in Figure 48–I.

Figure 48–I: An Example of the Single-Item Movement Table

Question 40	*Correct response* D	Pre					
	Post	*A*	*B*	*C*	*D*	*E*	*omit*
	A	3.2	0.0	1.4	3.7	0.9	0.9
	B	0.3	0.0	0.0	0.0	0.0	0.0
	C	0.9	0.0	0.0	0.3	0.6	0.0
	D	**30.0**	1.4	7.2	26.2	10.7	0.9
	E	3.5	0.9	0.3	4.3	1.4	0.6
	omit	0.0	0.0	0.0	0.0	0.6	0.0

A quick explanation of Figure 48–I is warranted. Obviously, the question number and the correct answer are in the upper left-hand corner. The highlighted column and row heading reiterate the correct response. The matrix element with the maximum value is noted in boldface. The first row displays what percentage of students answered *A* on the post-test and *A*, *B*, *C*, *D*, or *E* on the pre-test, the second row displays what percentage of students answered *B* on the post-test and *A*, *B*, *C*, *D*, or *E* on the pre-test, and so on.

Item #40 provides a case study for a typical question. As any instructor would hope, most of the students moved to choice *D* on the post-test. On the pre-test all the responses except *B* proved to be decent distractors, as they captured more than a few students' interest. Most students answered either *D* or *E* on the post-test.

9.7.2 Dual-Item Frequency Table

The Dual-Item Frequency Table (DIFT) compares student response patterns for two questions, for both the post-test and pre-test. Creation of the table is straightforward and similar to the SIMT.

Figure 48–II: An Example of the Dual-Item Frequency Table

Question &	*Question*
69	70
Correct responses	
A	A

			70					
			A	*B*	*C*	*D*	*E*	*omit*
Pre	69	*A*	7.5	9.8	0.9	1.4	0.6	0.0
		B	**20.5**	5.8	3.5	4.0	1.4	0.0
		C	2.3	4.6	3.5	0.9	0.3	0.3
		D	3.5	2.6	1.7	2.3	0.9	0.0
		E	1.4	0.9	0.9	0.9	7.2	0.0
		omit	0.0	0.0	0.0	0.0	0.0	10.1

			70					
			A	*B*	*C*	*D*	*E*	*omit*
Post	69	*A*	15.3	17.6	0.3	1.2	2.6	0.0
		B	**19.6**	16.1	0.9	2.9	1.2	0.0
		C	0.9	2.3	3.5	0.6	0.3	0.0
		D	1.7	2.3	0.6	1.4	0.6	0.3
		E	0.0	0.6	0.0	0.3	3.2	0.0
		omit	0.0	0.0	0.3	0.0	0.0	3.7

Figure 48–II requires a brief description of its elements. Obviously, this particular table is comparing the pre-test and post-test student response patterns for items 69 and 70. The layout and calculations are similar to the SIMT. The first row in a table indicates how many students answered *A* in the question number indicated on the left-hand side of the box in the upper left-hand corner. Each element in that row indicates how many students answered *A*, *B*, *C*, etc. in the question number on the right-hand side of the box in the upper left-hand corner. A highlighted column or row heading indicates that it contains the elements connected with the correct answer. The correct answers are also noted in the box in the upper left-hand corner. (For this particular example the correct answer to both questions is *A*, but that need not be the case.) A quick glance at a row or column indicates whether or not a particular item was a good distractor. If the sum of the

elements in a row or column is low, then the examinees did not choose that particular response and the investigator must consider its future on the exam.

For items 69 and 70, one can see that on the pre-test many students chose answer *B* (a reasonable choice) but then switched their answer to *A* in item 70. The second most popular combination was students answering *A* in item 69 and moving to *B* in item 70. These two patterns make up almost a third of the response patterns for items 69 and 70 on the pre-test. This expresses the fact that students felt that somehow the situation was reversed when the field moves toward the rings as opposed to when the ring moves toward the field. In slight contrast to these results are the post-test patterns.

On the post-test, clearly most students knew there would be a current induced in the wire, over two-thirds of the students indicated in items 69 and 70 that there would be a current induced. Over thirty-seven percent, however, wanted to say the situations were reversed in 69 and 70 and only thirty-one percent wanted to say that the current induced is the same. A subset of the latter choice pattern (fifteen percent of the whole group) got both questions completely correct.

An instructor can delicately probe students' understanding of questions using both the SIMT and DIFT. The investigator used them mainly to analyze questions and sets of questions and to weed out poor distractors.

GENDER BIASES IN TEST ITEMS

Up to this point, the investigator's statistical analysis of the exam has used the entire population of test takers for the pool of data. Individual scores, or groups, have not been separated to determine how they preformed relative to another group. One of the major challenges a test designer faces is trying to prevent items, or the whole test, from discriminating against one demographic group as compared to another. This chapter will address the issue of items that may or may not favor one gender versus the other.

Sections 49) and 50) introduce straightforward methods of detecting item biases: phi coefficient and differences in the item difficulty index. These two sections set the stage for more advanced item bias detection methods: sections 51) and 52). Section 51) introduces advanced chi-square techniques, while section 52) expands on the item difficulty index to find item discriminations. Section 10.1 will present the results of an f-test comparing the gains and post-test scores for males and females. The investigator used this test to determine if there was any significant variations between males and females overall test gains or scores. The last section will be an overview of all the results of this chapter to determine if any items fail a number of tests for bias. If that occurs, the investigator will offer remedies to alleviate the problems.

63) Phi Coefficient

The Phi Coefficient is used when scores from a dichotomously scored item are to be compared with a dichotomous criterion, such as gender (Crocker and Algina, 1986). The

Phi Coefficient is simply another method of calculating the Pearson Product Moment Correlation as discussed in Chapter 11. The steps taken to find phi coefficient are superficially different, and they do provide some insight that a straightforward application of the Pearson Product does not. Also the ease of calculation makes this a logical choice to compare a criterion, such as gender, to item responses.

The simplest way to calculate the Phi Coefficient is by utilizing the table below.

Table 10–I: Phi Coefficient Matrix

		Item *i* +	Item *i* -
Item *j*	-	*a*	*b*
	+	*c*	*d*

Table 10–I could be arranged so *Item i* is gender – "+" being female and "-" being male – and *Item j* is the item number in question – "+" is the number of correct responses and "-" is the number of incorrect responses. Summing the columns yields the number of females and males that took the exam and summing the rows yields the number of incorrect and correct responses to the item in question. To calculate the Phi Coefficient, ρ_{phi}, using the Table 10–I one employs:

$$\rho_{phi} = \frac{bc - ad}{\sqrt{(a+b)(a+c)(c+d)(b+d)}} . \qquad (10:1)$$

Since the phi coefficient is a correlation coefficient, it can have values between ±1.00. A strong bias towards females would be a value of $\rho_{phi}>0.6$, and a strong bias towards males would be a value of $\rho_{phi}<-0.6$. For the DEEM the average ρ_{phi} was 0.05 ($\sigma=0.07$) on the pre-test and 0.06 ($\sigma=0.08$) on the post-test, suggesting only the slightest bias towards the male test-takers. Table B–I, located in Appendix B, is a full listing of phi-coefficients for each item.

The investigator set an arbitrary criterion to pick out modestly biased items. Any item with $|\rho_{phi}|>0.2$ was flagged and considered a possible candidate as a biased question. Only three items on the post-test are considered biased using this technique (2, 43, and 53). All three were biased toward the males. No item had $|\rho_{phi}|>0.3$.

64) Item Difficulty Difference

The calculation of the item difficulty for the entire pool of students was discussed in Chapter 11. Briefly, the item difficulty for a dichotomously scored exam question is simply the number of examinees who answered the question correctly divided by the total number of examinees who answered the question (in effect, the average). The item difficulty has a range from 0 to 1. An item difficulty of 0 means the entire group answered the question incorrectly; an item difficulty of 1 means the entire group answered the question correctly. To use this statistic to uncover item discriminations, one calculates the difference between the item difficulties for females and males. The resulting number ranges between ±1. Here, a value of 1 implies that all the females answered the question correctly and all the males answered it incorrectly; a value of -1 implies the opposite scenario. The results of this calculation for Version 3.0 of the DEEM can be found in Appendix B, (Table B–I).

The average difference in the item difficulties for the DEEM was $\Delta p= -0.05$ ($\sigma =0.08$) on the pre-test and $\Delta p= -0.06$ ($\sigma=0.09$) on the post-test. These results suggest that the items were slightly easier for the male students. When one considers how small the difference is compared to the magnitude of the standard deviation the discrepancies seem negligible. Using a similar criterion for the phi coefficient (now $|\Delta p| > 0.2$), items that where modestly biased were again flagged and examined. On top of the three already picked out by the phi coefficient, numbers 12 and 41 also showed some bias towards the male test takers. Again, no item showed a value of $|\Delta p| > 0.3$.

65) Chi-Square Techniques

As one might expect, it is possible for a test creator to employ chi-square techniques to further analyze the questions for item bias. One can adapt chi-square techniques in many ways to meet this end. The phi-coefficient (discussed earlier) is one such adaptation. Camille (Camille, 1979) and Scheuneman (Scheuneman, 1979) have put forth more advanced techniques.

For one to use either of the approaches, it is necessary to first divide the group into subgroups, g_i. In the interest of testing for gender bias, these subgroups are obviously females (g_1) and males (g_2). Following that, the designer must divide both subgroups into intervals, *j*. In general, these intervals are ranges of total exam scores. (The following paragraph will provide details regarding the appropriate choice of the number of intervals and their sizes.) Next, one must calculate the number of people in each subgroup in each interval ($N_{g,j}$). Finally, one calculates the percentage of all the people, from all the subgroups, in the interval that answered the question correctly ($P_{.j}$) for a given item.

The designer's choice for the ranges of the intervals, *j*, are somewhat arbitrary. For example, on the DEEM it is conceivable to call every possible raw score an interval, making the total number of intervals seventy. This would be a mistake for two reasons. First, the number of examinees in each subgroup in each interval would be very small or possibly zero, making the analysis very questionable. Second, and most importantly (as will be demonstrated in a moment), if the number of examinees in an interval is zero, then calculation is invalid altogether. So, one wants to maximize the number of intervals (which corresponds to the number of degrees of freedom in the chi-square analysis), but at the same time one wants to keep the number of examinees in each interval greater than zero, and preferable much greater. “How much greater?” is a question the authors of several papers dealing with this topic have addressed. Scheuneman points out that for each interval there must be at least one incorrect score or else that interval is not valid. She later goes on to suggest that the number of correct responses in each interval should be at least ten to twenty, and the expected number of correct responses ($N_{g,j}P_{.j}$) should be at least five (Scheuneman, 1979). Another author offered similar suggestions for Camilli’s technique (Ironson, 1982).

With these suggestions in-mind, the investigator divided the post-test of the DEEM Version 3.0 into 2 intervals (*j*=2). The range of exam scores for interval 1 was 0-38, and for interval 2 was 39-70. The investigator kept the number of intervals small to ensure that most of the items would pass all of the aforementioned criteria. As it turned out, only a few of the more difficult questions did not pass all of the parameters. This is the reason the investigator carried out other methods of item bias detection. Chi-square techniques have the potential to be the strongest, but, unless the pool of examinees approaches 1000, some items may not be suitable.

For the DEEM, the investigator used the signed Scheuneman statistic:

$$\chi^2_{s_{signed}} = \sum_{j=1}^{J}\sum_{i=1}^{g} \frac{\left(O_{ij} - P_{.j}N_{ij}\right)\left|O_{1j} - P_{.j}N_{ij}\right|}{P_{.j}N_{ij}} \quad (10:2)$$

The investigator used this particular technique because of its intuitive approach and ease in interpreting its results. Equation 10:2 closely resembles a standard equation for chi-square analysis. This formula takes the number of people in a subgroup's interval who answered the question correctly and subtracts off the number one would expect to answer the question correctly, in that interval, based on the percentage of the overall population that correctly answered the question. That difference is multiplied by the absolute value of itself (hence the "signed Scheuneman statistic") and then divided by the expected value.

Despite the investigator breaking each population into only two intervals, there were several items that failed to meet the criteria outlined above. Six items (10, 13, 31, 45, 47, and 58) failed the test for having every single person in an interval answer the question correctly. Two items (26 and 36) had an expected outcome of fewer than five. Additionally, thirteen items (plus a borderline item) failed the softer criteria that there should be at least ten to twenty people in an interval. These were items 19, 24, 27,39, 42, 50, 51, 57, 59, 61, 63, 66, and 68. Item 23 was on the borderline. The remaining forty-eight items all showed a remarkable lack of item bias. None of the items was even close to the value of |5.99| that would indicate there is some gender bias in the items (Minium, 1978). This is the value for a 95% confidence level for a chi-square analysis with two degrees of freedom.

66) Delta Plot

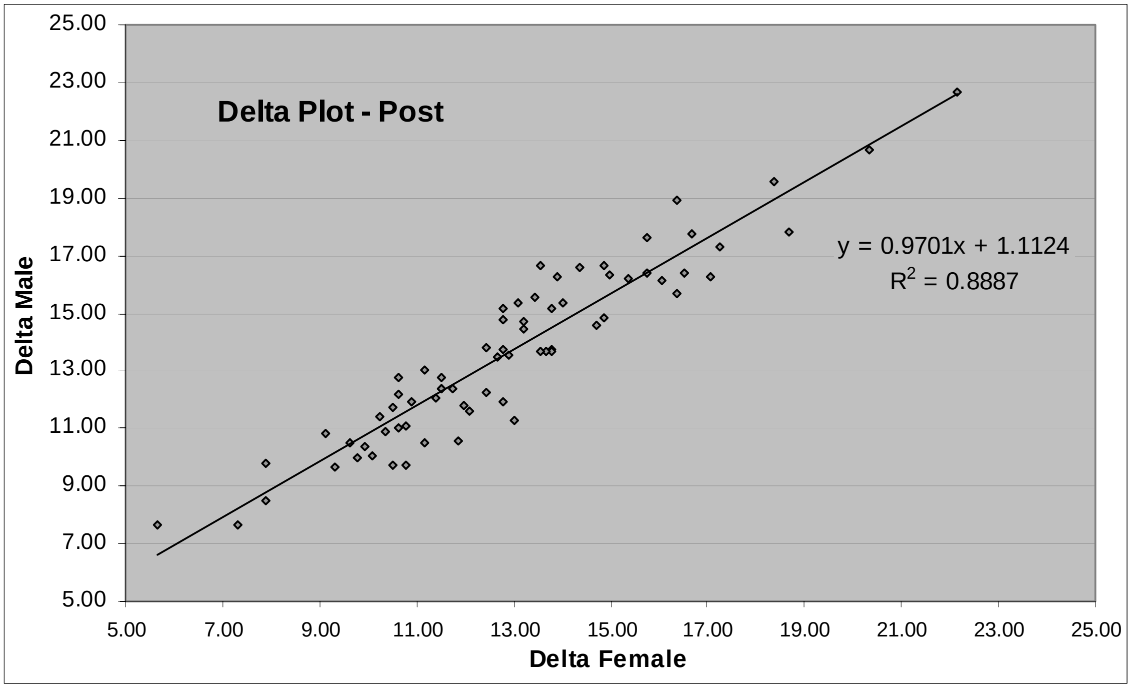
The Delta plot is a graphical measure of the item difficulty for one subgroup versus another. The measure of item difficulty that is used to create a Delta plot is given by the expression:

$$\Delta_g = 4z_g + 13, \qquad (10:3)$$

(Angoff and Ford, 1973), where z_g is value of the abscissa corresponding to the proportion of examinees that answers the item correctly for a normalized distribution (μ=0 and σ=1), and *g* is the index corresponding to the item number. The coefficient, *4*, and constant, *13*, are an arbitrarily chosen standard deviation and mean, respectively, for converting the data to a normal deviate (Angoff, 1982). As a hypothetical example of this formula: if 70% of the students answered question 1 correctly, then z_1=0.53; therefore, Δ_1=15.15. If 50% of the students answered question 2 correctly, then z_2=0.00; therefore, Δ_2 = 13. And, if 40% of the students answered question 3 correctly, then z_3=-0.25; therefore, Δ_3 = 12.0.

The Delta plot is a fast way to scan for item bias as it provides a convenient set of data to graph. One detects item biases by dividing the entire pool among the subgroups, in this case females and males. One then calculates $\Delta_{g,female}$ and $\Delta_{g,male}$ for each and every item. It is then a straightforward matter for one to plot one subgroup versus the other to see if any items deviate from the group. A regression line increases the visual power of this technique. Figure 10–I is such a plot for post-test of Version 3.0 of the DEEM.

Figure 10–I: Delta Plot of Post-Test Scores for Males vs. Females



Excel included the regression line (the thin, black line) and the equation of that line along with the correlation (R=0.88). It is obvious, both visually and numerically, that there is a strong correlation between the Delta scores of males and females. Very few items stray far from the regression line.

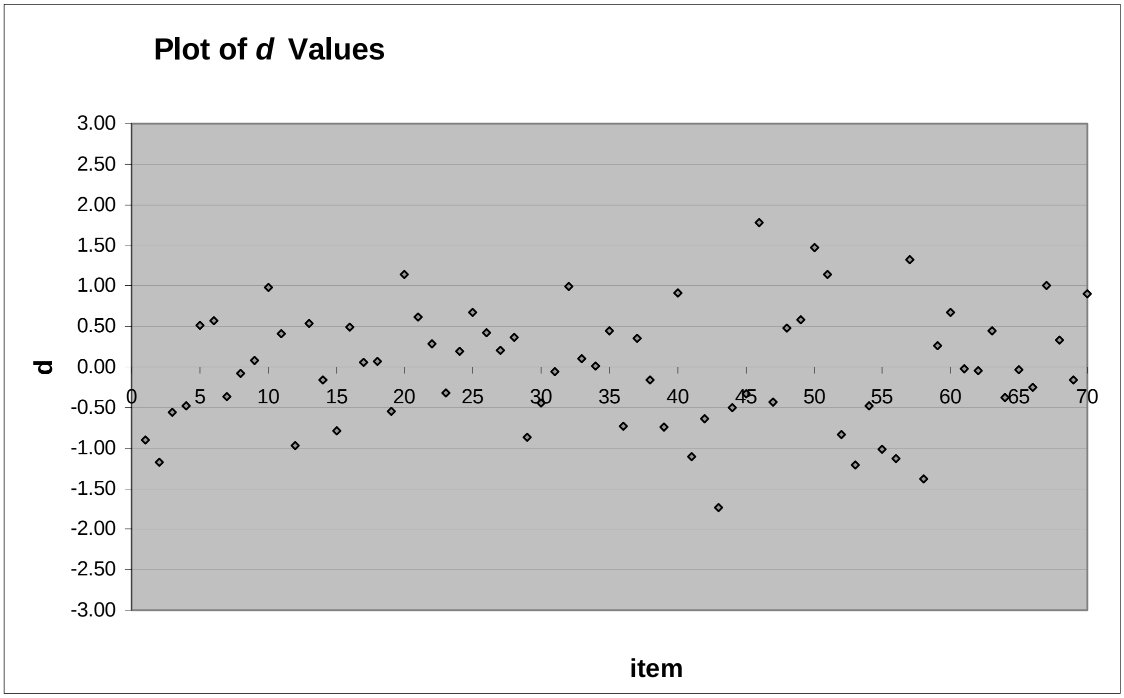
One encounters a small problem with this plot if a particular item is far removed from the regression line. It is not possible, by simple visual inspection, for a designer to determine what item is the one that is far removed. To determine the item number of a data point that is relatively far from the regression line, one has to obtain that point's *x* and *y* coordinates from the graph. Then one must go back to the data set and sift through each item until the one that corresponds to the previously determined *x* and *y* values is located. Although this may sound tedious, even with a set of data as large as 70 points, it was not difficult for the investigator to manually determine item numbers not bunched with the rest of the data set. However, to make this easier, and to remove some of the guesswork about whether one items' distance from the regression line is greater than another's, one can simply calculate the distance, d_g, of each item to the regression line. The equation to do this is:

$$d_g = \frac{m\Delta g_1 - \Delta g_2 + b}{\sqrt{m^2 + 1}}, \qquad (10:4)$$

where Δg_1 are the x-axis values, Δg_2 are the y-axis values, and *m* and *b* are the slope and the y-intercept of the Delta plot's regression line, respectively. The test designer can now plot the d_g's versus item numbers to ensure quick location of those items whose difference in gender difficulties vary greatly from the rest of the group. The author who put forth this procedure did not offer specific values of d_g that would indicate a real bias (Angoff, 1982).

Figure 10–II is a plot of the d_g values for the DEEM (post-test). It is reassuring to see nearly an equal number of positive and negative values, indicating that about the same number of items were slightly biased towards women as were biased towards men. When one couples Figure 10–I and Figure 10–II it is clear that none of the items seem to strongly bias against males or females. And only two items (43 and 46) warrant any attention whatsoever.

Figure 10–II: Plot of *d* Values vs. Item Number



10.1 t-Test on Percent Gain

The investigator preformed a t-test on the percent gains for males and females to determine if one gender group was more likely to have higher gains. The females had an average percent gain of 24.1%, with a standard deviation of 15.4%; the males had an average percent gain of 27.3%, with a standard deviation of 20.9%. The result of the t-test was T= -1.54. This value falls within the critical range of $-1.97 \leq t \leq 1.97$ for a 95% confidence level, impling that neither males nor females had significantly higher gains.

67) Summary of Results and Modifications

Very few of the items showed any large bias towards one gender or the other. The underlying reasons for this may be very complex. The investigator put forth one mundane reason, however, that may explain why electricity and magnetism enjoys this seeming lack of bias, as opposed to mechanics, which is permeated with it. Many examples in mechanics are related to sports, the military, or automobiles. These are all areas that have, for one reason or another, been traditionally dominated by men. Therefore, it is possible that males are more likely to be at ease with and, consequently, perhaps better at doing problems related to areas they are more familiar with than their female counterparts. Introductory electricity and magnetism is, relativity speaking, a “genderless” subject in the sense that instructors and textbooks rarely use examples from day to day experiences, and even less frequently use examples from areas such as sports, the military, automobiles, etc. Moreover, most of the examples are very abstract. It is possibly this very academic and abstract reasoning breaks down the typical gender differences and puts everyone on the same footing. It is assumed that few people, male or female, think about things such as vector or scalar fields before they are faced with it in an electricity and magnetism course. This issue seemingly “levels the playing field” for both men and women in electricity and magnetism. With that said an overview of the questionably biased items will be examined to clear up any problems that may exist.

Item #2 was biased towards males more than most, and the bias was detected in all the techniques; however, items 1-4 all showed a bias toward males. The reader should note that the set of items following 1-4 (5-9) showed a mixed bias. It does not seem it was the diagrams that caused any problems. Inspection of the other problems that relate to force on a charged particle in an electric field (42 and 43) shows that they also were biased towards the male test takers.

The other questions that showed some biases were items 52-62. These were questions probing students’ grasp of field superposition. Some of these questions showed some biases towards the male test takers, although there is no consistent pattern.

In conclusion, no modifications were made to the DEEM based on the results of gender biases. The only concept that showed an even slightly consistent bias toward male examinees was the force on a charged particle in a uniform electric field. When one considers that none of the tests for bias turned in any results that were statistically significant at all, a test designer would be remiss to make any changes based on the results.

68) CREATION OF VERSION 4.0

This chapter represents the apex of this work – the completion of the exam. Subsection 55) through subsection 78) uses Version 3.0 as a basis for the final version of the exam. Version 3.0 can be found in Appendix C and Version 4.0 is located in Appendix D. (The reader should note that after item 30 the numbering of Version 3.0 and Version 4.0 differs as a result of the investigator having deleted some items.) Each subsection introduces a small set of items, or an extended stem, and then discusses strengths, weaknesses, typographical errors, and expert review. From there, the subsection indicates what changes the investigator made to certain items or extended stems. The investigator made two changes to the exam as a whole and they will be mentioned here, briefly.

