# CHAPTER 3
# METHODS

Interpreting the results of this study will depend both on the chosen research methodology and on the instructional setting. This chapter will discuss the setting, including the specifics of each case, the research design, the measurement instruments, an overview of the data analysis, and the selection of the cohorts.

## Instructional Setting

This study was conducted during an entire academic year of introductory, calculus-based physics at a large, public research university during the 1995/1996 academic year. The university was on a quarter-system, requiring the year-long physics course to be sub-divided into three, 10-week long quarters. In this dissertation, an academic quarter will be referred to as a term. Successful completion of all three terms of the course was a requirement for any undergraduate student majoring in engineering, mathematics, or science. To meet this requirement, about 800 students enrolled in the course during the Fall term of 1995.

### Overview of Course Structure

Given the large enrollment, students could enroll into more manageable sections of about 150 students per section. For the 1995/1996 academic year, there were five sections in the first term, but four sections during the last two terms. A different lecturer with a unique team of teaching assistants (TAs) taught each section of the course. The lecturers were selected from the regular faculty of the Department of Physics and for the 1995/1996 academic year. The TAs were graduate students from the Department of Physics, with the exception of three advanced undergraduate physics majors who were employed as TAs for this course. In addition, every first-year TA participated in a 10-day orientation to acquaint them with the teaching techniques they would use throughout the course.

The lecturer and TAs were responsible for all aspects of their section of the course. These aspects included the lecture, discussion sessions, laboratories, quizzes, exams, homework, and grading. The basic structure of each of these aspects was the same for each section and had the following characteristics:

Lecture

The faculty lecturer taught the lecture and it met three times a week for 50 minutes in a large room with auditorium-style seating. The lecturer had a wide array of demonstration equipment available to enhance his lectures.

Discussion Sessions

The discussion sessions were led by TAs in small classrooms and met once a week for 50 minutes. In discussion sections, approximately 20 students (all from the same course section) solved the same problem in cooperative groups. Students were required to attend discussion sessions. The problem was prepared by either the lecturer or the TAs to be a context-rich problem. Context-rich problems are real-world problem situations which may contain one or more of the following characteristics: "(1) The problem statement does not always explicitly identify the unknown variables; (2) More information may be available than is needed to solve the problem; or (3) Information may be missing, but can easily be estimated or is 'common knowledge;' (4) Reasonable assumptions may need to be made to solve the problem" (Heller, Keith and Anderson, 1992, p. 630).

Laboratories

The laboratories met once a week for two hours and was a required part of the course. Students who did not pass the laboratory, could not pass the course. Part of the students' laboratory grade was based on attendance. Students were in the laboratory with the same people from their discussion sessions. The laboratories were taught by the same TA the students had for their discussion sessions using the same cooperative groups. The students were required to buy a laboratory manual, which was the same for all sections of the course. The laboratory manual was divided into three or four units which lasted for three or two weeks each. The units were based on broad content areas of physics such as Forces, Energy, or Electric Fields. Each unit was further divided into four to seven laboratory problems which required one to two hours to complete.

A laboratory problem required the students to quantitatively solve a context-rich problem and then compare their answer to results generated in the laboratory. The laboratory problems were designed to guide the students without step-by-step instructions. Each laboratory problem had the following parts: (1) a context-rich problem to give relevance to the laboratory problem; (2) a brief description of the equipment; (3) a prediction question that the student had to answer before the start of the laboratory which was based upon the context-rich problem and the equipment; (4) a series of "method questions" designed to aid the student in answering the prediction or in doing the analysis for the lab; (5) a set of exploration guidelines to help the students determine the measurement limitations and relevant precautions for their equipment; (6) a measurement section suggesting possible measurements for the students to make; (7) an analysis section which required the students to compare their prediction to their results; and (8) a conclusion section which asked the students to reflect upon what they learned doing the problem.

The three or four laboratory units were designed to be taught in conjunction with the course; so when the lecturer was discussing forces, the students in the laboratory were doing the forces unit. Due to a limited number laboratory rooms and equipment, every section of the course taught the same laboratory unit at the same time. However, each course section decided on the specific problems they used. The laboratory space constraint helped the various sections of the course retain a fairly uniform pace through the course curriculum.

Quizzes

The quizzes were written by the section lecturer and were administered three or four times a quarter. There were three parts to each quiz; a few multiple-choice questions, two context-rich problems, and a group problem. The group problem was given in the discussion session the day prior to the day of the quiz. All students in a group received the same grade for their solution to the problem.

Final exams

The final exams were handled differently than quizzes. Since the course was divided into three academic quarters and students were free to choose which section to enroll in each quarter, it was vital that each section of the course presented a similar curriculum. Toward this end, every student from each section was simultaneously given the same final examination at the end of each quarter. The final exam was written by all the lecturers from each section.