First, the exact layout of Version 3.0 is not the same one the students that took the exam saw. The investigator had to change the layout to accommodate the mandatory margins of the thesis. Second, in Version 1.0 the borders of the figures were a light shade of gray; in the later versions they were black. The investigator changed the figure borders back to light gray because it makes them less distracting and possibly prevents confusion about whether or not the edge of the figure is the edge of a box (Meltzer, 1998).

Section 79) will present a profile of the final version of the exam to demonstrate that all of the topics are covered equally and are uniformly spread out over the entire exam, and that the keys are evenly distributed among the options *A*-*E*.

69) Extended Stem for Items 1-4

The investigator changed the phrase "…the group of dashed arrows…" to "… the group of dotted arrows…" to better reflect the diagram. Two reviewers of Version 1.0 (Sherwood, 1996; and Styer, 1996) suggested that the phrase "electric field lines" be used in place of "electric lines of force." The investigator changed that phrase when creating Version 2.0.

70) Items 1-4

The first four items were very good questions. In particular, these item had the highest "primary-level" correlations of any other group of problems – 0.66 on the post-test. The only aspect the investigator changed was option *E* in items 3 and 4. They now read "Into the page" and "Out of the page" respectively, to account for the students that might incorrectly attempt a right-hand rule approach, and to rid the options of the phrase "None of these."

71) Extended Stem for Items 5-9

The investigator changed the phrase "…the group of dashed arrows…" to "… the group of dotted arrows…" to better reflect the diagram. Two reviewers of Version 1.0 (Sherwood, 1996; and Styer, 1996) suggested that the phrase "magnetic field lines" be used in place of "magnetic lines of force." The investigator changed that phrase when creating Version 2.0.

72) Items 5-9

These items are among the best on the exam. They were medium difficulty, yet strong discriminators, and they correlated well with the overall scores and with each other. Furthermore, the students found the distractors very appealing. The investigator only removed option *E*, “None of these,” in items 8 and 9 and reshuffled the options so option *A* is now an arrow pointing to the right. This new distractor is intended to reveal students that feel the force must be in the direction of the motion.

73) Extended Stem for Items 10-33

Two reviewers (Sherwood, 1996; and Styer ,1996) suggested that the thick, black lines “represent side views of long, *straight*, conducting wires.” The investigator changed that in the most recent version. Other reviewers (Lindberg, 1997; and Behringer, 1997) suggested that the use of a black dot to indicate a location in these items and the use of a cross in later items could be confusing. The investigator chose to indicate location with a cross. The exam reserves black dots to represents particles whose net charge is determined in the stem.

74) Items 10-15

One reviewer (Crowder, 1997) was concerned about the term “large” in the stem for item 12 (and presumably item 15). Since the stem indicates that large does not mean infinite, and the answer does not depend on one’s definition of how large “large” should be, the investigator saw no reason to change these stems. Another reviewer (Sherwood, 1996) suggested that “The phrase, ‘…if the dot is moved to a new position…,’ feels odd.” He suggested showing an additional diagram instead. The investigator did not add additional diagrams. However, the investigator changed the wording of this and related questions from Version 1.0 to Version 2.0, after another reviewer suggested it may be misleading (Fox, 1997). The investigator did replace the phrase “…if the dot is moved to a new position…” with “… if the cross were at a new position…” in the appropriate stems as well. (Note the change from “dot” to “cross”.) The investigator increased the distance between the charged particles and the crosses (formerly dots) in the figures and italicized words such as "greater" or "lesser" in the options. Finally, the investigator removed option *E*, "None of these," in items 11 and 14. They were poor distractors and did not offer any insight into the students' understanding of the items.

75) Extended Stem for items 16-27

The students seemed to have a general misunderstanding of the difference between magnitude and value. Since the exam was not designed to test these mathematical issues, the investigator included a brief summary of the difference between value and magnitude as an extended stem for items 16-23. It reads, "Items 16-23 ask questions about the value of a number, *not* its magnitude. For example, the value of 1 is greater than the value of -5." This extended stem is intended to clear up misunderstanding to improve the validity of the exam as a test of electricity and magnetism.

76) Items 16-23

The investigator changed some aspects of these problems. One reviewer suggested that the stems for items 16 and 20 imply that the answer will be a *discrete* value, but, since there is a lack of information, that is not possible (Lindberg, 1997). This made option *A* reasonable. To remedy this, the investigator changed the options for items 16 and 20 to:

a) The *exact* value can *not* be determined, but it must be *less* than 0.
b) The value is zero.
c) The *exact* value can *not* be determined, but it must be *greater* than 0.
d) The electrostatic potential is not defined at the cross.
e) There is not enough information to determine anything about the value.

The investigator removed option *E* from items 17 and 21. They were poor distractors and simply did not add anything to the interpretation of the students' responses.

77) Items 24-27

These four items did not prove to be the best test questions. Beside low discriminatory ability, their item reliability indices were low (<0.10) and correlated only modestly with each other (average primary correlation was 0.30). With the addition of the extended stem for these and the previous set of items, it is assumed that misconceptions regarding value and magnitude will be cleared up somewhat; this will only improve these items. There are only six items on the exam testing potential energy. If the investigator eliminated these four items then that would essentially mean abolishing the entire concept from the exam. This was not acceptable. To further clarify matters, the investigator changed the options for items 24 and 26 to read:

a) The *exact* value can *not* be determined, but it must be *less* than 0.
b) The value is zero.
c) The *exact* value can *not* be determined, but it must be *greater* than 0.
d) The electrostatic potential energy is only defined at specific location in space.
e) There is not enough information to determine anything about the value.

Finally, the investigator removed option *E* from items 25 and 27. They were very poor distractors and simply did not add anything to the interpretation of students' responses.

78) Items 28-30

On the pre-test these were all very strong questions. The distractors were particularly good. On the post-test, the students found these items relatively easy and, as a result, the items' statistics suffered. In an attempt to increase the variance of item 29, the investigator removed option *E* from all three items.

79) Items 31-33

The investigator removed these three items from the exam. They proved to have very poor discriminatory abilities, none of the distractors were plausible, and because the correct answer to 32 is “omit” it is difficult to machine grade. Their absence only improves the exam.

80) Items 34-37

The investigator considered these four questions to be novel and potentially very informative sources for students’ understanding of Maxwell’s equation. Typographical errors in the options prevented the items from reaching their full diagnostic potential. Even after the investigator fixed these, item 36 still had to be considered carefully. The diagram implies that there may be a current concealed by the box (option *A*) or there may be no current (option *C*). The correct answer is “Nothing can be said” (option *E*). A set of options where more than one answer is possibly correct only hurts the item, and the reliability of the exam. This item could be saved if the investigator chose to present it as a two-option item, but that would reverse the trend of multiple options on the exam. This question may prove a worthwhile discussion question; however, it is a poor multiple-choice test question, so it was removed.

81) Items 38 & 39

Item 38 proved to be a strong question; still, the investigator chose to modify the options to move them within the content domain of the exam and to reflect the overhaul of item 39. The investigator removed the parenthetical remark from item 38, as it does not add or detract from the plausibility of the options. Removing the phrase “alternating current” moved the option into the content domain of the exam, so the investigator replaced option *B* with the phrase, “A clockwise current is produced in the loop.” The investigator changed option *C* to read, “A counterclockwise current is produced in the loop.” Option *D* mentioned torque, which was also outside the content domain of the exam. The frequency table suggested that many students chose *D*, even post-instruction (16%), but very few choose *E* (2%) post-instruction, indicating that many students believe the loop will rotate. Consequently, the investigator rewrote *D* to read, “The loop will rotate clockwise in the plane of the page” and *E* to read, “The loop will rotate counterclockwise in the plane of the page.”

Item 39 was quite difficult, even for experts. Therefore, the investigator replaced item 39 with a question similar to 38. In this new question, the radius of the loop remained unchanged, but the strength of the magnetic field decreases. Thus, as in the previous problem, the flux through the loop decreases and a counterclockwise current is established in the loop. The options to the new item were the same as the revised item 38. The issue of charge separation was not included in these questions. The patterns of charge separation in a time-varying magnetic field can be quite complex, even in seemingly simple scenarios such as this (Aguirregabiria, et. al., 1992).

82) Items 40-43

The investigator updated these items only slightly. In all four questions, very few students chose *B* – five percent or less on the pre-test, and one percent or less on the post-test. The investigator removed option *B* and adjusted the figures and the options accordingly. The investigator then changed option *D* to read, “A and C result in a maximum force, but B does not.” Option *E* remained unchanged.

83) Extended stem for items 44-61

The investigator made this extended stem consistent with the one for items 10-33 by changing the symbols that represent positively and negatively charged particles to match the ones in the extended stem for items 10-33. The investigator changed the phrase "…long, conducting wires" to "…long, straight, conducting wires."

84) Extended stem for items 44-49

The investigator changed the word "charges" to the phrase "charged particles."

85) Items 44-49

Nearly all these items served the exam well. Items 44 and 45 had fine statistics; however, options *C* and *D* were not strong post-test distractors. The investigator did not remove them because those options were good pre-test distractors in items 47 and 48. Item 46 was a poor discriminator because a large number of the high scorers went for option *A*. This is the distractor that the investigator assumed most students would choose. Item 47 was a poor discriminator because it was too easy. The investigator did not remove this item because it makes items 48 and 49 possible. A comment by a reviewer on another set of items prompted the investigator to change the stems for items 46 and 49 (Crowder, 1997). The stem for item 46 now reads, "What is the relationship, if any, between the net electric field at the cross in question 44 and the cross in question 45?" The stem for item 49 was changed accordingly.

86) Items 50 & 51

These two items presented a challenge to the test creator. They both generated very poor statistics. They were hard problems with very poor discriminators (item 50 had a discriminatory value of -0.07), and the rest of their indicators were very poor. They did, however, correlate well with each other (0.67) and modestly well with other items dealing with potential (0.25), so that gave them some merit. All of the distractors proved worthy in item 50; in item 51, options *A* and *B* were poor distractors, but they will remain to keep the options in the two items parallel. Item 50 was one of the few questions that students did very well on in the pre-test (62% got it correct) and then did poorly on in the post-test (only 30% answered correctly). So as a diagnostic tool these items work fairly well. For that reason they remained on the exam. The investigator also rearranged options *B*, *C*, and *D* in both items to present a more logical order.

87) Items 52-57

Except for item 57, these items were very good test questions. The first five in this set were strong distractors and correlated well with test scores, each other, and items 28-30. Item 57 had similar problems to item 46. Many of the students (38%), particularly the higher-scoring ones, went for the obvious distractor, option *A*. Almost 12% of the examinees that chose *A* in item 49, chose *A* in item 57. Interestingly enough, 24% of the students that chose *A* in item 49 chose *C* (the correct response) in 54, but only 4% who chose *C* in 49 (the correct response) chose *A* in 54. Although item 57 detracts from the statistics of the exam as a whole, the insight that can be gained about students' understanding of the concepts earned it a place on the final version of the exam. One reviewer suggested that item 54 (and presumably item 57) could be misinterpreted to mean the magnetic field at the cross due to each of the wires in item 53 (and 56) (Crowder, 1997). Hence, the investigator change the stem for item 54 to read, "What is the relationship between the net magnetic field at the cross in question 52 and the cross in question 53?" The stem for item 57 was changed accordingly.

88) Items 58 & 59

Items 58 and 59 were good test questions. They both were adequate discriminators and correlated well with total scores. Item 58 did have one shortcoming: it was too easy in the post-test, and, consequently, most of the distractors were not very strong. The investigator used the DIFT to analyze the results more closely. It revealed the pattern one would expect. On the post-test, 60% of the examinee chose option *A* in item 58 (correct) and option *A* (incorrect) in item 59. To improve item 58, the distractors were changed back to the distractors from Version 1.0. Although these were weak as well, they were better than the ones on the latest version, and they present a more coherent pattern.

89) Items 60 & 61

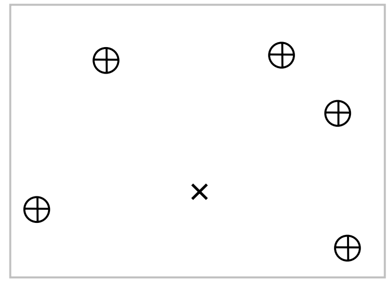
The investigator changed these two items. Item 60 proved to be misleading, even for the experts. The electric field does point in an upward direction. Therefore test takers of all abilities picked option *C*. Item 61 was not a very good item either. The investigator changed item 61 to a similar question from Version 1.0 that had good statistics – item difficulty of 0.56 and item discrimination of 0.53. The investigator changed item 60 so the problems mentioned previously were removed, and it still coupled with the new item 61.

The investigator changed items 60 and 61 to read:

For questions 60-61, assume the distances between the cross and the charged particles are all the same, however, the exact positions of the charged particles are *not* known.

60) For the figure to the right, what can be said about the direction net electric field at the cross?

a) The direction can *not* be determined *exactly*, but there is a net field at the cross.
b) The direction can be determined *exactly* and the magnitude of the field at the cross is *not* zero.
c) The field has no direction, because the magnitude of the field at the cross is zero.



61) For the figure to the right, what is the value of the electrostatic potential at the cross?

a) Less than 0. b) 0. c) Greater than 0.
d) There is not enough information.

90) Items 62-65

The investigator intended for these four questions to probe students' understanding of induced current. Originally, options about charge separation were included, however it was pointed out that in many of the items in Versions 1.0, 2.0, and 3.0, an argument could be made for charge separation (Sherwood, 1996; and Palmer, 1996). To remove any possible alternate interpretations of students' answers, the investigator removed option *A* from all four items. Also, to increase the variance, the investigator changed option *E* (now *D* with option *A* removed) to "No current is induced." Another reviewer suggested that the use of the phrase "The magnetic field is uniform" in the stems for items 62-65 is misleading (Lindberg, 1997). The reviewer felt that the field was non-uniform, because the loop moved from a region of magnetic field to a region of no magnetic field. The investigator changed those stems to read, "The magnetic field in the gray region is **uniform,"** and the stem for item 63 now reads, "The magnetic field in the gray region is **non-uniform**." The same reviewer thought the stem for 64 was misleading because it read, "Choose an answer that describes something that happens to the wire loop as it moves as shown in the figure to the right. The loop in the diagram, however, is not moving. The investigator did not find this contradictory, but, to prevent confusion or misinterpretation by the students the stem for item 64 now reads, “Choose an answer that describes something that happens to the stationary wire loop in the figure to the right.” Finally, in every option the investigator replaced the word “induced” with “produced.” All the items but 63 had strong statistics, and all of the distractors were appealing. The statistics of item 63 were low, most likely because non-uniform fields are not covered in-detail with the pool of students that took the exam.

The investigator darkened the gray regions to improve photocopying.

91) Items 66-68

On Version 3.0 of the DEEM, these questions produced weak statistics. Their difficulties ranged from 0.32 down to 0.20, and their discrimination indices ranged from 0.19 to 0.09. Items 66 and 67 were on Version 1.0 (as items 75 and 76) where they posted stronger statistics. The items’ difficulties were about the same, but the discrimination indices were 0.49 and 0.22, respectively. On both versions, the students found all of the options plausible. The investigator kept these items on the exam.

The investigator made the gray regions darker to improve photocopying.

92) Items 69-70

The investigator changed the third sentence in the stems of both of these items to, “Looking at the loop as it enters the non-uniform magnetic field directed into the page…” The investigator did this after it was suggested by a reviewer that students may get confused about the direction the magnetic field is pointed as they shift between the “TOP VIEW” and the “SIDE VIEW” (Lindberg, 1997). Item 70 was an extremely poor discriminator (0.01). This does not reflect the fact that, although a nearly equal number of high and low scorers answered the question correctly (about 35%), high scorers on the exam were much more likely to choose answer *B* than low scorers (51% versus 31%). These two questions also reveal an intriguing pattern of students' responses. Using the DIFT, the investigator realized that about 19% of the students were likely to change their answer from *A* to *B* (or vice versa) and nearly as many, about 16%, were likely to stay with either *A* or *B* in the two questions. So, about 70% of the students chose some combination of *A* or *B* in these two items, but only about 30% realized that the situations were the same, and only half of them got both answers correct.

93) Final Profile of the DEEM

It is important for a test creator to ensure that the final version of the exam still strongly samples the domain it was originally intended to test. Table 54–I reveals the pattern of concepts on the final version of the DEEM.

Table 54–I: Distribution of Concepts on the Final Version of the DEEM

Items	Concept
1-4	Force due to electric field
5-9	Force due to magnetic field
10-15	Electric field properties
16-23	Electrostatic Potential
24-27	Potential Energy
28-30	Magnetic Field properties
31-33	Maxwell's equations
34-35	Induced currents
36-37	Force due to magnetic field
38-39	Force due to electric field
40-45	Electric field properties
46-47	Potential Energy
48-53	Magnetic Field properties
54 & 56	Electric field properties
55 & 57	Electrostatic Potential
58-61	Induced currents
62-64	Maxwell's equations
65-66	Induced currents

Table 54–I indicates the concepts were evenly distributed through out the exam. One can outline the concepts in a slightly different way to show how many items were devoted to each concept. (See Table 54–II below.)

Table 54–II: Frequency of Concepts on the Final Version of the DEEM

<table>
<tr><th colspan="2">CONCEPT</th><th colspan="2">NUMBER OF ITEMS</th></tr>
<tr><td rowspan="2">Force on a charged particle in a (an)</td><td>Electric Field</td><td>6</td><td rowspan="2">13</td></tr>
<tr><td>Magnetic Field</td><td>7</td></tr>
<tr><td rowspan="2">Field properties</td><td>Electric</td><td>9</td><td rowspan="2">19</td></tr>
<tr><td>Magnetic</td><td>10</td></tr>
<tr><td colspan="2">Electrostatic Potential</td><td colspan="2">10</td></tr>
<tr><td colspan="2">Electrostatic Potential Energy</td><td colspan="2">6</td></tr>
<tr><td colspan="2">Maxwell’s Equations</td><td colspan="2">6</td></tr>
<tr><td colspan="2">Induced Currents</td><td colspan="2">8</td></tr>
</table>

Table 54–II proves there was an even presentation of concepts; no one topic monopolized the exam.

Finally, it is important to ensure that the keys (correct options) are reasonably distributed on the exam. In other words, there should be about the same number of *A*’s *B*’s, *C*’s, etc. that serve as the correct response.

Table 54–III: Frequency of Options Serving as Keys

Option	Frequency as a key
A	14
B	15
C	17
D	10
E	10

Table 54–III displays that options *A*, *B*, and *C* enjoyed being the keys nearly an equal number of times. *D* and *E* suffered slightly, but since some of the items do not have and option *E*, or options *D* and *E*, the distribution seems even.

With the final version of the exam complete, the test creator is prepared to draw conclusions regarding the exam and its strengths, as well as prepare it for publication.

"To write it, it took months; to conceive it – three minutes; to collect the data in it – all my life." – F. Scott Fitzgerald

CONCLUSION

94) Conclusion

This investigator has created a research-based diagnostic exam to test undergraduates' knowledge of basic concepts in electricity and magnetism (DEEM). The DEEM is a sixty-eight item multiple-choice exam to be administered in fifty minutes. The keys represent experts universally accepted answers concerning electromagnetic phenomena; the distractors are based on students' most common misconceptions of those same phenomena as gleaned from open-ended, interview questions with students that represented the population of examinees.

The DEEM is to be administered in a pre-test/post-test fashion. The pre-test is to be given to the students before any of the material on the exam is covered in the class and the post-test is to be given some time after all the material on the exam has been covered in the class. The investigator designed the exam so physics instructors may easily interpret patterns of students' responses to draw conclusions about their mental constructions of the principles of electricity and magnetism. The results of the pre-test can be used to establish a baseline of students' preinstructional knowledge and notions of basic concepts in electricity and magnetism. The results of the post-test can be used to assess individual student achievement and determine misconceptions that persists through instruction or ones that developed as a result of instruction. Furthermore, based in part on uses of other diagnostic exams employed by the physics education community in mechanics, once the DEEM has become widely accepted and used by the research community, it can be given to groups of students to gauge learning gains in various instructional environments.

The conceptual domain of the exam is composed of the following physical concepts: forces on charged particles in static, uniform electric and magnetic fields; properties of the electric and magnetic vector fields; properties of the scalar, electrostatic potential field, and electrostatic potential energy; generalized applications of Maxwell's equations; and induced currents with Lenz's law.

To verify the exam's reliability, the investigator used well-established statistical techniques and test creation strategies to create an exam that generates reproducible results. The investigator's calculation of coefficient alpha (α=0.74 for the post-test, Version 3.0) provides proof that the final version's reliability is sound. This statistic is certainly a lower bound for the reliability of the final version based on two facts. An earlier version of the exam produced a much higher result (α=0.88 and 0.93 for the pre-test and post-test respectively, Version 1.0), and the latest version of the exam has unquestionably improved stems and options, wording, optimized numbers and choices of options, diagrams, and design.

To establish the exam's validity, the investigator surveyed widely used introductory physics textbooks and experienced instructors to establish the content domain of the exam. Furthermore, the investigator disseminated versions of the exam to the physics education community for comments and criticism. The investigator included relevant critiques of items into the most recent version of the exam.

Finally, the investigator produced an exam that, given the pool of examinees, does not show any significant statistical biases based on gender.

95) Recommendations

The investigator has three recommendations to further improve and expand the uses of the DEEM

1. The reliability and validity of the exam for use on high school students should be established through field-testing and expert review, thus broadening the pool of potential examinees and creating a tool for a community of instructors that, as of now, is without one.

2. The exam's statistics should be examined in light of the individual concepts on the exam. For example, one could group all of the students' responses dealing with electrostatic forces together. In general, the reliability of each subsection would be lower than the reliability of the whole test. However, given the fragmented knowledge set most students have of physical concepts, it is conceivable that the reliability of smaller sets of questions may actually be higher than the entire exam. If this were the case, then instructors could safely partition the exam into smaller bits for research and assessment purposes.

3. The pool of examinees (Physics II students at Rensselaer Polytechnic Institute) may represent a homogenous population of students, with regards to electricity and magnetism. The reliability of the exam, i.e. the value of coefficient alpha, may very well be higher than what was measured at Rensselaer. As more data is collected from different colleges and universities, it would be prudent to recalculate coefficient alpha for a more heterogeneous group.

96) REFERENCES

Adrian, B. and Fuller, G. (1997) Electricity and Magnetism Fields According to College Students. Contributed Paper, 1997 Summer AAPT Meeting.

Aiken, L. Jr. (1976). Psychological Testing and Assessment. Boston, Massachusetts: Allyn and Bacon, Inc.

Alexopoulou, E. and Driver, R. (1997) Gender Differences in Small Group Discussion in Physics. International Journal of Science Education, 19, 4, 393-406.

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. (1985). Standards for Educational and Psychological Tests and Manuals. Washington D.C: American Psychological Association Press.

Angoff, W. (1982). Use of Difficulty and Discrimination Indices for Detecting Item Bias. In R. A. Berk, (Ed.), Handbook of Methods for Detecting Item Bias. Baltimore, Maryland: The Johns Hopkins University Press.

Angoff, W. H, and Ford, S. F. (1973). Item Race Interaction on a Test of Scholastic Aptitude. Journal of Educational Measurement, 18, 97-108.

Aubrecht II, G. and Aubrecht, J. (1983). Constructing Objective Tests. American Journal of Physics, 51, 7, 613-620.

Barnett-Foster, D. and Nagey, P. (1995). A Comparison of Undergraduate Test Response Strategies of Multiple-Choice and Constructed-Response Questions. The Alberta Journal of Educational Research, 41, 1, 18-35.

Baron, R. A. (1998). Psychology. Boston, Massachusetts: Allyn and Bacon, Inc.

Behringer, E. (1997). Personal communication via electronic mail. August 27, 1997.

Beichner, R. (1994). Testing Student Interpretation of Kinematics Graphs. American Journal of Physics, 62, 8, 750-762.

Berk, R. A. (1982). Handbook of Methods for Detecting Item Bias. United States of America: The Johns Hopkins University Press.

Bork, A. (1984). Letter to the Editor. American Journal of Physics, 52, 10, 873-874.

Camilli, G. (1979). A Critique of the Chi-Square Method for Assessing Item Bias. Unpublished paper, Laboratory of Educational Research, University of Colorado, Boulder. Referenced in Ironson (1982). Use of Chi-square and Latent Trait Approaches for Detecting Item Bias. Handbook for Dectecting Item Bias. In R. A. Berk (Ed.). Baltimore, Maryland: Johns Hopkins University Press.

Chambers, S. and Andre, T. (1997) Gender, Prior Knowledge, and Experience in Electricity and Conceptual Change Text Manipulations in Learning about Direct Current. Journal of Research in Science Teaching, 34, (2), 107-123.

Cohen, R., Eylon, B., and Ganiel, U. (1983). Potential Difference and Current in Simple Electric Circuits: A Study of Students' Concepts. American Journal of Physics, 51, 4, 407-412.

Cohen, R., Eylon, B., and Ganiel, U. (1984). Answer to Letter by M. Iona. American Journal of Physics, 52, 4, 392.

Crocker, L. and Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York, New York: CBS College Publishing.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. Psychometrika, 16, 297-334.

Crowder, J. (1997). Personal communication via electronic mail. August 20, 1997.

Downey, R. G. (1979). Item-Option Weighting of Achievement Tests: Comparative Study of Methods. Applied Psychological Measurement, 3, 453-462.

Eylon, B. and Ganiel, U. (1989). Macro-Micro Relationships: The Missing Link Between Electrostatics and Electrodynamics in Students' Reasoning. Preprint, International Journal of Science Education.

Fishbane, P. M., Gasiorowicz, S., and Thornton, S. (1996). Physics for Scientists and Engineers: Volume II. Upper Saddle River, New Jersey: Prentice Hall.

Fox, K. (1997). Personal communication via electronic mail. February 21, 1997.

Fredette, N. and Lochhead, J. (1980). Student Conceptions of Simple Circuits. The Physics Teacher, 18, March, 194-198.

Furst, E. (1964). Constructing Evaluation Instruments. United States of America: David McKay, Inc.

Galili, I. (1995). Mechanics Background Influences Students' Conceptions in Electromagnetism. International Journal of Science Education, 17, 3, 371-387.

Gilman, D. A., and Ferry, P. (1972). Increasing Test Reliability Through Self-Scoring Procedures. Journal of Educational Measurement, 9, 205-207.

Glass, G. and Hopkins, K. (1970). Statistical Methods in Education and Psychology. Englewoods Cliffs, New Jersey: Prentice-Hall, Inc.

Gregory, R. (1992). Psychological Testing: History, Principles and Applications. Needham Heights, Massachusetts: Allyn and Bacon, Inc.

Gulliksen, H. (1945). The Relation of Item Difficulty and Inter-item Correlation to Test Variance and Reliability. Psychometrika, 10, 79-91.

Guth, J. (1995). An In-depth Study of Two Individual Students' Understanding of Electric and Magnetic Fields. Research in Science Education, 25, (4), 479-490.

Gutwill, J., Frederiksen, J., and Ranney, M. (1996). Seeking the Causal Connection in Electricity: Shifting Among Mechanistic Perspectives. International Journal of Science Education, 18, 2, 143-162.

Guzzetti, B. and Williams, W. (1996) Gender, Text, and Discussion: Examining Intellectual Safety in the Science Classroom. Journal of Research in Science Teaching, 33, 1, 5-20.

Hake, R. (1998). Interactive-Engagement vs. Traditional Methods: A Six-Thousand Student Survey of Mechanics Test Data for Introductory Physics Courses. American Journal of Physics, 66, 1, 64-74.

Halliday, D., Resnick, R., and Walker, J. (1997). Fundamentals of Physics: Volume II. New York, New York: John Wiley & Sons, Inc.

Halloun, I. and Hestenes D. (1985b). Common Sense Concepts About Motion. American Journal of Physics, 53, 11, 1056-1065.

Halloun, I. and Hestenes, D. (1985a). The Initial Knowledge State of College Physics Students. American Journal of Physics, 53, 11, 1043-1055.

Hanna, G. S. (1975). Incremental Reliability and Validity of Multiple-Choice Tests With an Answer-Until-Correct Procedure. Journal of Educational Measurement, 12, 175-178.

Harding, J. (1979) Sex Difference in Examination Performance at 16+. Physics Education, 14, 280-284.

Harrington, R. (1996). Personal communication via electronic mail. August 23, 1996.

Hazel, E., Logan, P., and Gallagher, P. (1997) Equitable Assessment of Students in Physics: Importance of Gender and Language Background. International Journal of Science Education, 19, (4), 382-392.

Heller, P. and Huffman, D. (1995). Interpreting the Force Concept Inventory: A Reply to Hestenes and Halloun. The Physics Teacher, 33, November, 503 & 507-511.

Henderickson, G. F. (1971). The Effect of Differential Option Weighting on Multiple-Choice Objective Tests. Journal of Educational Measurement, 8, 291-296.

Hestenes, D. and Wells, M. (1992) A Mechanics Baseline Test. The Physics Teacher, 30, March, 159-166.

Hestenes, D., and Halloun, I. (1995). Interpreting the Force Concept Inventory: A Reply to Hestenes and Halloun. The Physics Teacher, 33, November, 502 & 504-506.

Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force Concept Inventory. The Physics Teacher, 30, March, 141-158.

Hobson, A. and Ghoshal, D. (1996). Flexible Scoring for Multiple-Choice Exams. The Physics Teacher, 34, May, 284

Horn, D. (1970). The Writing of Multiple-Choice Mathematics Test Items. Toronto, Ontario: The Ontario Institute for Studies in Education.

Hudson, H. and Hudson, C. (1981). Suggestions on the Construction of Multiple-Choice Tests. American Journal of Physics, 49, 9, 838-841.

Huffman, D., and Heller, P. (1995). What Does the Force Concept Inventory Really Measure? The Physics Teacher, 33, March, 138-143.

Iona, M. (1984). Multiple-Choice Questions. American Journal of Physics, 52, 4, 392.

Iona, M. (1986). In Self-Defense and Defense of Multiple-Choice Tests. American Journal of Physics, 54, 4, 295-296.

Ironson (1982). Use of Chi-square and Latent Trait Approaches for Detecting Item Bias. Handbook for Dectecting Item Bias. In R. A. Berk (Ed.). Baltimore, Maryland: Johns Hopkins University Press.

Ironson, G. (1982). Use of Chi-Square and Latent Trait Approaches for Detecting Item Bias. In R. A. Berk, (Ed.), Handbook of Methods for Detecting Item Bias. (pp. 117-160). United States of America: The Johns Hopkins University Press.

Kerlinger, F. (1986). Foundations of Behavioral Research. New York, New York: CBS College Publishing.

Lilley, C. (1989) Hairbrushes and Diamond Rings: Femininity in Physics. The Physics Teacher, 27, 9, 479-480.

Lindberg, V. (1997). Personal communication via electronic mail. September 10, 1997

Lord, F. M. (1977). Optimal Number of Choices per Item: A Comparison of Four Approaches. Journal of Educational Measurement, 14, 33-38.

Lorre, D. (1997). Personal communication via electronic mail. August 20, 1997.

Maloney, D., O'Kuma, T., Van Heuvelen, A., and Hieggelke, C. (1997) Developing Electricity and Magnetism Assessments. Contributed Paper, 1997 Summer AAPT Meeting

McDermott, L. and Shaffer, P. (1992). Research as a Guide for Curriculum Development: An Example from Introductory Electricity. Part I: Investigation of Student Understanding. American Journal of Physics, 60, 11, 994-1003.

McGrath, R. E. (1997). Understanding Statistics: A Research Perspective. United States of America: Addison-Wesley Educational Publishers Inc.

Mehrens, W. and Lehmann, I. (1984). Measurement and Evaluation in Education and Psychology. United States of America: Holt, Rinehart, and Winston, Inc.

Meltzer, A (1998). Personal communication. February, 13 1998.

Messick, S. (1980). Test Validity and the Ethics of Assessment. American Psychologist, 35, 1012-1027.

Minium, E. (1978). Statistical Reasoning in Psychology and Education. United States of America: John Wiley and Sons, Inc.

Nevo, B. (1985). Face Validity Revisited. Journal of Educational Measurement, 22, 287-293.

Odom, A. and Barrow, L. (1995). Development and Application of a Two-Tier Diagnostic Test Measuring Colleg Biology Students' Understanding of Diffusion and Osmosis after a Course of Instruction. Journal of Research in Science Teaching, 32, 1, 45-61.

Oosterhof, A. (1996). Developing and Using Classroom Assessment. Englewoods Cliffs, New Jersey: Prentice-Hall, Inc.

Palmer, L. (1996). Personal communication via electronic mail. August 29, 1996.

Patnaik, D., and Traub, R. E. (1973). Differential Weighting by Judged Degree of Correctness. Journal of Educational Measurement, 10, 281-286.

Pfeiffenberger, G. and Modu, C. (1977). A Validity Study of the Multiple-Choice Component of the Advanced Placement Physics C Examination. American Journal of Physics, 45, 11, 1066-1069.

Pinkerton, D. (1997). Dimensionality of the Force Concept Inventory: A Comparison of Factor and Rasch Analysis Techniques. Paper distributed accompanying a contributed talk at the 1997 Summer AAPT Meeting.

Popham, W. J. (1981). Modern Educational Measurement. Englewood Cliffs, New Jersey: Prentice-Hall

Prosser, M. (1994). A Phenomenographic Study of Students' Intuitive and Conceptual Understanding of Certain Electrical Phenomena. Instructional Science, 22, 189-205

Rachinger, W. (1973). Electricity and Magnetism: Diagnostic Tests. Sydney, Australia: John Wiley and Sons Australasia Pty. Ltd.

Rainson, S., Viennot, L., and Tranströmer, G. (1994). Students' Understanding of Superposition of Electric Fields. American Journal of Physics, 62, 11, 1026-1032.

Redish, E., Saul J., and Steinberg R. (1997). On the Effectiveness of Active-Engagement Microcomputer-Based Laboratories. American Journal of Physics, 65, 1, 45-54.

Reif, F. and Allen, S. (1992) Cognition for Interpreting Scientific Concepts: A Study of Acceleration. Cognition and Instruction, 9, 1, 1-44.

Salvia, J. and Hughes, C. (1990). Curriculum-Based Assessment: Testing What is Taught. New York, New York: MacMillan Publishing Company.

Sandin, T. (1985). On Not Choosing Multiple-Choice. American Journal of Physics, 53, 4, 299-300.

Scheuneman, J. (1979). A New Method for Assessing Bias in Test Items. Journal of Educational Measurement, 16, 143-152.

Scott, B. (1985). A Defense of Multiple-Choice Tests. American Journal of Physics, 53, 11, 1035.

Serway, R. A. (1994). Principles of Physics. United States of America: Saunders College Publishing.

Shaffer, P. and McDermott, L. (1992). Research as a Guide for Curriculum Development: An Example from Introductory Electricity. Part II: Design of Instructional Strategies. American Journal of Physics, 60, 11, 1003-1009.

Sherwood, B. (1996). Personal communication via electronic mail. September 7, 1996.

Sobolewski, S. (1996). Development of Multiple-Choice Test Items. The Physics Teacher, 34, February, 80-82.

Sokoloff D. and Thornton, R. (1996). Using Interactive Lecture Demonstrations to Create and Active Learning Environment. Preprint, submitted for publication in the Proceedings of ICUPE, September, 1996.

Stienberg, R. and Sabella, M. (1997). Performance on Multiple-Choice Diagnostics and Complementary Exam Problems. The Physics Teacher, 35, (3), 150-155.

Strosnider, K. (1997). A Controversial Study of Testing Finds the Gap Between Boys and Girls Is Shrinking. Chronicle of Higher Education, 43, 36, 34

Styer, D. (1996). Personal communication via electronic mail. August 23, 1996.

Thompson, N. (1979). Sex Differentials in Physics Education. Physics Education, 14, 285-288 & 317-320.

Thorton, R. and D. Sokoloff. (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, 4, 338-532.

Törnkvist, S., Pettersson, K.-A., and Tranströmer, G. (1993). Confusion by Representation: On Students' Comprehension of the Electric Field Concept. American Journal of Physics, 61, 4, 335-338

Traub, R. E., and Hambleton, R. K. (1972). The Effect of Scoring Instrcutions and Degree of Speededeness on the Validity and Reliablity of Multiple-Choice Tests. Educational and Psychological Measurement, 32, 737-758.

Varlashkin, P. (1987). Modified Multiple-Choice Testing. The Physics Teacher, 25, May, 37.

Varney, R. (1984). More Remarks on Multiple-Choice Questions. American Journal of Physics, 52, 12, 1069.

Viennot, L. and Rainson, S. (1992). Students' Reasoning About the Superposition of Electric Fields. International Journal of Science Education, 14, 4, 475-487.

Walford, G. (1981). Tracking Down Sexism in Physics Textbooks. Physics Education, 16, 261-265.

Wilson, J. M. (1994). The CUPLE Physics Studio. The Physics Teacher, 32, 12, 518-523.

APPENDIX A

Table A–I: Frequency Table

Frequency Table

Answer	Item	A		B		C		D		E		Omit	
		pre	post	pre	post	pre	post	pre	post	pre	post	pre	post
D	1	4.3	3.5	5.2	4.0	28.0	14.7	**46.1**	**62.0**	15.3	15.9	1.2	0.0
C	2	4.0	4.6	4.6	3.5	**54.8**	**64.6**	22.8	13.0	13.0	14.4	0.9	0.0
B	3	33.4	18.7	**39.8**	**62.8**	6.1	4.9	2.3	2.9	17.6	10.7	0.9	0.0
A	4	**51.6**	**69.2**	21.6	16.1	6.6	4.9	4.0	2.6	15.0	7.2	1.2	0.0
E	5	18.7	3.5	**24.2**	6.3	20.7	15.9	18.4	16.1	16.1	**57.6**	1.7	0.6
E	6	**28.2**	6.6	15.9	2.3	18.4	17.3	20.7	16.1	14.7	**57.1**	2.0	0.6
E	7	**25.1**	7.5	21.9	8.9	16.1	17.0	15.9	20.2	18.7	**44.7**	2.3	1.7
C	8	18.4	4.9	**27.1**	14.1	20.7	**41.8**	19.9	28.5	11.2	8.9	2.6	1.7
D	9	**31.7**	13.5	14.7	6.1	19.9	27.7	20.5	**42.4**	10.7	8.6	2.6	1.7
A	10	**36.0**	**80.7**	32.6	11.5	2.9	2.0	2.0	0.6	25.6	5.2	0.9	0.0
B	11	5.8	4.3	**56.2**	**78.1**	7.2	5.2	2.3	0.6	4.3	0.3	24.2	11.5
A	12	24.8	42.7	21.9	8.6	3.5	0.6	2.6	0.6	**45.5**	**46.7**	1.7	0.9
B	13	32.0	12.7	**40.9**	**80.4**	1.7	0.6	2.6	1.2	21.3	4.9	1.4	0.3
B	14	6.9	4.9	**59.1**	**77.2**	7.5	5.2	2.9	0.9	4.0	0.9	19.6	11.0
B	15	21.0	8.1	27.7	**45.5**	1.4	0.3	2.9	0.6	**43.5**	45.2	3.5	0.3
D	16	20.5	11.5	10.1	7.5	15.9	10.4	**38.0**	**56.8**	14.1	12.4	1.4	1.4
B	17	13.3	11.5	**35.4**	**53.6**	14.4	12.4	4.3	1.7	2.0	0.6	30.5	20.2
C	18	20.2	23.6	14.1	9.2	27.1	**39.2**	4.0	2.0	3.2	4.9	**31.4**	21.0
D	19	17.6	30.0	**28.0**	11.8	15.3	10.1	22.2	**38.3**	14.4	7.5	2.6	2.3
E	20	17.0	10.4	11.5	7.5	12.7	10.1	24.8	29.4	**31.1**	**41.5**	2.9	1.2
A	21	17.0	23.1	**37.2**	**43.5**	11.8	11.5	2.9	1.4	0.9	0.6	30.3	19.9
B	22	17.3	23.6	25.6	**30.3**	19.9	19.3	4.0	4.0	3.7	3.7	**29.4**	19.0
D	23	22.5	11.5	**26.5**	34.0	13.8	10.1	21.0	**34.9**	13.3	8.1	2.9	1.4
E	24	3.5	4.6	19.9	22.2	24.8	23.9	18.2	19.6	**32.3**	**28.8**	1.4	0.9
B	25	12.1	8.4	**45.8**	**42.9**	16.1	18.7	3.2	2.0	2.0	1.4	20.7	26.5
C	26	4.6	4.0	13.8	17.3	9.2	8.6	25.1	25.4	**44.7**	**43.5**	2.6	1.2
A	27	13.3	12.1	**44.1**	**37.2**	23.1	28.0	2.6	0.9	2.0	0.3	15.0	21.6
C	28	15.0	0.3	15.0	4.0	**38.9**	**85.9**	21.9	8.9	8.1	0.6	1.2	0.3
B	29	2.9	1.2	**66.6**	**84.4**	15.6	11.0	2.9	0.6	2.6	0.0	9.5	2.9
C	30	8.4	0.3	13.0	3.5	**30.5**	**54.5**	16.7	7.8	30.0	33.7	1.4	0.3
E	31	0.9	0.0	5.8	0.6	1.4	0.3	2.0	0.0	**88.8**	**99.1**	1.2	0.0
Omit	32	1.7	0.0	5.2	1.2	6.9	4.9	0.9	0.9	5.8	3.5	**79.5**	**89.6**
E	33	0.9	0.0	1.4	0.0	6.3	0.6	2.3	0.0	**83.9**	**97.1**	5.2	2.3
C	34	15.0	4.6	8.1	4.9	**59.4**	**79.0**	6.6	2.0	9.2	8.9	1.7	0.6

Frequency Table (cont.)

Answer	Item	A		B		C		D		E		Omit	
		pre	post	pre	post	pre	post	pre	post	pre	post	pre	post
A	35	**41.5**	**56.2**	25.1	22.2	9.5	2.6	11.2	13.8	11.2	4.6	1.4	0.6
E	36	19.9	39.2	13.0	6.1	**45.5**	**45.0**	6.3	1.4	13.0	7.5	2.3	0.9
D	37	19.6	29.1	19.0	14.7	**25.1**	16.7	15.0	**30.8**	17.9	7.8	3.5	0.9
C	38	19.3	17.6	13.8	8.9	**33.4**	**54.8**	23.1	16.4	8.4	1.7	2.0	0.6
E	39	8.9	6.1	14.4	8.4	25.6	29.4	**30.0**	**34.6**	15.9	18.2	5.2	3.5
D	40	**37.8**	10.1	2.3	0.3	8.9	1.7	34.6	**76.4**	14.1	11.0	2.3	0.6
E	41	19.0	17.6	3.7	0.0	7.5	0.9	15.3	14.7	**52.2**	**66.9**	2.3	0.0
E	42	**46.4**	**42.7**	4.0	1.4	7.8	2.6	20.2	26.5	17.9	26.2	3.7	0.6
E	43	12.7	10.7	4.6	0.3	8.6	2.3	14.7	11.2	**55.6**	**75.2**	3.7	0.3
A	44	21.3	**76.9**	13.8	7.8	3.5	0.9	2.0	0.6	**57.3**	13.5	2.0	0.3
A	45	**44.7**	**86.7**	40.6	11.0	4.3	0.9	3.2	0.6	4.9	0.6	2.3	0.3
C	46	32.0	**45.0**	9.5	9.2	**38.3**	37.5	9.8	4.6	8.1	2.6	2.3	1.2
E	47	6.9	0.9	3.7	1.4	10.1	2.0	3.7	1.4	**72.6**	**93.9**	2.9	0.3
B	48	**30.5**	10.4	29.7	**67.7**	6.9	4.3	5.2	1.7	24.2	15.6	3.5	0.3
C	49	34.0	22.8	7.8	1.4	**35.4**	**65.7**	6.1	3.2	13.3	6.3	3.5	0.6
D	50	3.7	15.3	5.5	8.1	20.2	**40.1**	**62.0**	30.3	4.6	5.2	4.0	1.2
C	51	3.7	1.2	1.7	0.6	33.7	22.2	**50.7**	**72.9**	6.6	2.3	3.5	0.9
E	52	6.1	6.9	12.1	7.2	20.5	4.3	13.8	2.6	**42.7**	**78.4**	4.9	0.6
B	53	24.8	11.5	**28.5**	**74.1**	22.5	5.8	11.0	3.2	7.5	4.6	5.8	0.9
C	54	25.9	19.0	11.2	6.3	**38.6**	**67.1**	10.1	3.5	8.6	3.2	5.5	0.9
B	55	8.4	3.2	29.4	**68.9**	8.1	1.7	6.9	1.7	**42.1**	23.6	5.2	0.9
B	56	20.7	9.8	**36.0**	**76.9**	22.2	7.2	7.5	1.4	8.4	3.7	5.2	0.9
C	57	22.5	**38.0**	19.0	30.8	**32.9**	22.8	10.7	5.2	9.2	2.0	5.8	1.2
A	58	**61.7**	**89.6**	17.6	3.5	3.5	1.7	2.6	0.6	9.5	3.7	5.2	0.9
C	59	**52.2**	**63.4**	8.1	4.0	25.1	24.8	9.2	6.1	0.3	0.3	5.2	1.4
E	60	19.6	11.8	13.0	6.3	16.1	35.2	11.0	6.1	**34.3**	**38.6**	6.1	2.0
A	61	27.7	24.8	8.6	6.3	**32.9**	**45.8**	21.6	19.6	3.7	2.0	5.5	1.4
D	62	22.2	5.2	11.0	3.5	**25.6**	**43.2**	25.1	41.5	8.4	4.0	7.8	2.6
D	63	13.5	6.6	**30.8**	**36.6**	21.3	24.8	16.7	22.2	9.5	7.5	8.1	2.3
E	64	21.6	16.1	12.7	5.5	13.8	7.5	13.5	6.6	**29.4**	**61.1**	8.9	3.2
E	65	14.7	3.5	7.5	3.5	21.6	19.6	23.3	17.9	**23.6**	**52.4**	9.2	3.2
D	66	15.3	7.8	**27.7**	12.4	17.3	23.9	12.1	**31.7**	18.4	20.5	9.2	3.7
C	67	13.5	4.9	**20.7**	15.0	19.3	**28.2**	17.0	25.4	20.2	22.8	9.2	3.7
C	68	12.4	8.6	21.3	11.5	16.1	19.6	11.8	17.9	**29.1**	**38.0**	9.2	4.3
A	69	20.2	36.9	**35.2**	**40.6**	11.8	7.5	11.0	6.9	11.2	4.0	10.7	4.0
A	70	**35.2**	37.5	23.9	**38.9**	10.4	5.5	9.5	6.3	10.4	7.8	10.7	4.0

Table A–II: Upper and Lower Frequency Table

Key	Item	Frequency Table (post)											
		A		B		C		D		E		Omit	
		upper	lower	upper	lower	upper	lower	upper	lower	upper	lower	upper	lower
D	1	1.06	5.32	1.06	7.45	6.38	31.91	86.17	34.04	5.32	21.28	0.00	0.00
C	2	2.13	7.45	0.00	6.38	87.23	38.30	6.38	24.47	4.26	23.40	0.00	0.00
B	3	5.32	32.98	84.04	34.04	0.00	11.70	0.00	6.38	10.64	14.89	0.00	0.00
A	4	91.49	43.62	4.26	27.66	1.06	11.70	0.00	5.32	3.19	11.70	0.00	0.00
E	5	2.13	5.32	2.13	8.51	4.26	26.60	2.13	29.79	89.36	29.79	0.00	0.00
E	6	3.19	6.38	1.06	4.26	4.26	32.98	2.13	27.66	89.36	28.72	0.00	0.00
E	7	2.13	8.51	3.19	14.89	10.64	25.53	7.45	31.91	76.60	17.02	0.00	2.13
C	8	0.00	9.57	4.26	19.15	67.02	23.40	22.34	36.17	6.38	8.51	0.00	3.19
D	9	3.19	17.02	1.06	12.77	21.28	36.17	68.09	24.47	6.38	6.38	0.00	3.19
A	10	95.74	57.45	3.19	28.72	0.00	4.26	0.00	0.00	1.06	9.57	0.00	0.00
B	11	0.00	6.38	92.55	62.77	2.13	11.70	0.00	1.06	0.00	0.00	5.32	18.09
A	12	72.34	18.09	5.32	15.96	0.00	1.06	0.00	1.06	21.28	62.77	1.06	1.06
B	13	4.26	29.79	94.68	58.51	0.00	0.00	0.00	1.06	1.06	9.57	0.00	1.06
B	14	0.00	6.38	92.55	61.70	1.06	11.70	1.06	0.00	0.00	3.19	5.32	17.02
B	15	5.32	15.96	75.53	20.21	0.00	1.06	0.00	0.00	19.15	61.70	0.00	1.06
D	16	3.19	23.40	4.26	15.96	4.26	17.02	76.60	28.72	10.64	12.77	1.06	2.13
B	17	7.45	14.89	78.72	24.47	5.32	15.96	1.06	1.06	0.00	1.06	7.45	42.55
C	18	11.70	32.98	9.57	6.38	63.83	10.64	2.13	2.13	4.26	4.26	8.51	43.62
D	19	22.34	27.66	7.45	17.02	3.19	15.96	62.77	21.28	3.19	13.83	1.06	4.26
E	20	2.13	22.34	5.32	15.96	5.32	14.89	23.40	27.66	63.83	17.02	0.00	2.13
A	21	35.11	8.51	48.94	32.98	7.45	12.77	1.06	1.06	0.00	0.00	7.45	44.68
B	22	13.83	31.91	56.38	8.51	13.83	13.83	5.32	2.13	2.13	3.19	8.51	40.43
D	23	7.45	17.02	22.34	31.91	5.32	15.96	59.57	17.02	3.19	15.96	2.13	2.13
E	24	3.19	7.45	17.02	12.77	22.34	26.60	12.77	25.53	43.62	25.53	1.06	2.13
B	25	7.45	8.51	55.32	40.43	14.89	21.28	2.13	4.26	0.00	2.13	20.21	23.40
C	26	2.13	7.45	17.02	7.45	18.09	1.06	17.02	29.79	44.68	51.06	1.06	3.19
A	27	21.28	9.57	39.36	38.30	17.02	32.98	2.13	0.00	0.00	0.00	20.21	19.15
C	28	0.00	1.06	0.00	5.32	96.81	75.53	3.19	15.96	0.00	1.06	0.00	1.06
B	29	0.00	2.13	96.81	69.15	3.19	21.28	0.00	2.13	0.00	0.00	0.00	5.32
C	30	1.06	0.00	2.13	2.13	79.79	37.23	4.26	11.70	12.77	48.94	0.00	0.00
E	31	0.00	0.00	0.00	2.13	0.00	1.06	0.00	0.00	100.00	96.81	0.00	0.00
Omit	32	0.00	0.00	0.00	4.26	1.06	8.51	0.00	2.13	1.06	6.38	97.87	78.72
E	33	0.00	0.00	0.00	0.00	0.00	2.13	0.00	0.00	98.94	94.68	1.06	3.19
C	34	0.00	6.38	1.06	10.64	84.04	69.15	2.13	4.26	11.70	8.51	1.06	1.06
A	35	68.09	41.49	13.83	32.98	0.00	5.32	13.83	15.96	3.19	4.26	1.06	0.00
E	36	38.30	41.49	1.06	5.32	46.81	43.62	0.00	2.13	12.77	6.38	1.06	1.06
D	37	37.23	22.34	11.70	17.02	6.38	25.53	38.30	28.72	5.32	5.32	1.06	1.06
C	38	14.89	26.60	6.38	9.57	65.96	32.98	11.70	26.60	0.00	3.19	1.06	1.06
E	39	2.13	9.57	9.57	11.70	9.57	47.87	40.43	20.21	32.98	7.45	5.32	3.19
D	40	4.26	17.02	0.00	1.06	1.06	2.13	92.55	64.89	2.13	13.83	0.00	1.06
E	41	4.26	27.66	0.00	0.00	0.00	1.06	6.38	29.79	89.36	41.49	0.00	0.00
E	42	29.79	40.43	2.13	3.19	0.00	8.51	19.15	31.91	46.81	15.96	2.13	0.00
E	43	1.06	22.34	0.00	1.06	0.00	6.38	4.26	18.09	94.68	51.06	0.00	1.06
A	44	95.74	51.06	1.06	18.09	0.00	1.06	0.00	0.00	3.19	28.72	0.00	1.06
A	45	98.94	68.09	1.06	26.60	0.00	1.06	0.00	2.13	0.00	1.06	0.00	1.06
C	46	48.94	43.62	5.32	11.70	44.68	30.85	1.06	5.32	0.00	6.38	0.00	2.13
E	47	0.00	0.00	0.00	4.26	0.00	6.38	0.00	1.06	100.00	87.23	0.00	1.06
B	48	7.45	15.96	82.98	48.94	3.19	7.45	0.00	2.13	6.38	24.47	0.00	1.06
C	49	12.77	32.98	0.00	3.19	86.17	44.68	0.00	3.19	1.06	13.83	0.00	2.13
D	50	8.51	18.09	6.38	7.45	50.00	27.66	31.91	39.36	3.19	4.26	0.00	3.19
C	51	0.00	0.00	0.00	1.06	29.79	28.72	69.15	62.77	1.06	4.26	0.00	3.19
E	52	1.06	12.77	2.13	11.70	0.00	9.57	0.00	5.32	96.81	58.51	0.00	2.13
B	53	2.13	21.28	95.74	53.19	1.06	9.57	0.00	6.38	1.06	6.38	0.00	3.19
C	54	9.57	36.17	2.13	11.70	85.11	39.36	0.00	6.38	3.19	4.26	0.00	2.13
B	55	0.00	5.32	92.55	41.49	1.06	5.32	1.06	3.19	5.32	42.55	0.00	2.13
B	56	2.13	19.15	95.74	55.32	1.06	14.89	1.06	1.06	0.00	7.45	0.00	2.13
C	57	45.74	37.23	32.98	24.47	17.02	26.60	4.26	5.32	0.00	4.26	0.00	2.13
A	58	100.00	74.47	0.00	10.64	0.00	4.26	0.00	1.06	0.00	7.45	0.00	2.13
C	59	51.06	65.96	2.13	6.38	43.62	12.77	3.19	10.64	0.00	0.00	0.00	4.26
E	60	7.45	19.15	1.06	17.02	38.30	24.47	1.06	8.51	50.00	27.66	2.13	3.19

Key	Item	Frequency Table (post)											
		A		*B*		*C*		*D*		*E*		*Omit*	
		upper	*lower*	*upper*	*lower*	*upper*	*lower*	*upper*	*lower*	*upper*	*lower*	*upper*	*lower*
A	*61*	32.98	21.28	5.32	9.57	39.36	42.55	18.09	21.28	4.26	1.06	0.00	4.26
D	*62*	0.00	8.51	2.13	4.26	38.30	50.00	56.38	25.53	3.19	7.45	0.00	4.26
D	*63*	2.13	14.89	36.17	29.79	23.40	29.79	30.85	14.89	7.45	7.45	0.00	3.19
E	*64*	3.19	25.53	3.19	9.57	1.06	12.77	4.26	10.64	87.23	38.30	1.06	3.19
E	*65*	2.13	4.26	2.13	4.26	8.51	35.11	8.51	20.21	77.66	31.91	1.06	4.26
D	*66*	6.38	11.70	7.45	18.09	26.60	22.34	40.43	22.34	18.09	21.28	1.06	4.26
C	*67*	2.13	9.57	7.45	19.15	36.17	21.28	31.91	22.34	21.28	21.28	1.06	6.38
C	*68*	7.45	13.83	3.19	17.02	26.60	18.09	23.40	12.77	38.30	34.04	1.06	4.26
A	*69*	63.83	17.02	25.53	52.13	0.00	9.57	7.45	9.57	1.06	6.38	2.13	5.32
A	*70*	37.23	36.17	51.06	30.85	1.06	8.51	5.32	9.57	3.19	10.64	2.13	4.26

Table A–III: Item Analysis Table

Item	Item Difficulty		Item Discrimination		Item Variance		Modified Point Biserial Correlation		Item Reliability Index	
	pre	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*
1	0.46	0.62	0.28	0.52	0.25	0.24	0.22	0.38	0.11	0.18
2	0.55	0.65	0.33	0.49	0.25	0.23	0.22	0.36	0.11	0.17
3	0.40	0.63	0.23	0.50	0.24	0.23	0.14	0.37	0.07	0.18
4	0.52	0.69	0.35	0.48	0.25	0.21	0.21	0.37	0.10	0.17
5	0.16	0.58	0.19	0.61	0.14	0.24	0.22	0.42	0.08	0.21
6	0.15	0.57	0.16	0.62	0.13	0.25	0.20	0.42	0.07	0.21
7	0.19	0.45	0.16	0.59	0.15	0.25	0.16	0.45	0.06	0.23
8	0.21	0.42	0.15	0.43	0.16	0.24	0.13	0.34	0.05	0.17
9	0.20	0.42	0.22	0.44	0.16	0.24	0.20	0.35	0.08	0.17
10	0.36	0.81	0.43	0.38	0.23	0.16	0.35	0.35	0.17	0.14
11	0.56	0.78	0.57	0.30	0.25	0.17	0.36	0.22	0.18	0.09
12	0.25	0.43	0.32	0.54	0.19	0.24	0.34	0.41	0.14	0.20
13	0.41	0.80	0.40	0.36	0.24	0.16	0.31	0.33	0.15	0.13
14	0.59	0.77	0.54	0.31	0.24	0.18	0.35	0.23	0.17	0.10
15	0.28	0.46	0.35	0.55	0.20	0.25	0.32	0.40	0.14	0.20
16	0.38	0.57	0.41	0.48	0.24	0.25	0.28	0.32	0.14	0.16
17	0.35	0.54	0.39	0.54	0.23	0.25	0.28	0.33	0.13	0.17
18	0.27	0.39	0.39	0.53	0.20	0.24	0.27	0.38	0.12	0.18
19	0.22	0.38	0.27	0.42	0.17	0.24	0.20	0.29	0.08	0.14
20	0.31	0.41	0.40	0.46	0.21	0.24	0.30	0.32	0.14	0.16
21	0.17	0.23	0.15	0.26	0.14	0.18	0.08	0.23	0.03	0.10
22	0.26	0.30	0.37	0.47	0.19	0.21	0.26	0.37	0.11	0.17
23	0.21	0.35	0.23	0.43	0.17	0.23	0.20	0.33	0.08	0.16
24	0.32	0.29	0.19	0.19	0.22	0.21	0.13	0.13	0.06	0.06
25	0.46	0.43	0.22	0.15	0.25	0.25	0.12	0.10	0.06	0.05
26	0.09	0.09	0.02	0.17	0.08	0.08	0.10	0.26	0.03	0.07
27	0.13	0.12	0.01	0.12	0.11	0.11	0.00	0.11	0.00	0.04
28	0.39	0.86	0.40	0.21	0.24	0.12	0.32	0.22	0.16	0.08
29	0.67	0.84	0.31	0.28	0.22	0.13	0.21	0.27	0.10	0.10
30	0.31	0.54	0.37	0.42	0.21	0.25	0.28	0.30	0.13	0.15
31	0.89	0.99	0.20	0.03	0.10	0.01	0.20	0.10	0.06	0.01
32	0.79	0.90	0.28	0.19	0.16	0.09	0.18	0.15	0.07	0.05
33	0.84	0.97	0.24	0.04	0.14	0.03	0.22	0.07	0.08	0.01
34	0.59	0.79	0.35	0.15	0.24	0.17	0.18	0.09	0.09	0.04
35	0.41	0.56	0.26	0.27	0.24	0.25	0.13	0.16	0.07	0.08
36	0.13	0.07	-0.05	0.07	0.11	0.07	-0.06	0.12	-0.02	0.03
37	0.15	0.31	0.10	0.09	0.13	0.21	0.07	0.06	0.03	0.03
38	0.33	0.55	0.38	0.34	0.22	0.25	0.25	0.24	0.12	0.12
39	0.16	0.18	0.06	0.26	0.13	0.15	0.00	0.26	0.00	0.10
40	0.35	0.76	0.27	0.28	0.23	0.18	0.17	0.22	0.08	0.10
41	0.52	0.67	0.39	0.49	0.25	0.22	0.26	0.33	0.13	0.15
42	0.18	0.26	0.10	0.31	0.15	0.19	0.15	0.22	0.06	0.10
43	0.56	0.75	0.39	0.44	0.25	0.19	0.26	0.32	0.13	0.14
44	0.21	0.77	0.32	0.45	0.17	0.18	0.32	0.40	0.13	0.17
45	0.45	0.87	0.37	0.31	0.25	0.11	0.26	0.34	0.13	0.12
46	0.38	0.37	0.26	0.13	0.24	0.23	0.13	0.04	0.06	0.02
47	0.73	0.94	0.32	0.13	0.20	0.06	0.24	0.19	0.11	0.05
48	0.30	0.68	0.33	0.34	0.21	0.22	0.32	0.26	0.14	0.12
49	0.35	0.66	0.40	0.41	0.23	0.23	0.30	0.31	0.14	0.15
50	0.62	0.30	0.14	-0.07	0.24	0.21	0.01	-0.09	0.00	-0.04

Item	Item Difficulty		Item Discrimination		Item Variance		Modified Point Biserial Correlation		Item Reliability Index	
	pre	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*
51	0.34	0.22	0.03	0.01	0.22	0.17	-0.02	0.00	-0.01	0.00
52	0.43	0.78	0.52	0.38	0.24	0.17	0.38	0.32	0.19	0.13
53	0.29	0.74	0.53	0.43	0.20	0.19	0.43	0.35	0.19	0.15
54	0.39	0.67	0.45	0.46	0.24	0.22	0.31	0.35	0.15	0.16
55	0.29	0.69	0.44	0.51	0.21	0.21	0.38	0.37	0.17	0.17
56	0.36	0.77	0.60	0.40	0.23	0.18	0.46	0.36	0.22	0.15
57	0.33	0.23	0.21	-0.09	0.22	0.18	0.07	-0.12	0.03	-0.05
58	0.62	0.90	0.43	0.26	0.24	0.09	0.31	0.31	0.15	0.09
59	0.25	0.25	0.15	0.31	0.19	0.19	0.08	0.26	0.04	0.11
60	0.34	0.39	0.22	0.22	0.23	0.24	0.10	0.14	0.05	0.07
61	0.28	0.25	0.13	0.12	0.20	0.19	0.09	0.09	0.04	0.04
62	0.25	0.41	0.21	0.31	0.19	0.24	0.15	0.22	0.07	0.11
63	0.17	0.22	0.07	0.16	0.14	0.17	0.10	0.15	0.04	0.06
64	0.29	0.61	0.30	0.50	0.21	0.24	0.23	0.36	0.10	0.18
65	0.24	0.52	0.19	0.46	0.18	0.25	0.18	0.31	0.08	0.16
66	0.12	0.32	0.09	0.19	0.11	0.22	0.08	0.15	0.03	0.07
67	0.19	0.28	0.06	0.15	0.16	0.20	0.00	0.14	0.00	0.06
68	0.16	0.20	0.01	0.09	0.14	0.16	-0.03	0.07	-0.01	0.03
69	0.20	0.37	0.11	0.47	0.16	0.23	0.09	0.33	0.04	0.16
70	0.35	0.37	0.27	0.01	0.23	0.23	0.19	-0.01	0.09	-0.01
Ave.	**0.35**	**0.53**	**0.27**	**0.32**	**0.20**	**0.19**	**0.20**	**0.25**	**0.09**	**0.11**

Table A–IV is the matrix of correlations between items on Version 3.0 of the DEEM. The row and column headings are the item numbers. Highlighted groups of cells on the diagonal represent items that have “primary-level” relationship; highlighted groups of cells that are off the diagonal represent items that have a “secondary-” or “tertiary-level” relationship.

Table A–IV: Correlation Table

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1.00																							
2	0.86	1.00																						
3	0.61	0.54	1.00																					
4	0.58	0.61	0.75	1.00																				
5	0.06	0.04	0.14	0.13	1.00																			
6	0.09	0.03	0.15	0.14	0.98	1.00																		
7	0.18	0.19	0.14	0.15	0.50	0.50	1.00																	
8	0.10	0.07	0.16	0.10	0.28	0.27	0.23	1.00																
9	0.11	0.09	0.09	0.09	0.32	0.32	0.26	0.82	1.00															
10	0.20	0.17	0.21	0.16	0.07	0.09	0.10	0.13	0.12	1.00														
11	0.02	0.07	-0.03	0.01	0.05	0.06	0.08	0.05	0.10	0.24	1.00													
12	0.23	0.24	0.22	0.24	0.14	0.15	0.14	0.18	0.16	0.38	0.15	1.00												
13	0.20	0.17	0.21	0.16	0.11	0.11	0.12	0.14	0.11	0.86	0.16	0.38	1.00											
14	0.07	0.09	0.02	0.05	0.01	0.04	0.11	0.07	0.10	0.17	0.71	0.19	0.16	1.00										
15	0.23	0.23	0.20	0.18	0.15	0.15	0.14	0.13	0.13	0.36	0.16	0.85	0.41	0.15	1.00									
16	0.05	0.05	0.08	0.07	0.10	0.10	0.07	0.09	0.11	0.10	0.07	0.08	0.08	0.04	0.09	1.00								
17	0.08	0.11	0.07	0.10	0.11	0.12	0.07	0.06	0.06	0.09	0.12	0.09	0.06	0.14	0.11	0.58	1.00							
18	0.03	0.01	0.02	0.02	0.09	0.09	0.10	0.11	0.15	0.12	0.08	0.19	0.07	0.06	0.20	0.63	0.55	1.00						
19	0.09	0.06	0.07	0.09	0.15	0.16	0.17	0.11	0.08	0.03	0.07	0.11	0.03	0.13	0.06	0.26	0.19	0.23	1.00					
20	0.08	0.06	0.14	0.14	0.08	0.09	0.06	0.07	0.11	0.09	0.08	0.10	0.08	0.04	0.12	0.58	0.38	0.49	0.23	1.00				
21	0.09	0.09	0.07	0.14	0.07	0.07	0.04	0.11	0.15	0.06	-0.02	0.12	0.08	0.02	0.13	0.31	0.29	0.21	0.09	0.40	1.00			
22	0.05	0.03	0.07	0.06	0.12	0.13	0.05	0.14	0.18	0.10	0.06	0.23	0.10	0.07	0.20	0.46	0.39	0.68	0.20	0.68	0.29	1.00		
23	0.05	0.06	0.04	0.10	0.20	0.21	0.18	0.12	0.11	0.05	0.07	0.16	0.07	0.14	0.11	0.24	0.17	0.19	0.82	0.21	0.09	0.20	1.00	
24	0.05	0.06	0.07	0.08	0.03	0.00	0.03	0.03	0.02	-0.03	0.00	-0.05	-0.04	-0.05	-0.05	0.05	0.01	0.10	0.10	-0.08	-0.02	-0.05	0.10	1.00
25	0.08	0.07	0.13	0.15	0.05	0.04	0.03	0.00	-0.01	-0.05	0.01	0.01	-0.07	0.04	0.01	-0.03	-0.03	0.02	0.05	-0.01	-0.05	-0.01	0.07	0.60
26	0.11	0.12	0.19	0.16	0.04	0.04	0.07	0.13	0.15	0.10	0.01	0.00	0.07	0.00	0.05	0.00	-0.02	0.05	0.05	0.12	0.08	0.11	0.01	0.32
27	-0.04	-0.02	-0.01	-0.04	0.03	0.02	0.00	0.01	0.00	0.05	0.09	-0.02	0.03	0.05	0.07	-0.05	0.04	-0.01	-0.02	-0.01	0.15	0.01	-0.01	0.21
28	0.26	0.22	0.24	0.23	0.14	0.15	0.21	0.08	0.05	0.01	0.07	0.02	-0.01	0.02	-0.01	0.00	0.07	0.02	0.13	0.06	0.03	-0.02	0.12	-0.02
29	0.07	0.10	0.15	0.09	0.18	0.16	0.19	0.07	0.09	0.05	0.20	0.11	0.07	0.22	0.11	0.07	0.06	0.15	0.00	0.07	-0.01	0.16	0.01	0.08
30	0.13	0.13	0.11	0.14	0.15	0.15	0.22	0.09	0.10	-0.05	0.01	0.40	-0.07	0.03	0.35	0.11	0.10	0.22	0.15	0.08	0.05	0.10	0.16	0.06
31	-0.01	-0.07	-0.01	0.01	0.11	0.11	0.08	0.02	0.02	0.03	0.03	0.08	0.03	0.10	0.09	0.11	0.04	0.07	0.07	0.08	0.05	0.06	0.00	0.06
32	0.10	0.04	0.09	0.06	0.09	0.11	0.19	0.00	0.04	-0.07	0.03	0.08	-0.03	0.13	0.08	0.10	0.08	0.06	0.05	0.09	0.03	0.04	0.07	0.05
33	0.08	0.05	0.08	0.07	0.03	0.06	0.09	-0.06	-0.06	0.05	0.03	-0.06	0.05	0.07	-0.08	-0.01	0.01	0.03	0.03	-0.03	-0.07	0.00	0.02	0.00
34	0.05	0.11	0.04	0.07	-0.01	0.01	0.01	-0.02	-0.02	0.05	0.17	0.03	0.03	0.18	0.06	-0.02	0.10	0.07	-0.01	0.03	-0.05	0.06	0.01	-0.03
35	0.15	0.12	0.10	0.18	0.01	0.02	0.03	0.05	0.04	0.23	0.04	0.09	0.19	-0.01	0.11	0.10	0.06	0.07	0.03	0.05	0.03	0.06	-0.01	0.05
36	0.07	0.03	0.06	0.07	0.13	0.11	0.12	0.11	0.09	-0.11	-0.06	0.02	-0.08	-0.03	0.03	-0.02	-0.02	0.02	0.09	-0.02	0.08	0.05	0.09	0.06
37	-0.06	-0.07	-0.11	-0.15	0.12	0.13	0.15	0.12	0.12	-0.05	0.01	-0.05	-0.05	0.04	-0.07	0.03	-0.02	0.09	0.05	-0.01	-0.05	0.06	0.01	0.08
38	0.07	0.05	0.12	0.10	0.19	0.21	0.14	0.12	0.11	0.04	0.06	0.12	0.00	0.06	0.11	0.08	0.07	0.11	0.09	0.06	0.06	0.07	0.11	-0.02
39	0.08	0.05	0.16	0.14	0.10	0.11	0.18	0.16	0.17	0.14	-0.02	0.20	0.14	-0.01	0.23	-0.01	-0.01	0.00	0.01	0.00	-0.03	-0.02	0.03	0.06
40	0.12	0.04	-0.01	0.03	0.28	0.30	0.30	0.06	0.12	0.05	0.03	0.08	0.03	0.04	0.06	0.08	0.07	0.09	0.12	0.14	-0.05	0.12	0.14	0.07
41	0.17	0.18	0.10	0.09	0.33	0.30	0.28	0.16	0.18	0.14	0.07	0.14	0.16	0.14	0.10	0.00	-0.02	0.01	0.10	0.02	0.07	0.01	0.13	0.02
42	0.02	0.00	0.11	0.02	0.14	0.15	0.15	0.13	0.11	-0.02	0.05	0.17	-0.02	0.09	0.18	0.12	0.11	0.22	0.15	0.07	0.14	0.12	0.13	0.04
43	0.21	0.22	0.14	0.14	0.21	0.20	0.17	0.16	0.19	0.13	0.07	0.14	0.17	0.12	0.11	0.04	0.04	0.09	0.05	0.01	0.08	0.07	0.13	0.03
44	0.19	0.20	0.16	0.17	0.10	0.11	0.19	0.10	0.07	0.43	0.14	0.24	0.37	0.16	0.20	0.06	0.18	0.12	-0.02	0.04	0.09	0.06	0.08	0.03
45	0.11	0.12	0.16	0.13	0.06	0.07	0.13	0.11	0.04	0.50	0.02	0.20	0.47	0.03	0.20	0.05	0.13	0.10	-0.01	0.02	0.05	0.04	0.07	0.04
46	-0.02	-0.01	0.02	0.04	0.01	0.00	0.00	-0.11	-0.16	0.06	0.06	0.04	0.10	0.04	0.07	0.00	-0.03	0.02	-0.02	0.02	0.00	0.05	0.01	-0.03
47	-0.02	-0.04	0.05	-0.01	0.03	0.05	0.08	0.14	0.12	0.09	0.01	0.02	0.09	0.09	-0.01	0.05	0.13	0.08	0.03	0.04	0.08	0.11	0.01	0.00
48	0.04	0.03	0.04	0.02	0.04	0.05	0.12	0.18	0.14	0.27	0.17	0.11	0.23	0.11	0.12	0.13	0.20	0.11	0.01	0.08	0.09	0.05	0.05	-0.02
49	0.06	0.05	0.06	0.04	0.03	0.05	0.22	0.10	0.09	0.12	0.13	0.02	0.07	0.10	0.06	0.15	0.20	0.17	-0.07	0.07	0.12	0.09	0.01	0.04
50	0.03	-0.01	-0.05	0.01	-0.03	-0.04	-0.12	0.00	-0.04	-0.09	-0.08	-0.06	-0.09	-0.09	-0.07	0.04	-0.05	-0.03	0.23	0.06	0.01	0.00	0.18	0.07
51	-0.04	-0.05	-0.09	-0.08	0.04	0.03	-0.08	0.03	-0.01	-0.07	0.00	-0.07	-0.09	-0.06	-0.06	0.05	-0.06	0.01	0.18	0.11	0.02	0.12	0.12	0.15
52	0.11	0.15	0.13	0.12	0.14	0.12	0.19	0.06	0.04	0.06	0.04	0.10	0.06	-0.03	0.13	0.08	0.09	0.06	0.00	0.02	0.05	0.04	0.06	0.02
53	0.12	0.15	0.14	0.15	0.13	0.12	0.21	0.13	0.12	0.11	0.10	0.15	0.06	0.06	0.13	-0.01	0.08	0.06	-0.03	0.03	-0.07	0.10	-0.01	-0.07
54	0.12	0.14	0.10	0.09	0.10	0.09	0.18	0.02	0.07	0.11	0.09	0.08	0.15	0.04	0.12	0.07	0.14	0.10	-0.07	0.03	0.08	0.09	0.01	-0.03
55	0.19	0.23	0.20	0.21	0.18	0.17	0.27	0.09	0.10	0.13	0.05	0.08	0.08	0.02	0.08	0.05	0.11	0.04	-0.03	0.04	-0.05	0.02	0.03	0.03
56	0.13	0.21	0.15	0.17	0.18	0.16	0.26	0.12	0.11	0.10	0.04	0.07	0.09	0.01	0.09	0.03	0.07	0.05	-0.03	0.03	-0.07	0.08	0.03	0.00
57	-0.10	-0.13	-0.09	-0.11	-0.04	-0.01	-0.09	-0.04	-0.08	-0.07	-0.08	-0.01	-0.03	0.00	-0.05	0.00	-0.02	0.00	0.07	-0.01	0.03	0.05	0.05	-0.04
58	0.01	0.06	0.17	0.12	0.11	0.09	0.15	0.06	0.06	0.19	0.09	0.08	0.17	0.11	0.05	0.08	0.16	0.12	0.05	0.04	0.03	0.10	0.07	0.03
59	0.06	0.08	0.00	-0.01	0.14	0.12	0.13	0.14	0.13	0.03	0.08	0.11	-0.02	0.04	0.08	0.19	0.12	0.20	0.17	0.21	0.10	0.20	0.20	0.12
60	0.00	0.01	0.03	0.06	0.02	0.01	0.11	0.02	0.03	0.03	0.13	0.05	0.00	0.05	0.04	-0.02	-0.07	0.03	0.01	0.02	-0.01	0.02	0.03	0.07
61	0.00	-0.01	0.03	-0.05	0.06	0.05	0.07	0.04	0.03	0.09	0.05	0.04	0.06	0.01	0.06	0.08	0.04	0.07	0.17	0.09	0.05	0.09	0.11	0.00
62	0.11	0.07	0.04	0.07	0.08	0.09	0.08	0.06	0.13	0.07	-0.02	0.07	0.09	-0.02	0.05	0.00	0.02	0.01	0.00	0.01	0.05	0.03	0.05	0.01
63	0.15	0.09	0.05	0.04	0.01	0.01	0.06	0.08	0.08	0.00	-0.12	0.03	0.05	-0.04	0.03	0.06	0.01	0.05	0.05	0.04	0.04	0.07	0.02	-0.05
64	0.14	0.14	0.13	0.16	0.32	0.30	0.27	0.12	0.17	0.09	0.06	0.10	0.13	0.07	0.09	0.02	0.03	-0.03	0.05	0.05	0.07	0.00	0.08	0.14
65	0.10	0.09	0.08	0.04	0.25	0.27	0.29	0.05	0.06	0.00	0.03	0.17	-0.03	-0.02	0.19	-0.04	0.10	0.06	-0.01	0.01	0.03	0.01	0.02	0.02
66	0.05	0.05	0.00	0.04	-0.04	-0.05	-0.03	0.11	0.09	0.10	0.05	0.06	0.12	0.10	0.07	-0.03	0.03	0.04	0.13	-0.01	0.07	0.05	0.14	0.05
67	0.03	0.04	0.09	0.03	-0.02	-0.02	-0.05	0.01	0.01	0.06	0.15	0.07	0.04	0.10	0.08	0.07	0.11	0.07	-0.03	0.12	0.11	0.14	-0.02	0.02
68	0.00	0.02	0.00	0.02	0.03	0.02	-0.01	-0.01	0.00	-0.03	0.09	0.01	-0.05	0.04	0.03	0.06	0.01	0.03	0.03	0.04	0.06	0.09	0.00	0.04
69	0.12	0.14	0.11	0.15	0.12	0.11	0.19	0.19	0.12	0.12	-0.03	0.10	0.09	0.04	0.12	0.05	0.15	0.07	0.12	0.02	0.05	0.08	0.13	0.11
70	-0.06	-0.01	-0.01	0.01	-0.13	-0.13	-0.04	-0.02	-0.01	-0.04	-0.01	-0.05	-0.01	-0.01	-0.04	-0.01	0.00	-0.06	-0.03	0.00	0.01	-0.06	-0.04	0.01

	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
25	1.00																						
26	0.27	1.00																					
27	0.20	0.29	1.00																				
28	0.05	0.07	0.07	1.00																			
29	0.12	0.05	0.11	0.05	1.00																		
30	0.06	-0.01	0.04	0.38	0.04	1.00																	
31	-0.04	0.03	0.03	-0.04	0.05	0.04	1.00																
32	-0.03	-0.03	-0.02	0.00	0.11	0.16	0.27	1.00															
33	0.05	0.05	0.01	0.08	0.16	0.02	-0.02	0.28	1.00														
34	0.00	0.03	0.04	0.05	0.07	0.07	0.03	-0.01	0.04	1.00													
35	0.00	0.09	0.13	0.09	0.02	0.08	-0.02	-0.05	0.02	0.17	1.00												
36	0.08	-0.01	0.06	0.08	0.12	0.08	0.03	0.06	-0.02	-0.26	-0.10	1.00											
37	0.03	0.02	0.00	0.04	0.06	0.10	0.06	0.04	-0.03	-0.05	-0.15	0.07	1.00										
38	-0.07	0.05	0.00	0.03	0.06	0.05	0.04	0.03	0.05	0.06	0.06	-0.03	0.04	1.00									
39	0.01	0.17	0.05	0.08	0.08	0.15	-0.04	0.01	0.04	0.06	0.04	0.04	0.03	0.07	1.00								
40	0.06	0.03	0.02	0.09	0.12	0.08	0.09	0.12	0.19	-0.02	0.04	0.06	0.03	0.09	0.05	1.00							
41	0.07	0.09	0.09	0.10	0.10	0.12	0.00	0.12	0.03	0.06	0.00	0.08	0.11	0.06	0.13	0.00	1.00						
42	0.03	0.05	0.04	0.00	0.02	0.18	0.06	0.03	-0.05	-0.11	-0.05	0.20	0.06	0.08	0.09	-0.10	0.20	1.00					
43	0.03	0.08	0.05	0.04	0.12	0.05	0.02	0.07	-0.02	0.10	-0.01	0.06	0.05	0.00	0.06	0.07	0.63	0.25	1.00				
44	0.00	0.12	0.08	0.03	0.18	0.05	0.10	0.02	0.07	0.04	0.14	0.08	0.00	0.16	0.12	0.11	0.18	0.05	0.19	1.00			
45	-0.04	0.09	0.07	-0.04	0.07	0.00	0.15	0.01	0.03	0.01	0.10	0.01	-0.05	0.07	0.12	0.04	0.12	0.00	0.13	0.69	1.00		
46	-0.01	-0.07	0.00	-0.10	0.00	0.00	0.07	0.07	-0.04	0.01	0.07	-0.06	-0.07	-0.09	-0.01	-0.03	-0.10	0.05	-0.01	-0.07	0.09	1.00	
47	-0.05	0.08	-0.05	0.00	0.09	-0.04	0.11	-0.01	0.03	-0.01	0.07	0.03	0.12	0.01	0.03	0.06	0.10	-0.04	0.05	0.23	0.29	-0.05	1.00
48	-0.07	0.04	-0.01	0.04	0.06	0.00	0.00	-0.01	-0.12	0.01	0.04	0.06	0.05	0.07	0.05	-0.01	0.08	-0.02	0.05	0.38	0.40	0.06	0.21
49	-0.07	0.09	0.03	0.06	0.09	0.10	-0.07	0.07	-0.05	0.00	0.06	-0.02	0.07	0.14	0.12	-0.02	0.07	0.06	0.09	0.25	0.18	0.20	0.17
50	0.09	0.04	-0.01	-0.08	-0.18	-0.03	-0.01	-0.04	0.00	-0.11	-0.01	-0.12	-0.09	-0.09	-0.07	-0.08	-0.07	0.01	-0.07	-0.24	-0.11	0.14	-0.15
51	0.14	0.13	0.01	-0.06	-0.10	-0.01	-0.03	0.00	0.05	-0.12	0.00	-0.07	-0.07	-0.09	0.02	-0.05	-0.02	0.06	-0.03	-0.22	-0.14	0.12	-0.07
52	0.03	0.11	-0.06	0.15	0.14	0.11	-0.05	-0.04	0.04	0.06	-0.01	0.04	0.05	0.09	0.14	0.09	0.09	0.03	0.07	0.19	0.17	-0.07	0.25
53	-0.06	0.11	-0.02	0.16	0.27	0.07	-0.06	0.01	0.09	0.11	-0.01	0.12	0.00	0.11	0.18	0.07	0.07	0.04	0.13	0.27	0.18	-0.06	0.18
54	-0.11	0.08	0.03	0.07	0.16	0.10	-0.07	0.06	0.06	0.06	0.04	-0.01	0.02	0.10	0.15	0.02	0.11	-0.02	0.17	0.19	0.18	0.15	0.16
55	0.06	0.12	0.02	0.28	0.23	0.14	-0.06	0.02	0.18	0.05	0.02	0.10	0.03	0.13	0.14	0.18	0.10	-0.04	0.16	0.25	0.18	-0.08	0.17
56	0.03	0.14	0.01	0.15	0.24	0.10	-0.05	0.02	0.19	0.12	0.03	0.10	0.04	0.11	0.15	0.15	0.05	-0.03	0.07	0.30	0.23	-0.09	0.18
57	-0.01	-0.02	0.03	-0.10	-0.05	-0.07	-0.02	0.00	-0.07	-0.02	0.04	-0.02	0.02	-0.02	-0.02	-0.02	-0.04	0.05	-0.01	0.00	0.03	0.18	-0.04
58	0.01	0.04	0.07	0.00	0.30	0.09	0.07	0.04	0.11	0.10	0.04	0.06	0.04	0.09	0.04	0.01	0.14	-0.03	0.13	0.31	0.31	0.01	0.31
59	0.12	0.08	0.09	-0.02	-0.03	0.02	0.05	0.06	-0.02	-0.03	-0.04	0.01	0.01	0.01	0.08	0.08	0.19	0.08	0.16	0.03	0.05	0.07	0.03
60	0.11	0.16	0.03	0.12	0.11	0.06	0.07	0.04	0.03	-0.01	-0.02	0.09	0.06	-0.06	0.06	0.11	0.03	0.07	0.04	0.10	0.08	0.06	0.05
61	0.04	0.01	0.05	-0.02	-0.09	-0.04	0.05	-0.02	-0.02	-0.06	-0.04	0.04	0.06	0.03	0.08	-0.04	0.05	0.10	0.04	-0.16	-0.03	0.09	0.01
62	-0.08	0.09	-0.03	0.09	0.04	0.05	0.08	0.00	-0.03	0.08	0.11	0.00	0.05	0.15	0.12	0.11	0.11	0.00	0.14	0.17	0.12	0.04	0.04
63	-0.03	0.03	0.01	0.00	-0.08	0.08	0.05	0.02	-0.03	0.07	0.04	0.03	0.02	0.05	0.02	0.10	0.05	0.03	0.11	0.06	0.09	0.06	-0.01
64	0.06	0.12	0.04	0.00	0.16	0.07	0.05	0.10	0.04	0.11	0.08	0.05	0.07	0.20	0.12	0.15	0.20	0.10	0.16	0.14	0.14	0.02	0.12
65	0.02	0.07	-0.04	0.09	0.04	0.13	0.04	0.07	-0.06	0.13	0.04	0.07	0.00	0.20	0.24	0.23	0.11	0.15	0.12	0.10	0.12	0.01	0.15
66	-0.03	0.05	0.05	-0.01	0.04	0.05	0.00	0.07	-0.07	0.00	0.05	0.06	0.03	0.08	0.11	-0.12	0.12	0.07	0.05	0.08	0.03	-0.04	0.02
67	0.01	0.03	0.06	0.05	0.06	0.05	-0.08	-0.02	0.03	0.01	0.05	-0.03	-0.03	0.09	0.05	0.11	-0.05	0.02	-0.04	0.05	0.02	0.00	0.02
68	0.00	0.05	0.08	0.03	-0.03	0.10	0.05	0.00	-0.05	0.08	0.00	-0.03	0.05	0.10	0.03	-0.02	-0.01	0.02	-0.07	0.03	0.04	-0.01	-0.03
69	0.11	0.17	0.10	0.14	0.02	0.14	0.01	0.03	-0.05	0.01	0.13	0.15	0.10	0.15	0.07	0.03	0.18	0.16	0.11	0.18	0.14	0.03	0.02
70	-0.09	-0.01	0.00	-0.01	-0.05	-0.02	0.01	0.07	0.13	-0.05	0.07	0.05	-0.05	0.07	0.07	0.01	-0.10	0.04	-0.11	-0.01	-0.07	0.03	-0.03

	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
48	1.00																						
49	0.49	1.00																					
50	-0.16	-0.20	1.00																				
51	-0.17	-0.14	0.67	1.00																			
52	0.15	0.21	-0.16	-0.14	1.00																		
53	0.13	0.22	-0.24	-0.06	0.49	1.00																	
54	0.17	0.46	-0.21	-0.10	0.36	0.43	1.00																
55	0.11	0.18	-0.22	-0.12	0.66	0.62	0.38	1.00															
56	0.09	0.17	-0.19	-0.09	0.54	0.74	0.32	0.74	1.00														
57	-0.01	0.06	0.08	0.01	-0.48	-0.24	-0.07	-0.48	-0.27	1.00													
58	0.15	0.13	-0.11	-0.09	0.23	0.19	0.27	0.22	0.22	0.00	1.00												
59	0.08	0.04	0.22	0.30	0.11	0.07	0.00	0.05	-0.02	-0.04	0.00	1.00											
60	-0.02	0.02	-0.07	-0.05	0.14	0.10	0.04	0.14	0.14	-0.08	0.10	0.08	1.00										
61	-0.10	-0.02	0.25	0.24	-0.02	-0.03	0.05	-0.05	-0.07	0.05	-0.02	0.35	-0.10	1.00									
62	0.11	0.15	-0.17	-0.10	0.06	0.12	0.27	0.10	0.09	-0.01	0.13	-0.06	0.10	-0.01	1.00								
63	0.06	0.11	0.03	-0.02	0.03	0.02	0.05	0.00	0.05	0.11	0.02	-0.02	0.00	-0.02	0.37	1.00							
64	0.06	0.10	-0.05	0.00	0.24	0.20	0.17	0.26	0.22	-0.12	0.17	0.10	0.17	-0.02	0.17	0.06	1.00						
65	0.07	0.18	-0.10	-0.09	0.29	0.27	0.24	0.22	0.19	-0.07	0.13	0.03	0.13	-0.07	0.20	0.11	0.34	1.00					
66	0.09	0.11	0.04	0.04	0.03	0.04	0.15	0.00	0.02	0.03	0.07	0.10	-0.01	0.05	0.04	0.07	0.07	0.07	1.00				
67	0.04	0.05	0.06	0.08	0.02	0.02	0.04	-0.02	0.01	0.03	0.11	0.06	0.07	-0.03	0.08	0.05	-0.02	0.03	0.05	1.00			
68	0.08	-0.01	0.01	0.03	0.01	-0.01	0.08	-0.06	0.01	0.04	0.00	0.04	-0.03	0.05	0.01	0.07	-0.02	-0.01	-0.04	0.37	1.00		
69	0.14	0.15	-0.05	-0.03	0.13	0.18	0.13	0.19	0.18	-0.03	0.10	0.06	0.12	-0.02	0.13	0.01	0.21	0.19	0.19	-0.04	-0.05	1.00	
70	-0.08	0.08	0.01	0.05	0.04	0.04	0.15	0.07	0.07	-0.05	-0.03	-0.07	0.02	0.01	0.13	0.12	0.02	0.06	0.00	0.03	-0.05	0.06	1.00

APPENDIX B

Table B–I contains data relating to gender biases for each item on the DEEM (Version 3.0). At the bottom of each column the mean (μ) and the standard deviation of each column are listed. For every element, a number greater then zero indicates a bias against males, a zero is no detectable gender bias, and a number less than zero indicates a bias against females.

The *Chi-square* columns lists the results of a Signed Scheuneman Statistic with two intervals (i.e. two degrees of freedom). The intervals were different for the pre- and the post-test. The investigator utilized only the post-test data when making a decision concerning the bias of an item. The darkened cells are data that did not meet the minimum criteria to make them statistically sound. They should be disregarded and other techniques should be consulted when determining the bias for that item. The averages and the standard deviations, listed at the bottom of the column, are only for those data points that met the criteria for sound statistics. No mean or standard deviation was calculated for the pre-test since none of the items were tested for their soundness.

Details concerning the techniques used to calculate the numbers in Table B–I can be found in Chapter 12.

Table B–I: Results for Item Bias Tests

Item	Phi Coefficent		Item Difficulty		Chi-square		Delta	
	"+" is female		*female-male*		Scheuneman (+ is f)		"+" is female	
	pre	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*
1	-0.02	-0.17	-0.02	-0.19	0.00	-1.00	0.16	-0.90
2	0.00	-0.21	0.00	-0.23	0.14	-1.33	0.20	-1.18
3	-0.01	-0.13	-0.01	-0.14	0.02	-0.29	0.29	-0.55
4	0.06	-0.12	0.07	-0.12	0.55	-0.22	0.73	-0.48
5	0.06	0.00	0.05	0.00	2.14	0.68	1.22	0.51
6	0.07	0.01	0.06	0.01	2.69	0.99	1.33	0.57
7	-0.10	-0.11	-0.09	-0.12	-1.06	-0.12	-0.35	-0.36
8	-0.03	-0.07	-0.03	-0.08	-0.43	-0.06	0.35	-0.08
9	0.03	-0.05	0.02	-0.06	1.42	0.01	0.83	0.09
10	-0.07	0.06	-0.08	0.05	-0.56	0.33	-0.19	0.98
11	-0.11	0.00	-0.13	0.00	-0.56	0.09	-0.67	0.41
12	-0.08	-0.18	-0.08	-0.20	-0.48	-1.37	-0.18	-0.97
13	-0.06	0.01	-0.07	0.01	-0.18	0.14	-0.15	0.54
14	-0.04	-0.07	-0.05	-0.07	-0.14	-0.23	-0.15	-0.16
15	-0.04	-0.16	-0.04	-0.18	-0.05	-0.96	0.15	-0.78
16	-0.11	0.00	-0.12	0.00	-0.87	0.18	-0.52	0.49
17	-0.12	-0.06	-0.13	-0.06	-1.01	-0.09	-0.60	0.06
18	-0.09	-0.06	-0.10	-0.06	-0.97	0.00	-0.31	0.07
19	-0.06	-0.13	-0.06	-0.14	-0.61	-1.58	0.01	-0.55
20	-0.07	0.08	-0.08	0.08	-0.55	1.52	-0.13	1.14
21	0.01	0.00	0.01	0.00	0.15	0.37	0.76	0.61
22	-0.09	-0.03	-0.09	-0.03	-0.60	0.19	-0.28	0.29
23	-0.08	-0.10	-0.07	-0.11	-1.27	-0.78	-0.15	-0.32
24	0.04	-0.04	0.04	-0.04	0.40	-0.08	0.79	0.20
25	-0.03	0.02	-0.03	0.02	-0.10	0.24	0.09	0.68
26	0.02	-0.02	0.01	-0.01	0.20	0.27	0.97	0.42
27	-0.02	-0.04	-0.01	-0.03	-0.04	-1.44	0.53	0.20
28	-0.15	-0.01	-0.17	0.00	-1.71	0.03	-0.84	0.37
29	-0.11	-0.15	-0.12	-0.12	-0.41	-0.23	-0.76	-0.87
30	-0.12	-0.12	-0.13	-0.14	-1.93	-0.30	-0.57	-0.45
31	-0.08	-0.02	-0.06	0.00	-0.16	0.00	-0.98	-0.06
32	-0.04	0.05	-0.03	0.04	-0.15	0.08	-0.37	0.99
33	-0.01	-0.02	-0.01	-0.01	0.00	-0.01	-0.19	0.10
34	-0.07	-0.05	-0.08	-0.05	-0.59	-0.02	-0.38	0.01
35	-0.02	-0.01	-0.02	-0.01	-0.03	0.04	0.22	0.44
36	0.01	-0.09	0.00	-0.06	0.01	-0.80	0.78	-0.73
37	-0.10	-0.02	-0.08	-0.03	-1.25	0.19	-0.40	0.35
38	-0.03	-0.08	-0.03	-0.09	-0.02	-0.24	0.21	-0.16
39	-0.04	-0.13	-0.03	-0.11	-0.82	-1.07	0.27	-0.74
40	-0.07	0.05	-0.08	0.05	-0.33	0.34	-0.15	0.92
41	-0.10	-0.20	-0.12	-0.21	-2.17	-1.02	-0.56	-1.11
42	-0.10	-0.13	-0.09	-0.13	-1.58	-2.52	-0.41	-0.64
43	-0.13	-0.27	-0.15	-0.27	-2.63	-1.70	-0.82	-1.73

Item	Phi Coefficent		Item Difficulty		Chi-square		Delta	
	"+" is female		*female-male*		Scheuneman (+ is f)		"+" is female	
	pre	*post*	*pre*	*post*	*pre*	*post*	*pre*	*post*
44	-0.20	-0.11	-0.18	-0.11	-4.39	-0.12	-1.52	-0.50
45	-0.08	-0.08	-0.09	-0.06	-0.51	-0.03	-0.32	-0.33
46	0.02	0.15	0.02	0.17	0.35	2.15	0.55	1.77
47	-0.05	-0.07	-0.05	-0.04	-0.16	-0.02	-0.36	-0.44
48	-0.02	0.00	-0.03	0.00	0.16	0.33	0.28	0.48
49	-0.07	0.01	-0.07	0.01	-0.64	0.24	-0.13	0.58
50	0.10	0.11	0.11	0.12	0.89	0.95	0.96	1.47
51	0.04	0.06	0.04	0.06	0.78	0.69	0.74	1.14
52	-0.12	-0.15	-0.14	-0.14	-2.59	-0.35	-0.65	-0.84
53	-0.24	-0.20	-0.25	-0.20	-6.24	-0.83	-1.93	-1.21
54	-0.13	-0.12	-0.15	-0.13	-1.88	-0.31	-0.69	-0.48
55	-0.18	-0.18	-0.19	-0.19	-3.65	-0.84	-1.16	-1.01
56	-0.27	-0.19	-0.29	-0.18	-6.24	-0.72	-2.02	-1.13
57	0.02	0.09	0.02	0.08	0.45	0.47	0.61	1.32
58	-0.06	-0.19	-0.07	-0.13	-0.36	-0.32	-0.32	-1.38
59	0.08	-0.04	0.08	-0.03	1.24	-0.03	1.22	0.26
60	-0.04	0.02	-0.04	0.02	-0.21	0.51	0.11	0.68
61	0.05	-0.07	0.05	-0.06	0.73	-1.77	0.88	-0.02
62	0.08	-0.07	0.08	-0.08	2.54	-0.23	1.22	-0.05
63	-0.02	-0.02	-0.02	-0.01	-0.13	0.05	0.48	0.44
64	0.04	-0.11	0.04	-0.12	0.74	-0.33	0.79	-0.37
65	-0.04	-0.07	-0.03	-0.08	-0.07	-0.09	0.25	-0.04
66	-0.08	-0.09	-0.06	-0.10	-1.39	-1.00	-0.14	-0.25
67	-0.02	0.05	-0.02	0.05	-0.03	0.59	0.41	1.00
68	-0.03	-0.03	-0.02	-0.02	-0.25	0.01	0.40	0.34
69	-0.04	-0.08	-0.03	-0.09	-0.09	-0.09	0.28	-0.16
70	-0.04	0.04	-0.04	0.05	-0.05	0.49	0.13	0.91
μ	-0.05	-0.06	-0.05	-0.06		-0.08	0.00	0.00
sd	0.07	0.08	0.08	0.09		0.67	0.69	0.73

Table B–II: Frequency Table Displaying Gender Differences

Item	Frequency Table											
	A		*B*		*C*		*D*		*E*		*Omit*	
	f	*m*	*f*	*m*	*f*	*m*	*f*	*m*	*f*	*m*	*f*	*m*
1	6.7	2.3	3.3	4.3	14.4	14.8	47.8	66.9	27.8	11.7	0.0	0.0
2	4.4	4.7	5.6	2.7	47.8	70.4	15.6	12.1	26.7	10.1	0.0	0.0
3	26.7	16.0	52.2	66.5	8.9	3.5	5.6	1.9	6.7	12.1	0.0	0.0
4	60.0	72.4	20.0	14.8	11.1	2.7	3.3	2.3	5.6	7.8	0.0	0.0
5	3.3	3.5	5.6	6.6	20.0	14.4	13.3	17.1	57.8	57.6	0.0	0.8
6	7.8	6.2	1.1	2.7	13.3	18.7	20.0	14.8	57.8	56.8	0.0	0.8
7	7.8	7.4	10.0	8.6	20.0	16.0	24.4	18.7	35.6	47.9	2.2	1.6
8	4.4	5.1	20.0	12.1	35.6	44.0	26.7	29.2	11.1	8.2	2.2	1.6
9	20.0	11.3	4.4	6.6	25.6	28.4	37.8	44.0	10.0	8.2	2.2	1.6
10	84.4	79.4	7.8	12.8	2.2	1.9	1.1	0.4	4.4	5.4	0.0	0.0
11	4.4	4.3	77.8	78.2	6.7	4.7	0.0	0.8	0.0	0.4	11.1	11.7
12	27.8	47.9	7.8	8.9	1.1	0.4	0.0	0.8	63.3	40.9	0.0	1.2
13	10.0	13.6	81.1	80.2	0.0	0.8	3.3	0.4	5.6	4.7	0.0	0.4
14	6.7	4.3	72.2	79.0	8.9	3.9	0.0	1.2	0.0	1.2	12.2	10.5
15	7.8	8.2	32.2	50.2	0.0	0.4	1.1	0.4	58.9	40.5	0.0	0.4
16	12.2	11.3	10.0	6.6	7.8	11.3	56.7	56.8	11.1	12.8	2.2	1.2
17	14.4	10.5	48.9	55.3	12.2	12.5	2.2	1.6	1.1	0.4	21.1	19.8
18	22.2	24.1	8.9	9.3	34.4	40.9	3.3	1.6	7.8	3.9	23.3	20.2
19	35.6	28.0	16.7	10.1	7.8	10.9	27.8	42.0	8.9	7.0	3.3	1.9
20	8.9	10.9	11.1	6.2	7.8	10.9	22.2	31.9	47.8	39.3	2.2	0.8
21	23.3	23.0	42.2	44.0	8.9	12.5	1.1	1.6	2.2	0.0	22.2	19.1
22	18.9	25.3	27.8	31.1	18.9	19.5	7.8	2.7	6.7	2.7	20.0	18.7
23	14.4	10.5	41.1	31.5	6.7	11.3	26.7	37.7	10.0	7.4	1.1	1.6
24	5.6	4.3	20.0	23.0	27.8	22.6	20.0	19.5	25.6	30.0	1.1	0.8
25	7.8	8.6	44.4	42.4	20.0	18.3	1.1	2.3	2.2	1.2	24.4	27.2
26	6.7	3.1	14.4	18.3	7.8	8.9	27.8	24.5	42.2	44.0	1.1	1.2
27	10.0	12.8	36.7	37.4	33.3	26.1	0.0	1.2	0.0	0.4	20.0	22.2
28	0.0	0.4	3.3	4.3	85.6	86.0	10.0	8.6	1.1	0.4	0.0	0.4
29	4.4	0.0	75.6	87.5	17.8	8.6	1.1	0.4	0.0	0.0	1.1	3.5
30	0.0	0.4	3.3	3.5	44.4	58.0	5.6	8.6	45.6	29.6	1.1	0.0
31	0.0	0.0	1.1	0.4	0.0	0.4	0.0	0.0	98.9	99.2	0.0	0.0
32	0.0	0.0	2.2	0.8	2.2	5.8	1.1	0.8	2.2	3.9	92.2	88.7
33	0.0	0.0	0.0	0.0	1.1	0.4	0.0	0.0	96.7	97.3	2.2	2.3
34	5.6	4.3	7.8	3.9	75.6	80.2	3.3	1.6	6.7	9.7	1.1	0.4
35	55.6	56.4	22.2	22.2	1.1	3.1	14.4	13.6	5.6	4.3	1.1	0.4
36	45.6	37.0	6.7	5.8	41.1	46.3	2.2	1.2	3.3	8.9	1.1	0.8
37	24.4	30.7	16.7	14.0	22.2	14.8	28.9	31.5	6.7	8.2	1.1	0.8
38	17.8	17.5	13.3	7.4	47.8	57.2	14.4	17.1	4.4	0.8	2.2	0.0
39	6.7	5.8	11.1	7.4	35.6	27.2	33.3	35.0	10.0	21.0	3.3	3.5
40	7.8	10.9	1.1	0.0	2.2	1.6	80.0	75.1	8.9	11.7	0.0	0.8
41	24.4	15.2	0.0	0.0	0.0	1.2	24.4	11.3	51.1	72.4	0.0	0.0
42	43.3	42.4	1.1	1.6	2.2	2.7	35.6	23.3	16.7	29.6	1.1	0.4

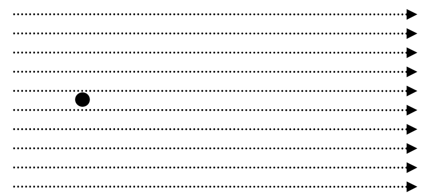
Item	Frequency Table											
	A		*B*		*C*		*D*		*E*		*Omit*	
	f	*m*	*f*	*m*	*f*	*m*	*f*	*m*	*f*	*m*	*f*	*m*
43	23.3	6.2	1.1	0.0	2.2	2.3	17.8	8.9	55.6	82.1	0.0	0.4
44	68.9	79.8	6.7	8.2	1.1	0.8	1.1	0.4	22.2	10.5	0.0	0.4
45	82.2	88.3	15.6	9.3	1.1	0.8	0.0	0.8	1.1	0.4	0.0	0.4
46	33.3	49.0	10.0	8.9	50.0	33.1	2.2	5.4	3.3	2.3	1.1	1.2
47	2.2	0.4	1.1	1.6	5.6	0.8	0.0	1.9	91.1	94.9	0.0	0.4
48	12.2	9.7	67.8	67.7	3.3	4.7	2.2	1.6	14.4	16.0	0.0	0.4
49	21.1	23.3	1.1	1.6	66.7	65.4	1.1	3.9	10.0	5.1	0.0	0.8
50	16.7	14.8	8.9	7.8	30.0	43.6	38.9	27.2	4.4	5.4	1.1	1.2
51	1.1	1.2	1.1	0.4	26.7	20.6	66.7	75.1	3.3	1.9	1.1	0.8
52	10.0	5.8	7.8	7.0	8.9	2.7	5.6	1.6	67.8	82.1	0.0	0.8
53	14.4	10.5	58.9	79.4	12.2	3.5	6.7	1.9	6.7	3.9	1.1	0.8
54	26.7	16.3	8.9	5.4	57.8	70.4	2.2	3.9	4.4	2.7	0.0	1.2
55	4.4	2.7	54.4	73.9	2.2	1.6	2.2	1.6	35.6	19.5	1.1	0.8
56	12.2	8.9	63.3	81.7	14.4	4.7	3.3	0.8	5.6	3.1	1.1	0.8
57	33.3	39.7	28.9	31.5	28.9	20.6	4.4	5.4	3.3	1.6	1.1	1.2
58	80.0	93.0	7.8	1.9	3.3	1.2	1.1	0.4	7.8	2.3	0.0	1.2
59	65.6	62.6	5.6	3.5	22.2	25.7	3.3	7.0	1.1	0.0	2.2	1.2
60	14.4	10.9	2.2	7.8	35.6	35.0	6.7	5.8	40.0	38.1	1.1	2.3
61	20.0	26.5	7.8	5.8	40.0	47.9	26.7	17.1	4.4	1.2	1.1	1.6
62	8.9	3.9	4.4	3.1	45.6	42.4	35.6	43.6	5.6	3.5	0.0	3.5
63	10.0	5.4	37.8	36.2	22.2	25.7	21.1	22.6	8.9	7.0	0.0	3.1
64	16.7	16.0	10.0	3.9	7.8	7.4	12.2	4.7	52.2	64.2	1.1	3.9
65	5.6	2.7	6.7	2.3	14.4	21.4	25.6	15.2	46.7	54.5	1.1	3.9
66	6.7	8.2	15.6	11.3	24.4	23.7	24.4	34.2	26.7	18.3	2.2	4.3
67	3.3	5.4	16.7	14.4	32.2	26.8	14.4	29.2	30.0	20.2	3.3	3.9
68	12.2	7.4	21.1	8.2	17.8	20.2	7.8	21.4	38.9	37.7	2.2	5.1
69	30.0	39.3	46.7	38.5	7.8	7.4	7.8	6.6	4.4	3.9	3.3	4.3
70	41.1	36.2	31.1	41.6	8.9	4.3	6.7	6.2	10.0	7.0	2.2	4.7

APPENDIX C

This is Version 3.0 of the DEEM. This appendix presents a modified version from what was administered to the students; it has been formatted to fit the specifications of the thesis.

For questions 1-4 the group of dashed arrows represents electric field lines.
What is the direction, if any, of the force on the particle due to the field if...

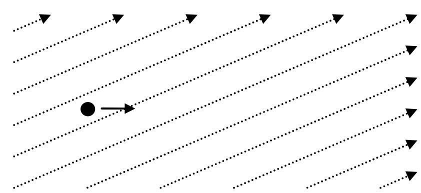
the particle is initially at rest and



1) the particle has a net negative charge?
   a) ↑ b) ↓ c) → d) ← e) There is no net force.

2) the particle has a net positive charge?
   a) ↑ b) ↓ c) → d) ← e) There is no net force.

the particle is initially moving to the right and

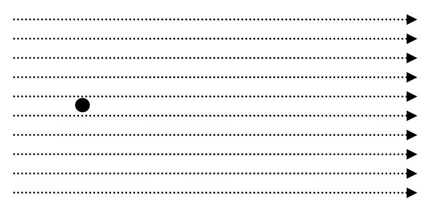


3) the particle has a net negative charge?
   a) ↗ b) ↙ c) → d) ← e) None of these.

4) the particle has a net positive charge?
   a) ↗ b) ↙ c) → d) ← e) None of these.

For questions 5-9 the group of dashed arrows represents magnetic field lines.
What is the direction, if any, of the force on the particle due to the field if...

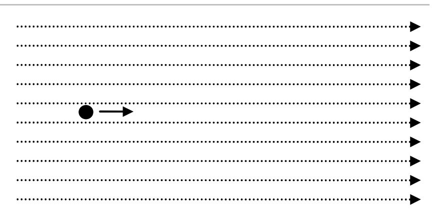
the particle is initially at rest and



5) the particle has a net negative charge?
   a) → b) ← c) Into the page. d) Out of the page.
   e) There is no net force.

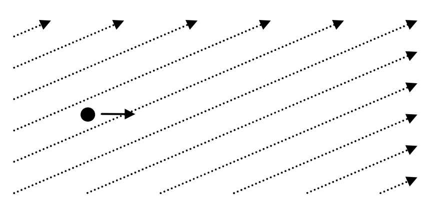
6) the particle has a net positive charge?
   a) → b) ← c) Into the page. d) Out of the page.
   e) There is no net force.

the particle is initially moving to the right and



7) the particle has a net negative charge?
   a) → b) ← c) Into the page. d) Out of the page.
   e) There is no net force.

the particle is initially moving to the right and



8) the particle has a net negative charge?
   a) ↗ b) ↙ c) Into the page. d) Out of the page.
   e) There is no net force.

9) the particle has a net positive charge?
   a) ↗ b) ↙ c) Into the page. d) Out of the page.
   e) There is no net force.

For questions 10-33 the dots, • , represent points in space (they are not particles).

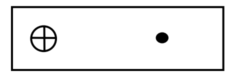
The circles represent particles with a net negative charge, ⊖, or a net positive charge, ⊕. The magnitudes of the net charges are all equal.

The thick, black lines, ——, represent side-views of long, conducting wires. An "**I**" with a solid arrow next to it (e.g. **I**→) indicates the direction of current along the wire in the plane of the page. No "**I**" in the figure indicates that there is no current along the wire. The magnitudes of the currents are all equal.

Any and all fields which may be present in each figure are due only to charges or currents in the figure.

Assume that the value of the electrostatic potential equals zero an infinite distance away from a point charge.

10) For the figure in the box to the right, the net electric field at the dot points...



a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

11) **If you chose answer "*e*" in question 10 then leave line #11 on your answer sheet blank, skip this question and go directly to question 12**. Relative to the magnitude of electric field at the dot in the figure for question 10, what would happen to the magnitude of the electric field at the dot, if the dot is moved to a new position farther to the right?

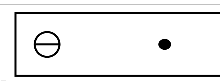
a) The magnitude of the electric field at the dot would be greater.
b) The magnitude of the electric field at the dot would be less.
c) The magnitude of the electric field at the dot would not change.
d) There is not enough information.
e) None of these.

12) What would be the direction of the electric field, if any, at the dot in question 10 if the dot is moved to the right a large, finite distance away?

a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

---

13) For the figure in the box to the right, the net electric field at the dot points...



a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

14) **If you chose answer "*e*" in question 13 then leave line #14 on your answer sheet blank, skip this question and go directly to question 15**. Relative to the magnitude of electric field at the dot in the figure for question 13, what would happen to the magnitude of the electric field at the dot, if the dot is moved to a new position farther to the right?

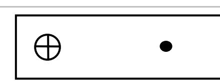
a) The magnitude of the electric field at the dot would be greater.
b) The magnitude of the electric field at the dot would be less.
c) The magnitude of the electric field at the dot would not change.
d) There is not enough information.
e) None of these.

15) What would be the direction of the electric field, if any, at the dot in question 13 if the dot is moved to the right a large, finite distance away?

a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

---

16) What is the value of the electrostatic potential at the black dot in the figure to the right?



a) Not enough information. b) The electrostatic potential is not defined at the dot.
c) 0. d) Greater than 0. e) Less than 0.

17) **If you chose answer "*a*" or "*b*" in question 20 then leave lines #21 and #22 on your answer sheet blank, skip questions 21 and 22 and go directly to question 23**. Relative to the value of electrostatic potential at the dot in the figure for question 16, what would happen to the value of the electrostatic potential at the dot, if the dot is moved to a new position farther to the right?

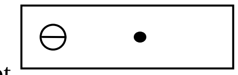
a) The value of the electrostatic potential at the dot would be greater.
b) The value of the electrostatic potential at the dot would be less.
c) The value of the electrostatic potential at the dot would not change.
d) There is not enough information.
e) None of these.

18) What would be the value of the electrostatic potential at the dot in question 16 if the dot was moved to the right a large, finite distance away?
a) It would be 0. b) It would be negative and nearly 0. c) It would be positive and nearly 0.
d) It would be negative and very large e) It would be positive and very large.

19) What is the direction of the electrostatic potential at the dot in question 16?
a) → b) ←
c) Since the electrostatic potential at the dot is zero, it can not have a direction.
d) Electrostatic potential does not have a direction.
e) The electrostatic potential is undefined at the dot.

---

20) What is the value of the electrostatic potential at the black dot in the figure to the right?



a) Not enough information. b) The electrostatic potential is not defined at the dot.
c) 0. d) Greater than 0. e) Less than 0.

21) **If you chose answer "*a*" or "*b*" in question 20 then leave lines #21 and #22 on your answer sheet blank, skip questions 21 and 22 and go directly to question 23**. Relative to the value of electrostatic potential at the dot in the figure for question 16, what would happen to the value of the electrostatic potential at the dot, if the dot is moved to a new position farther to the right?

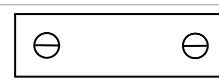
a) The value of the electrostatic potential at the dot would be greater.
b) The value of the electrostatic potential at the dot would be less.
c) The value of the electrostatic potential at the dot would not change.
d) There is not enough information.
e) None of these.

22) What would be the value of the electrostatic potential at the dot in question 20 if the dot was moved to the right a large, finite distance away?
a) It would be 0. b) It would be negative and nearly 0. c) It would be positive and nearly 0.
d) It would be negative and very large e) It would be positive and very large.

23) What is the direction of the electrostatic potential at the dot in question 20?
a) → b) ←
c) Since the electrostatic potential at the dot is zero, it can not have a direction.
d) Electrostatic potential does not have a direction.
e) The electrostatic potential is undefined at the dot.

---

24) What can be said about the value of the electrostatic potential energy for the system of charged particles in the figure to the right?



a) There is not enough information to say anything about the value of the electrostatic potential energy.
b) The electrostatic potential energy is only defined at a specific location in space.
c) Less than 0. d) 0. e) Greater than 0.

25) **If you chose answer "*a*" or "*b*" in question 24 then leave line #25 on your answer sheet blank, skip this question and go directly to question 26**. Relative to the electrostatic potential energy of the system in figure 24, what would happen to the value of the electrostatic potential energy if the two charged particles are moved to a new position farther apart?

a) The value of the electrostatic potential energy would be greater.
b) The value of the electrostatic potential energy would be less.
c) The value of the electrostatic potential energy would not change.
d) There is not enough information.
e) None of these.

---

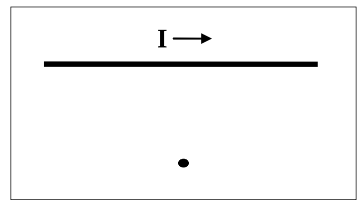
26) What can be said about the value of the electrostatic potential energy for the system of charged particles in the figure to the right?

a) There is not enough information to say anything about the value of the electrostatic potential energy.
b) The electrostatic potential energy is only defined at a specific location in space.
c) Less than 0. d) 0. e) Greater than 0.

27) **If you chose answer "*a*" or "*b*" in question 26 then leave line #27 on your answer sheet blank, skip this question and go directly to question 28**. Relative to the electrostatic potential energy of the system in figure 26, what would happen to the value of the electrostatic potential energy if the two charged particles are moved to a new position farther apart?

a) The value of the electrostatic potential energy would be greater.
b) The value of the electrostatic potential energy would be less.
c) The value of the electrostatic potential energy would not change.
d) There is not enough information.
e) None of these.

28) For the figure in the box to the right, the net magnetic field at the dot points...



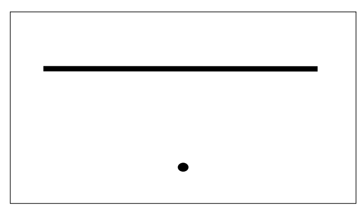
a) → b) ↑ c) Into the page. d) Out of the page.
e) There is no net magnetic field.

29) **If you chose answer "*e*" in question 28 then leave line #29 on your answer sheet blank, skip this question and go directly to question 30**. Relative to the magnitude of magnetic field at the dot in the figure for question 28, what would happen to the magnitude of the magnetic field at the dot, if the dot is moved to a new position farther down (in the plane of the page)?

a) The magnitude of the magnetic field at the dot would be greater.
b) The magnitude of the magnetic field at the dot would be less.
c) The magnitude of the magnetic field at the dot would not change.
d) There is not enough information.
e) None of these.

30) What would be the direction of the magnetic field, if any, at the dot in question 28 if the dot was moved farther down (in the plane of the page) a large, finite distance away?
a) → b) ← c) ↑ d) ↓ e) There would be no net magnetic field.

31) For the figure in the box to the right, the net magnetic field at the dot points...



a) → b) ↑ c) Into the page. d) Out of the page.
e) There is no net magnetic field.

32) If you chose answer "*e*" in question 31 then leave line #32 on your answer sheet blank, skip this question and go directly to question 33. Relative to the magnitude of magnetic field at the dot in the figure for question 31, what would happen to the magnitude of the magnetic field at the dot, if the dot is moved to a new position farther down (in the plane of the page)?

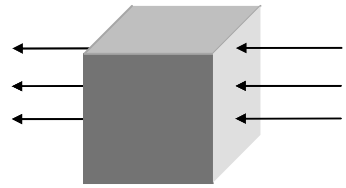
a) The magnitude of the magnetic field at the dot would be greater.
b) The magnitude of the magnetic field at the dot would be less.
c) The magnitude of the magnetic field at the dot would not change.
d) There is not enough information.
e) None of these.

33) What would be the direction of the magnetic field, if any, at the dot in question 31 if the dot was moved farther down (in the plane of the page) a large, finite distance away?
a) → b) ← c) ↑ d) ↓ e) There would be no net magnetic field.

For questions 34-37 the hollow cube in the figure is **imaginary**; it is only there to conceal a region of space which may or may not contain something that could produce electric or magnetic fields. There also may be other objects outside the cube which are not in the figure that could be producing the electric or magnetic fields in the figures.

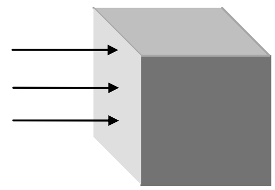
For the next two figures the arrows represent electric field lines.

34) In the figure to the right, what can be said about what the imaginary, hollow cube is concealing?



a) There is a net negative charge in the cube.
b) There is a net positive charge in the cube.
c) There is a no net charge in the cube.
d) This is an impossible situation.
e) Nothing can be said.

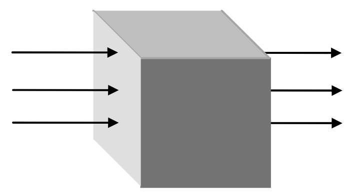
35) In the figure to the right, what can be said about what the imaginary, hollow cube is concealing?



a) There is a net negative charge in the cube.
b) There is a net positive charge in the cube.
c) There is a no net charge in the cube.
d) This is an impossible situation.
e) Nothing can be said.

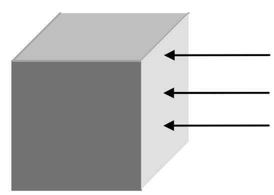
For the next two figures the arrows represent magnetic field lines

36) In the figure to the right, what can be said about what the imaginary, hollow cube is concealing?



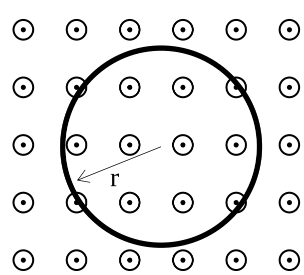
a) There is a net current looping up along the back and down along the front of the cube.
b) There is a net current looping down along the back and up along the front of the cube.
c) There is a no net current in the cube.
d) This is an impossible situation.
e) Nothing can be said.

37) In the figure to the right, what can be said about what the imaginary, hollow cube is concealing?



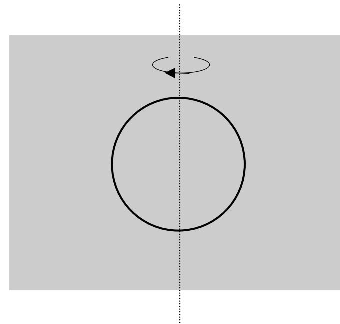
a) There is a net current looping up along the back and down along the front of the cube.
b) There is a net current looping down along the back and up along the front of the cube.
c) There is a no net current in the cube.
d) This is an impossible situation.
e) Nothing can be said

38) A thin, circular, conducting loop, in the plane of the page, is immersed in a uniform magnetic field pointed perpendicularly out of the page (see figure to the right). As the radius, *r*, of the loop decreases, which answer below describes something that happens to the loop (other than it simply gets smaller)?



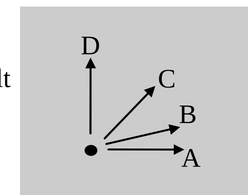
a) Nothing.
b) An alternating current is established in the loop.
c) A direct, counterclockwise current is established in the loop in the plane of the page.
d) There is a net torque exerted on the loop causing it to rotate counterclockwise in the plane of the page.
e) There is a net force exerted on the loop to the right.

39) A thin, circular, conducting loop is rotating in a uniform magnetic field pointing perpendicularly out of the page. The dotted line in the figure represents the imaginary axis that the loop is rotating about and the arrow indicates the direction the loop is rotating. Over a period of time that spans several revolutions what happens to the loop? At the instant of time in the figure, the loop is in the plane of the page.



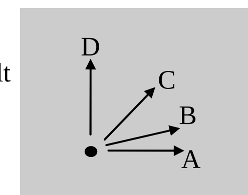
a) As long as the rotation speed is constant, the loop feels a net force to the left.
b) Nothing
c) A direct current is induced in the wire. At the instant of time in the figure, the current is counterclockwise
d) An alternating current is induced in the wire. At the instant of time in the figure, the magnitude of the current is at a maximum and is counterclockwise
e) An alternating current is induced in the wire. At the instant of time in the figure, the current is zero.

40) The figure to the right represents a region of a uniform **magnetic field pointed to the right**. If the black dot is a positively charged particle, what arrow would represent the particle's velocity, at this instant, which would result in the maximum force on the particle?



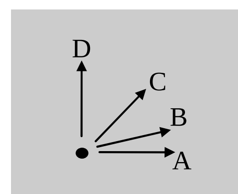
a) A b) B c) C d) D e) All the same

41) The figure to the right represents a region of a uniform **magnetic field pointed into the page**. If the black dot is a positively charged particle, what arrow would represent the particle's velocity, at this instant, which would result in the maximum force on the particle?



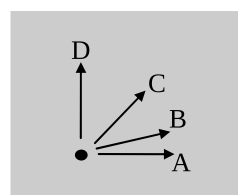
a) A b) B c) C d) D e) All the same

42) The figure to the right represents a region of a uniform **electric field pointed to the right**. If the black dot is a positively charged particle, what arrow would represent the particle's velocity, at this instant, which would result in the maximum force on the particle?



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

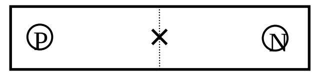
43) The figure to the right represents a region of a uniform **electric field pointed into the page**. If the black dot is a positively charged particle, what arrow would represent the particle's velocity, at this instant, which would result in the maximum force on the particle?



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

For questions 44-61 the crosses, ✕, represent a location in space.
The circles containing an "N" (Ⓝ) represent particles with a net negative charge. The circles containing an "P" (Ⓟ) represent particles with a net positive charge. The magnitudes of the net charges are all equal.
The filled-in circles, ●, represent cross-sectional views of long, conducting wires. An "**I**" with a the word "**in**" or "**out**" next to it indicates the direction of current along the wire into or out of the plane of the page respectively. The magnitudes of the currents are all equal.
Any and all fields which may be present in each figure are due only to charges or currents in the figure.
Assume the value of the electrostatic potential equals zero an infinite distance away from a point charge.
For questions 44-49 the vertical, dashed line in the figures represents the midpoint between the two charges.

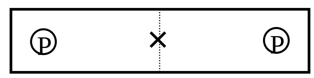
44) For the figure in the box to the right, the net electric field at the cross points...
a) → b) ← c) ↑ d) ↓ e) There is no net electric field.



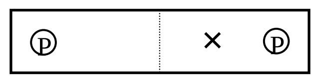
45) For the figure in the box to the right, the net electric field at the cross points...
a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

46) What is the relationship, if any, of the magnitudes of the electric fields in the above two questions?
a) They are both equal. b) The first is larger. c) The second is larger. d) Not enough information. e) There is no relationship.

47) For the figure in the box to the right, the net electric field at the cross points...
a) → b) ← c) ↑ d) ↓ e) There is no net electric field.



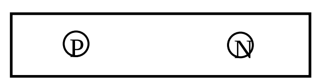
48) For the figure in the box to the right, the net electric field at the cross points...
a) → b) ← c) ↑ d) ↓ e) There is no net electric field.



49) What is the relationship, if any, of the magnitudes of the electric fields in the above two questions?
a) They are both equal. b) The first is larger. c) The second is larger. d) Not enough information. e) There is no relationship.

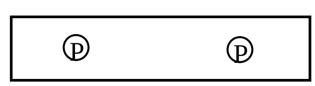
For questions consider only the space contained within the finite area of the figure.

50) For the figure in the box to the right, where is the electrostatic potential zero?



a) Somewhere to the left of the two particles.
b) Somewhere to the right of the two particles.
c) Nowhere in the figure.
d) Exactly halfway between the two particles.
e) More information is needed.

51) For the figure in the box to the right, where is the electrostatic potential zero?

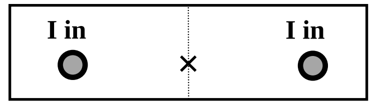


a) Somewhere to the left of the two particles.
b) Somewhere to the right of the two particles.
c) Nowhere in the figure.
d) Exactly halfway between the two particles.
e) More information is needed.

For questions 52-57 the vertical, dashed line in the figures represents the midpoint between the two wires.

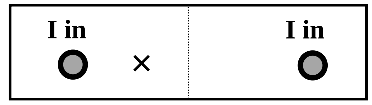
52) For the figure in the box to the right, the net magnetic field at the cross points...

a) ← b) ↓ c) Into the page.
d) Out of the page. e) There is no net magnetic field.



53) For the figure in the box to the right, the net magnetic field at the cross points...

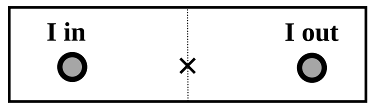
a) ← b) ↓ c) Into the page.
d) Out of the page. e) There is no net magnetic field.



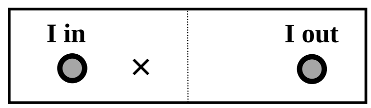
54) What is the relationship, if any, of the magnitudes of the magnetic fields in the above two questions?

a) They are both equal b) The first is larger c) The second is larger
d) There is not enough information e) There is no relationship.

55) For the figure in the box to the right, the net magnetic field at the cross points...
a) ← b) ↓ c) Into the page.
d) Out of the page. e) There is no net magnetic field.



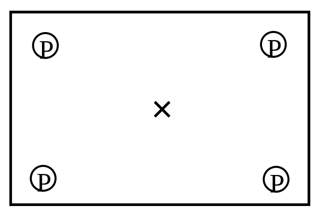
56) For the figure in the box to the right, the net magnetic field at the cross points...
a) ← b) ↓ c) Into the page.
d) Out of the page. e) There is no net magnetic field.



57) What is the relationship, if any, of the magnitudes of the magnetic fields in the above two questions?
a) They are both equal. b) The first is larger. c) The second is larger.
d) There is not enough information. e) There is no relationship.

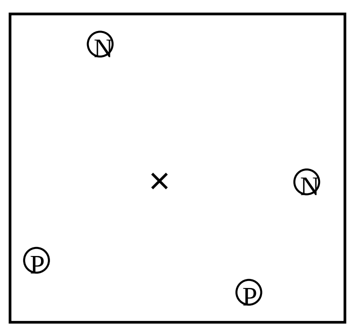
For questions 58-59 assume the charged particles are at the corners of an imaginary rectangle and are equidistant from the cross, however, the exact distance is not known.

58) For the figure in box to the right, the net electric field at the cross points...
a) There is no net electric field. b) Into the page.
c) ↑ d) ↓ e) Not enough information.



59) What is the value of the electrostatic potential at the cross in the figure to the right?
a) 0. b) Less than 0. c) Greater than 0. d) Not enough information.

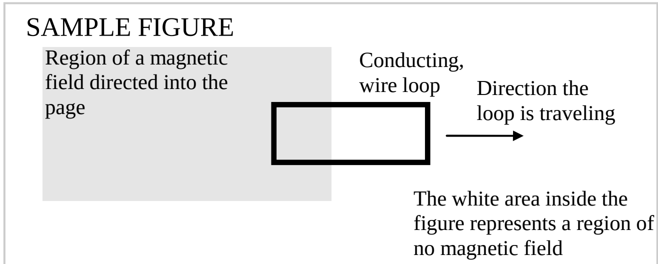
For questions 60-61 assume the distances between the cross and the charges are all the same, however, the exact positions of the charges are not known.



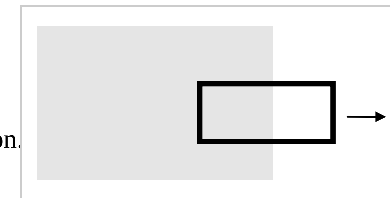
60) For the figure in box to the right, the net electric field at the cross points...
a) There is no net electric field. b) Into the page.
c) ↑ d) ↓ e) Not enough information.

61) What is the value of the electrostatic potential at the cross in the figure to the right?
a) 0. b) Less than 0. c) Greater than 0. d) Not enough information.

For questions 62-65 the thick, black rectangles represent conducting, wire loops. Assume the wire to be of negligible thickness. The gray areas represent regions of magnetic field directed perpendicularly into the page. The black arrows represent the direction of motion of the wire loops. No arrow indicates that the loop is not moving. Look at the sample figure below for an example.

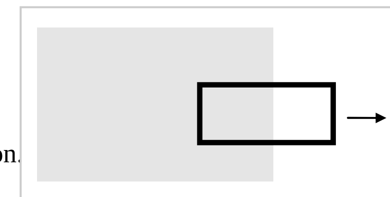


62) Choose an answer that describes something that happens to the wire loop as it moves as shown in the figure to the right? The magnetic field is **uniform**.



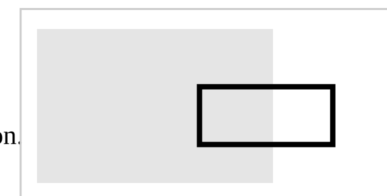
a) Positive charge builds up on the left side of the loop.
b) There is not enough information given to answer the question.
c) A counterclockwise current is induced.
d) A clockwise current is induced.
e) None of these.

63) Choose an answer that describes something that happens to the wire loop as it moves as shown in the figure to the right? The magnetic field is **non-uniform**.



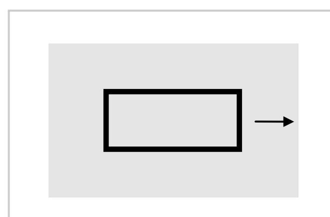
a) Positive charge builds up on the left side of the loop.
b) There is not enough information given to answer the question.
c) A counterclockwise current is induced.
d) A clockwise current is induced.
e) None of these.

64) Choose an answer that describes something that happens to the wire loop as it moves as shown in the figure to the right? The magnetic field is **uniform**.



a) Positive charge builds up on the left side of the loop.
b) There is not enough information given to answer the question.
c) A counterclockwise current is induced.
d) A clockwise current is induced.
e) None of these.

65) Choose an answer that describes something that happens to the wire loop as it moves as shown in the figure to the right? The magnetic field is **uniform**.



a) Positive charge builds up on the left side of the loop.
b) There is not enough information given to answer the question.
c) A counterclockwise current is induced.
d) A clockwise current is induced.
e) None of these.

66) The gray area represents a region of a uniform magnetic field directed perpendicularly into the page. The magnetic field is decreasing as a function of time. What can be said about the region outside the magnetic field?

a) There is a net charge building up above and below the field in a plane perpendicular to the page.
b) There is a net charge building up around the field in the plane of the page.
c) In the plane of the page there is an counterclockwise electric field.
d) In the plane of the page there is a clockwise electric field.
e) Nothing.

67) The gray area represents a region of a uniform electric field directed perpendicularly into the page. The electric field is decreasing as a function of time. What can be said about the region outside the electric field?

a) There is a net current building up above and below the field in a plane perpendicular to the page.
b) There is a net current building up around the field in the plane of the page.
c) In the plane of the page there is an counterclockwise magnetic field.
d) In the plane of the page there is a clockwise magnetic field.
e) Nothing.

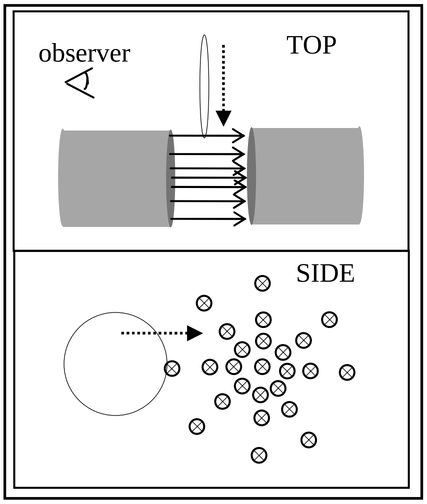
68) The gray area represents a region of a non-uniform electric field directed perpendicularly into the page. The electric field is decreasing as a function of time. What can be said about the region outside the electric field?

a) There is a net current building up above and below the field in a plane perpendicular to the page.
b) There is a net current building up around the field in the plane of the page.
c) In the plane of the page there is an counterclockwise magnetic field.
d) In the plane of the page there is a clockwise magnetic field.
e) Nothing.

69) An observer watches a thin, circular, conducting, wire loop enter a region of a uniform magnetic field (see Figure A- TOP VIEW). The thick, dotted, black arrow in both views indicates the direction the loop is moving. Looking at the loop **as it enters** the magnetic field (see Figure A- SIDE VIEW), which answer below describes something, if anything, that happens to the loop as viewed by the observer?

a) A counterclockwise current is created in the wire loop.
b) A clockwise current is created in the wire loop.
c) Nothing could happen to the wire loop as it enters any magnetic field.
d) Nothing happens because the field is uniform.
e) No answer can be given based on the information given.

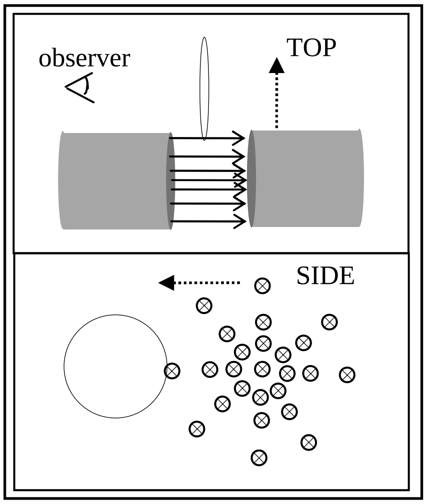
Figure A



70) An observer watches a region of a uniform magnetic field pass across a thin, circular, conducting, wire loop (see Figure B- TOP VIEW). The thick, dotted, black arrow in both views indicates the direction the magnetic field is moving. Looking at the loop, **as the uniform magnetic field begins to passes across it** (see Figure B- SIDE VIEW), which answer below describes something, if anything, that happens to the loop as viewed by the observer?

a) A counterclockwise current is created in the wire loop.
b) A clockwise current is created in the wire loop.
c) Nothing could happen to the wire loop as it enters any magnetic field.
d) Nothing happens because the field is uniform.
e) No answer can be given based on the information given.

Figure B

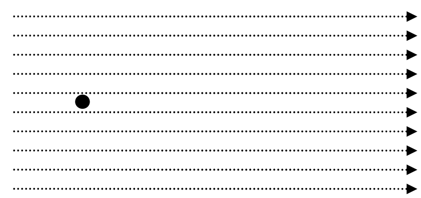


APPENDIX D

This appendix contains the final version of the exam as outlined in chapter 13.

For questions 1-4 the group of dotted arrows represents electric field lines.
What is the direction, if any, of the force on the particle due to the field if...

the particle is initially at rest and

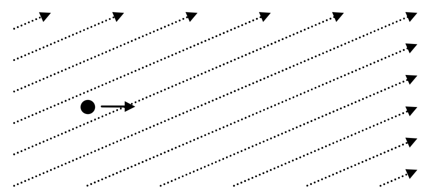


1) the particle has a net negative charge?
   a) ↑ b) ↓ c) → d) ← e) There is no net force.

2) the particle has a net positive charge?
   a) ↑ b) ↓ c) → d) ← e) There is no net force.

---

the particle is initially moving to the right and

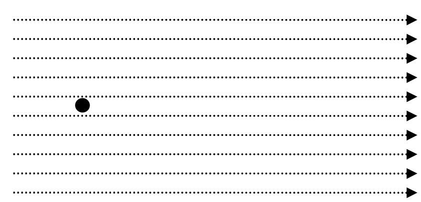


3) the particle has a net negative charge?
   a) ↗ b) ↙ c) → d) ← e) Into the page.

4) the particle has a net positive charge?
   a) ↗ b) ↙ c) → d) ← e) Out of the page.

For questions 5-9 the group of dotted arrows represents magnetic field lines.
What is the direction, if any, of the force on the particle due to the field if...

the particle is initially at rest and

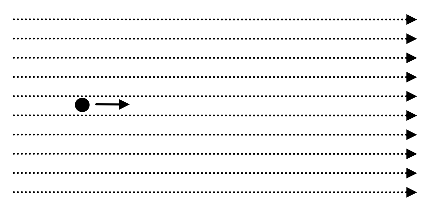


5) the particle has a net negative charge?
   a) → b) ← c) Into the page. d) Out of the page.
   e) There is no net force.

6) the particle has a net positive charge?
   a) → b) ← c) Into the page. d) Out of the page.
   e) There is no net force.

---

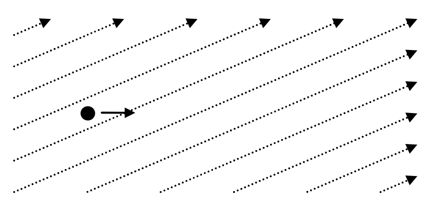
the particle is initially moving to the right and



7) the particle has a net negative charge?
   a) → b) ← c) Into the page. d) Out of the page.
   e) There is no net force.

---

the particle is initially moving to the right and



8) the particle has a net negative charge?
   a) → b) ↗ c) ↙
   d) Into the page. e) Out of the page.

9) the particle has a net positive charge?
   a) → b) ↗ c) ↙
   d) Into the page. e) Out of the page.

For questions 10-30 the crosses, ×, represent a location in space.
The circles represent particles with a net negative charge , ⊖ , or a net positive charge, ⊕. The magnitudes of the net charges are all equal.
The thick, black lines, ▬▬▬▬, represent side-views of long, straight, conducting wires. An "**I**" with a solid arrow next to it (e.g. **I**→) indicates the direction of current along the wire in the plane of the page. The magnitudes of the currents are all equal.
Any and all fields which may be present in each figure are due only to charges or currents in the figure.
Assume that the value of the electrostatic potential equals zero an infinite distance away from a point charge.

10) For the figure in the box to the right, the net electric field at the cross points...

a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

11) **If you chose answer "*e*" in question 10 then leave line #11 on your answer sheet blank, skip this question and go directly to question 12**.
Relative to the magnitude of electric field at the cross in the figure for question 10, what would happen to the magnitude of the electric field at the cross if the cross were farther to the right?

a) The magnitude of the electric field at the cross would be *greater*.
b) The magnitude of the electric field at the cross would be *less*.
c) The magnitude of the electric field at the cross would *not change*.
d) There is not enough information.

12) What would be the direction of the electric field, if any, at the cross in question 10 if the cross were farther to the right, a large, finite distance?
a) → b) ← c) ↑ d) ↓ e) There would be no net electric field.

13) For the figure in the box to the right, the net electric field at the cross points...

a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

14) **If you chose answer "*e*" in question 13 then leave line #14 on your answer sheet blank, skip this question and go directly to question 15**.
Relative to the magnitude of electric field at the cross in the figure for question 13, what would happen to the magnitude of the electric field at the cross if the cross were farther to the right?

a) The magnitude of the electric field at the cross would be *greater*.
b) The magnitude of the electric field at the cross would be *less*.
c) The magnitude of the electric field at the cross would *not change*.
d) There is not enough information.

15) What would be the direction of the electric field, if any, at the cross in question 10 if the cross were farther to the right, a large, finite distance?
a) → b) ← c) ↑ d) ↓ e) There would be no net electric field.

Items 16-23 ask questions about the values of a numbers, *not* their magnitudes. For example, the value of 1 is greater than the value of -5.

16) What is the value of the electrostatic potential at the cross in the figure to the right?

a) The *exact* value can *not* be determined, but it must be *less* than 0.
b) The value is zero.
c) The *exact* value can *not* be determined, but it must be *greater* than 0.
d) The electrostatic potential is not defined at the cross.
e) There is not enough information to determine anything about the value.

17) **If you chose answer "*d*" or "*e*" in question 16 then leave lines #17 and #18 on your answer sheet blank, skip questions 17 and 18 and go directly to question 19**. Relative to the value of electrostatic potential at the cross in the figure for question 16, what would happen to the value of the electrostatic potential at the cross, if the cross were farther to the right?

a) The value of the electrostatic potential at the cross would be *greater*.
b) The value of the electrostatic potential at the cross would be *less*.
c) The value of the electrostatic potential at the cross would *not change*.
d) There is not enough information.

18) What would be the value of the electrostatic potential at the cross in question 16 if the cross were farther to the right, a large, finite distance?

a) It would be 0. b) It would be negative and nearly 0. c) It would be positive and nearly 0.
d) It would be negative and very large e) It would be positive and very large.

19) What is the direction of the electrostatic potential at the cross in question 16?

a) → b) ←
c) Since the electrostatic potential at the cross is zero, it can not have a direction.
d) Electrostatic potential does not have a direction.
e) The electrostatic potential is undefined at the cross.

---

20) What is the value of the electrostatic potential at the cross in the figure to the right?

a) The *exact* value can *not* be determined, but it must be *less* than 0.
b) The value is zero.
c) The *exact* value can *not* be determined, but it must be *greater* than 0.
d) The electrostatic potential is not defined at the cross.
e) There is not enough information to determine anything about the value.

21) **If you chose answer "*d*" or "*e*" in question 20 then leave lines #21 and #22 on your answer sheet blank, skip questions 21 and 22 and go directly to question 23**. Relative to the value of electrostatic potential at the cross in the figure for question 20, what would happen to the value of the electrostatic potential at the cross, if the cross were farther to the right?

a) The value of the electrostatic potential at the cross would be *greater*.
b) The value of the electrostatic potential at the cross would be *less*.
c) The value of the electrostatic potential at the cross would *not change*.
d) There is not enough information.

22) What would be the value of the electrostatic potential at the cross in question 20 if the cross were farther to the right, a large, finite distance?

a) It would be 0. b) It would be negative and nearly 0. c) It would be positive and nearly 0.
d) It would be negative and very large e) It would be positive and very large.

23) What is the direction of the electrostatic potential at the cross in question 20?

a) → b) ←
c) Since the electrostatic potential at the cross is zero, it can not have a direction.
d) Electrostatic potential does not have a direction.
e) The electrostatic potential is undefined at the cross.

---

24) What can be said about the value of the electrostatic potential energy for the system of charged particles in the figure to the right?

a) The *exact* value can *not* be determined, but it must be *less* than 0.
b) The value is zero.
c) The *exact* value can *not* be determined, but it must be *greater* than 0.
d) The electrostatic potential energy is only defined at a specific location in space.
e) There is not enough information to determine anything about the value.

25) **If you chose answer "*a*" or "*b*" in question 24 then leave line #25 on your answer sheet blank, skip this question and go directly to question 26**. Relative to the electrostatic potential energy of the system in figure 24, what would happen to the value of the electrostatic potential energy if the two charged particles are moved to a new position farther apart?

a) The value of the electrostatic potential energy would be *greater*.
b) The value of the electrostatic potential energy would be *less*.
c) The value of the electrostatic potential energy would *not change*.
d) There is not enough information.

---

26) What can be said about the value of the electrostatic potential energy for the system of charged particles in the figure to the right?

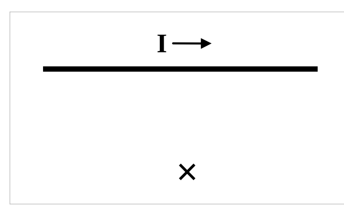
a) The *exact* value can *not* be determined, but it must be *less* than 0.
b) The value is zero.
c) The *exact* value can *not* be determined, but it must be *greater* than 0.
d) The electrostatic potential energy is only defined at a specific location in space.
e) There is not enough information to determine anything about the value.

27) I**f you chose answer "*a*" or "*b*" in question 26 then leave line #27 on your answer sheet blank, skip this question and go directly to question 28**. Relative to the electrostatic potential energy of the system in figure 26, what would happen to the value of the electrostatic potential energy if the two charged particles are moved to a new position farther apart?

a) The value of the electrostatic potential energy would be *greater*.
b) The value of the electrostatic potential energy would be *less*.
c) The value of the electrostatic potential energy would *not change*.
d) There is not enough information.

---

28) For the figure in the box to the right, the net magnetic field at the cross points...



a) → b) ↑ c) Into the page. d) Out of the page.
e) There is no net magnetic field.

29) **If you chose answer "*e*" in question 28 then leave line #29 on your answer sheet blank, skip this question and go directly to question 30**. Relative to the magnitude of magnetic field at the cross in the figure for question 28, what would happen to the magnitude of the magnetic field at the cross, if the cross were down (in the plane of the page)?

a) The magnitude of the magnetic field at the cross would be *greater*.
b) The magnitude of the magnetic field at the cross would be *less*.
c) The magnitude of the magnetic field at the cross would *not change*.
d) There is not enough information.

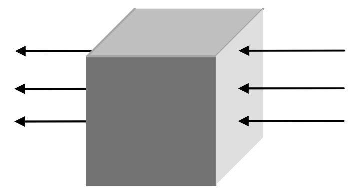
30) What would be the direction of the magnetic field, if any, at the cross in question 28 if the cross were farther down (in the plane of the page) a large, finite distance away?

a) → b) ↑ c) Into the page. d) Out of the page.
e) There is no net magnetic field.

For questions 31-33 the hollow cube in the figure is **imaginary**; it is only there to conceal a region of space which may or may not contain something that could produce electric or magnetic fields. There also may be other objects outside the cube, which are not in the figure, which could be producing the electric or magnetic fields in the figures.

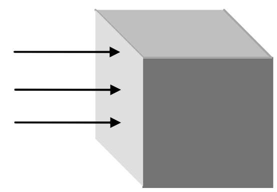
For the next two figures the arrows represent electric field lines.

31) In the figure to the right, what can be said about what the imaginary, hollow cube is concealing?



a) There is a net negative charge in the cube.
b) There is a net positive charge in the cube.
c) There is a no net charge in the cube.
d) This is an impossible situation.
e) Nothing can be said.

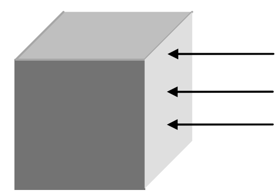
32) In the figure to the right, what can be said about what the imaginary, hollow cube is concealing?



a) There is a net negative charge in the cube.
b) There is a net positive charge in the cube.
c) There is a no net charge in the cube.
d) This is an impossible situation.
e) Nothing can be said.

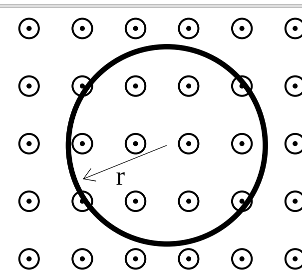
For the next figure the arrows represent magnetic field lines

33) In the figure to the right, what can be said about what the imaginary, hollow cube is concealing?



a) There is a net current looping up along the back and down along the front of the cube.
b) There is a net current looping down along the back and up along the front of the cube.
c) There is a no net current in the cube.
d) This is an impossible situation.
e) Nothing can be said

Questions 34 and 35 refer to the figure to the right. A thin, circular, conducting loop, in the plane of the page, is immersed in a uniform, magnetic field pointed perpendicularly out of the page.



34) If the magnitude of the magnetic field *does not change*, but the radius, *r*, of the loop *decreases*, then which answer below describes something that happens to the loop?

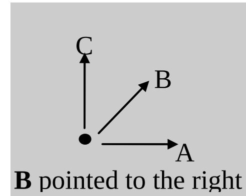
a) Nothing.
b) A clockwise current is produced in the loop.
c) A counterclockwise current is established in the loop.
d) The loop will rotate clockwise in the plane of the page.
e) The loop will rotate counterclockwise in the plane of the page.

35) If the radius, *r*, of the loop *does not change*, but the magnitude of the magnetic field *decreases*, then which answer below describes something that happens to the loop?

a) Nothing.
b) A clockwise current is produced in the loop.
c) A counterclockwise current is established in the loop.
d) The loop will rotate clockwise in the plane of the page.
e) The loop will rotate counterclockwise in the plane of the page.

36) The figure to the right represents a region of a uniform **magnetic field, B, pointed to the right**. If the black dot is a positively charged particle, what arrow would represent the particle's velocity, at this instant, which would result in the maximum force on the particle?

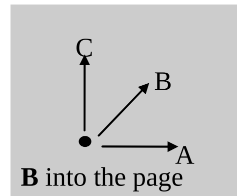
a) A b) B c) C
d) A and C would result in a maximum force, B would not.
e) All the same.



B pointed to the right

37) The figure to the right represents a region of a uniform **magnetic field, B, pointed into the page**. If the black dot is a positively charged particle, what arrow would represent the particle's velocity, at this instant, which would result in the maximum force on the particle?

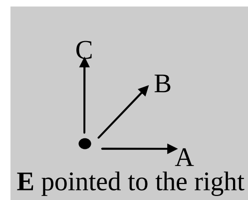
a) A b) B c) C
d) A and C would result in a maximum force, B would not.
e) All the same.



B into the page

38) The figure to the right represents a region of a uniform **electric field, E, pointed to the right**. If the black dot is a positively charged particle, what arrow would represent the particle's velocity, at this instant, which would result in the maximum force on the particle?

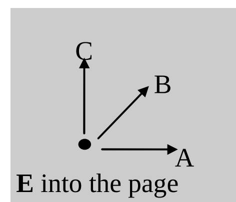
a) A b) B c) C
d) A and C would result in a maximum force, B would not.
e) All the same.



E pointed to the right

39) The figure to the right represents a region of a uniform **electric field, E, pointed into the page**. If the black dot is a positively charged particle, what arrow would represent the particle's velocity, at this instant, which would result in the maximum force on the particle?

a) A b) B c) C
d) A and C would result in a maximum force, B would not.
e) All the same.



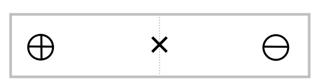
E into the page

For questions 40-57 the crosses, ×, represent a location in space.
The circles represent particles with a net negative charge ,⊖ , or a net positive charge, ⊕. The magnitudes of the net charges are all equal.
The filled-in circles, ●, represent cross-sectional views of long, straight, conducting wires. An "**I**" with a the word "**in**" or "**out**" next to it indicates the direction of current along the wire into or out of the plane of the page respectively. The magnitudes of the currents are all equal.
Any and all fields which may be present in each figure are due only to charges or currents in the figure.
Assume the value of the electrostatic potential equals zero an infinite distance away from a point charge.

For questions 40-47 the vertical, dashed line in the figures represents the midpoint between the two charged particles

40) For the figure in the box to the right, the net electric field at the cross points...

a) → b) ← c) ↑ d) ↓ e) There is no net electric field.



41) For the figure in the box to the right, the net electric field at the cross points...

a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

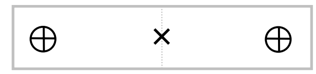
42) What is the relationship, if any, between the net electric field at the cross in question 40 and the cross in question 41?

a) They are both equal. b) The first is larger. c) The second is larger.

d) There is not enough information.

e) There is no relationship.

43) For the figure in the box to the right, the net electric field at the cross points...
a) → b) ← c) ↑ d) ↓ e) There is no net electric field.



44) For the figure in the box to the right, the net electric field at the cross points...
a) → b) ← c) ↑ d) ↓ e) There is no net electric field.

45) What is the relationship, if any, between the net electric field at the cross in question 43 and the cross in question 44?
a) They are both equal. b) The first is larger. c) The second is larger.
d) There is not enough information. e) There is no relationship.

For questions 46 and 47 only consider the space contained within the finite area of the figure.

46) For the figure in the box to the right, where is the electrostatic potential zero?

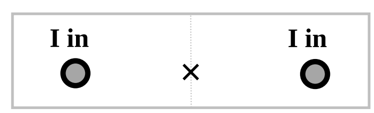
a) Somewhere to the left of the two particles.
b) Exactly halfway between the two particles.
c) Somewhere to the right of the two particles.
d) Nowhere in the figure.
e) More information is needed.

47) For the figure in the box to the right, where is the electrostatic potential zero?

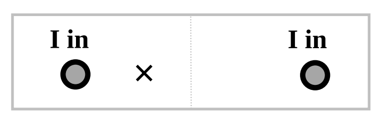
a) Somewhere to the left of the two particles.
b) Exactly halfway between the two particles.
c) Somewhere to the right of the two particles.
d) Nowhere in the figure.
e) More information is needed.

For questions 48-53 the vertical, dashed line in the figures represents the midpoint between the two wires.

48) For the figure in the box to the right, the net magnetic field at the cross points...
a) ← b) ↓ c) Into the page.
d) Out of the page. e) There is no net magnetic field.

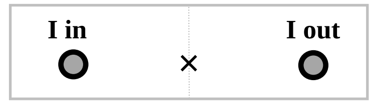


49) For the figure in the box to the right, the net magnetic field at the cross points...
a) ← b) ↓ c) Into the page.
d) Out of the page. e) There is no net magnetic field.

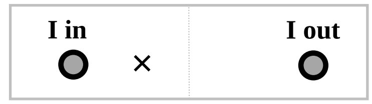


50) What is the relationship, if any, between the net magnetic field at the cross in question 48 and the cross in question 49?
a) They are both equal b) The first is larger c) The second is larger
d) There is not enough information e) There is no relationship.

51) For the figure in the box to the right, the net magnetic field at the cross points...
a) ← b) ↓ c) Into the page.
d) Out of the page. e) There is no net magnetic field.



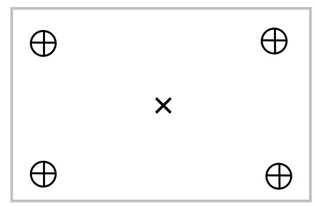
52) For the figure in the box to the right, the net magnetic field at the cross points...
a) ← b) ↓ c) Into the page.
d) Out of the page. e) There is no net magnetic field.



53) What is the relationship, if any, between the net magnetic field at the cross in question 51 and the cross in question 52?
a) They are both equal. b) The first is larger. c) The second is larger.
d) There is not enough information. e) There is no relationship.

For questions 58-59 assume the charged particles are at the corners of an imaginary rectangle and are equidistant from the cross, however, the exact distance is not known.

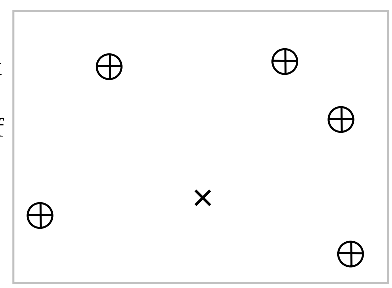
54) For the figure in box to the right, the net electric field at the cross points...
a) Into the page b) Out of the page c) ↑
d) There is not enough information e) There is no net electric field.



55) What is the value of the electrostatic potential at the cross in the figure to the right?
a) 0. b) Less than 0. c) Greater than 0. d) Not enough information.

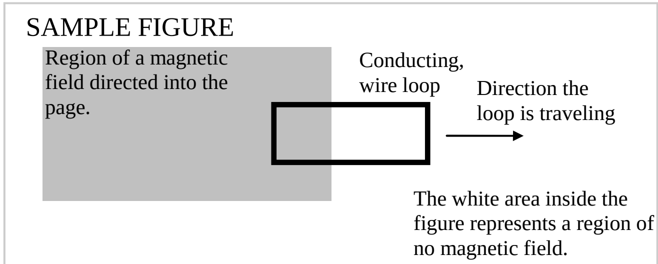
For questions 60-61 assume the distances between the cross and the charged particles are all the same, however, the exact position of the charged particles is *not* known.

56) For the figure to the right, what can be said about the direction net electric field at the cross?
a) The direction can *not* be determined *exactly*, but there is a net field at the cross.
b) The direction can be determined *exactly* and the magnitude of the field at the cross is *not* zero.
c) The field has no direction, because the magnitude of the field at the cross is zero.

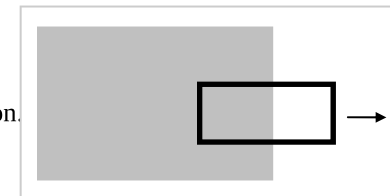


57) For the figure to the right, what is the value of the electrostatic potential at the cross?
a) Less than 0. b) 0. c) Greater than 0.
d) Not enough information.

For questions 58-65 the thick, black rectangles represent conducting, wire loops. Assume the wire to be of negligible thickness. The gray areas represent regions of magnetic field directed perpendicularly into the page. The black arrows represent the direction of motion of the wire loops. No arrow indicates that the loop is not moving. Look at the sample figure below for an example.

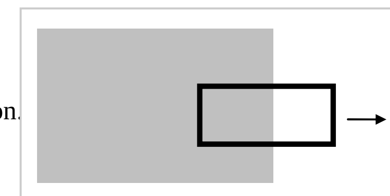


58) Choose an answer that describes something that happens to the wire loop as it moves as shown in the figure to the right? The magnetic field in the gray region is **uniform**.



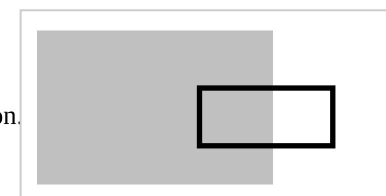
a) There is not enough information given to answer the question.
b) A counterclockwise current is produced.
c) A clockwise current is produced.
d) No current is produced.

59) Choose an answer that describes something that happens to the wire loop as it moves as shown in the figure to the right? The magnetic field in the gray region is **non-uniform**.



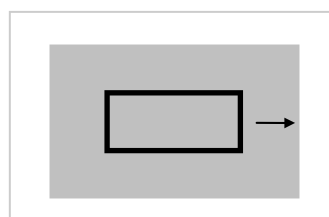
a) There is not enough information given to answer the question.
b) A counterclockwise current is produced.
c) A clockwise current is produced.
d) No current is produced.

60) Choose an answer that describes something that happens to the *stationary* wire loop in the figure to the right? The magnetic field in the gray region is **uniform**.



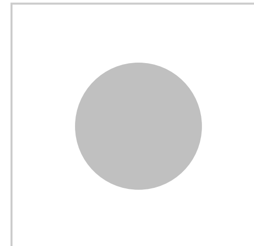
a) There is not enough information given to answer the question.
b) A counterclockwise current is produced.
c) A clockwise current is produced.
d) No current is produced.

61) Choose an answer that describes something that happens to the wire loop as it moves as shown in the figure to the right? The magnetic field in the gray region is **uniform**.



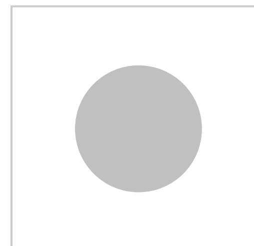
a) There is not enough information given to answer the question.
b) A counterclockwise current is produced.
c) A clockwise current is produced.
d) No current is produced.

62) The gray area represents a region of a uniform magnetic field directed perpendicularly into the page. The magnetic field is decreasing as a function of time. What can be said about the region outside the magnetic field?



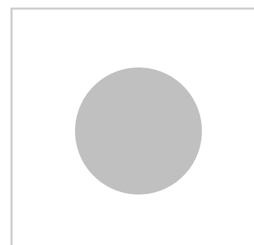
a) There is a net charge building up above and below the field in a plane perpendicular to the page.
b) There is a net charge building up around the field in the plane of the page.
c) In the plane of the page there is an counterclockwise electric field.
d) In the plane of the page there is a clockwise electric field.
e) Nothing.

63) The gray area represents a region of a uniform electric field directed perpendicularly into the page. The electric field is decreasing as a function of time. What can be said about the region outside the electric field?



a) There is a net current building up above and below the field in a plane perpendicular to the page.
b) There is a net current building up around the field in the plane of the page.
c) In the plane of the page there is an counterclockwise magnetic field.
d) In the plane of the page there is a clockwise magnetic field.
e) Nothing.

64) The gray area represents a region of a non-uniform electric field directed perpendicularly into the page. The electric field is decreasing as a function of time. What can be said about the region outside the electric field?

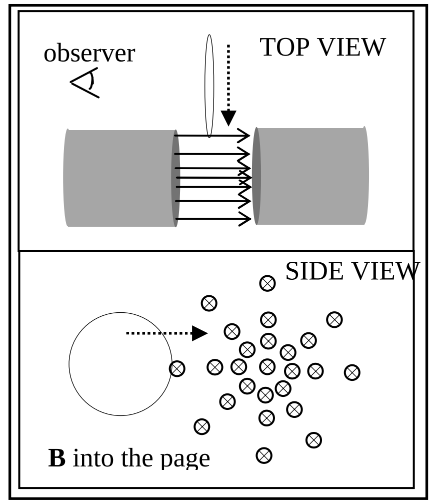


a) There is a net current building up above and below the field in a plane perpendicular to the page.
b) There is a net current building up around the field in the plane of the page.
c) In the plane of the page there is an counterclockwise magnetic field.
d) In the plane of the page there is a clockwise magnetic field.
e) Nothing.

65) An observer watches a thin, circular, conducting, wire loop enter a region of a *non-uniform magnetic field* (see Figure A- TOP VIEW). The thick, dotted, black arrow in both views indicates the direction the loop is moving. Looking at the loop as **it enters** the *non-uniform magnetic field* directed into the page (see Figure A- SIDE VIEW), which answer below describes something, if anything, that happens to the loop as viewed by the observer?

a) A counterclockwise current is created in the wire loop.
b) A clockwise current is created in the wire loop.
c) Nothing could happen to the wire loop as it enters any magnetic field.
d) Nothing happens because the field is non-uniform.
e) No answer can be given based on the information given.

Figure A



66) An observer watches a region of a non-uniform magnetic field pass across a thin, circular, conducting, wire loop (see Figure B- TOP VIEW). The thick, dotted, black arrow in both views indicates the direction the magnetic field is moving. Looking at the loop, as the *non-uniform magnetic field* **begins to passes across it** (see Figure B- SIDE VIEW), which answer below describes something, if anything, that happens to the loop as viewed by the observer?

a) A counterclockwise current is created in the wire loop.
b) A clockwise current is created in the wire loop.
c) Nothing could happen to the wire loop as it enters any magnetic field.
d) Nothing happens because the field is non-uniform.
e) No answer can be given based on the information given.

Figure B