Homework

The assigned homework for the course was identical across every lecture section. The homework was assigned from the same textbook (Fishbane, Gasiorowicz & Thornton, 1994) that all the students were required to buy, but the homework was not collected or graded. The students were told that one of the homework problems from the textbook would be on the quizzes or final exam, written as a context-rich problem. The students understood that it was in their best interest to do the homework to help them receive a better course grade.

Grading

The grading for the course was handled by the same TAs who taught in the discussion sections and the laboratories. The amount of guidance given to each TA about grading was different in each section. In addition, every lecture section used the same uncurved, grading scale.

Within this common course framework, each lecturer had complete control over the details of their section, such as specific quiz questions, discussion session problems, or problems for the laboratory. The lecturer also taught the lecture alone allowing for his individual preferences for lecture topics, illuminating examples, or demonstrations.

## Specific details of setting in the EPS section

One of the lecturers (KH) in the 1995/1996 academic year took this autonomy a step further. This lecturer was familiar with much of the problem-solving literature reviewed in Chapter 2 of this dissertation and chose to explicitly teach a modified version of the Minnesota Problem-solving Strategy to his students. This lecture section is referred to as the EPS section in this study.

Minnesota Problem-solving Strategy

The Minnesota Problem-solving Strategy was developed by Heller & Heller (1995). The strategy detailed five specifics steps based upon how experts solve real physics problems (as opposed to exercises) and was influenced by the works of Reif & Heller (1982) and Larkin (1983). The five steps of the Minnesota Problem-solving Strategy are: (1) Focus the problem, (2) describe the physics, (3) plan the solution, (4) execute the plan, and (5) evaluate the answer.

To Focus the problem, the students are asked visualize and sketch the events described in the problem statement. From this first interpretation of the problem statement, the students are asked to write a simple statement of what they need to solve for (in words, not equations or numbers) and to decide which physics principles they'll need to use. The goal is to put enough information into this step of the solution so that referring back to the problem statement becomes unnecessary.

The next step of the strategy, Describe the physics, requires the students to turn their personal-language, qualitative understanding of the problem into the language of physics. Their earlier rough sketches become physics representations (such as motion diagrams, vectors, or force diagrams) containing a well-defined coordinate system. Their statement of what needs to be solved for is written with variables. The choice of required physics is expressed in fundamental principles. This step is successfully completed when the students do not need to refer back to the Focus the problem step.

Once the physics description is complete, the students then need to Plan the solution. The purpose of this step is to create an outline of equations to be solved algebraically to produce an answer. Often this outline will solve several sub-problems on route to the single expression that answers the problem.

With the successful completion of the solution plan, the students Execute the plan. This simply requires substituting the known quantities from the problem statement into the final algebraic expression produced in the third step. Finally, the students should Evaluate their solution. Typically this includes checking to see that the answer is properly stated, that the units are correct, and that the answer is reasonable.

The EPS instructor added the Minnesota Problem-solving Strategy to the overall course (described above) with the following modifications to the course:

EPS Lecture

The students were told that the purpose of the lecture was four-fold in the EPS course syllabus, although only the first two are relevant to this study. The first goal was to "show [the students] the motivation for constructing the fundamental physics concepts addressed in this course and show how they are connected to other physics concepts and to the real world" (Heller, 1995). This goal was accomplished in the lecture by lecture content and demonstrations. Most lectures would begin by showing the students where the concepts being discussed fit into the larger hierarchy of physics and why this concept was unique. For example, the concept of momentum conservation was introduced after the students were reminded that nature could be described by conservation theories. However, energy conservation is not complete; it gives no direction information. Physicists have created a vector quantity now called momentum. After this introduction, the EPS instructor proceeded to do a few demonstrations of inelastic collisions that showed the limited use of energy conservation, but the success of momentum conservation. These demonstrations frequently involved the students making predictions of what they expect to see, sharing their prediction with their neighbor, and then voting on predicted behavior prior to the actual demonstration. Thus EPS instructor implemented the first goal of the lecture by introducing the specific concepts framed in the larger hierarchy of physics and the frequent use of demonstrations which actively involved the students.

The second goal of lecture was to "show [the students] examples of how to apply a logical and organized problem-solving technique to problems" (Heller, 1995). This was achieved through modeling the Minnesota Problem-solving Strategy during every worked solution of a context-rich problem in lecture. Modeling, in this context, was similar in use as to cognitive apprenticeship (Collins, Brown, & Newman, 1989) where the preferred performance was demonstrated in great detail to the students. For every example worked in lecture, the EPS instructor worked meticulously through the five steps of the strategy. On some occasions, he stopped the solution after completing only the first few steps, but never was a problem worked without starting from the first step. This modeling of the strategy was re-enforced by occasionally having the students complete one of the problem-solving strategy steps on a notecard to be collected in lecture. These notecards were inspected for good faith effort by the TAs and used to award points to the students for attendance.

<u>EPS Homework</u>

Even though the homework was the same in every section, the EPS students were encouraged to purchase the *Competent Problem Solver* (Heller & Heller, 1995). This supplemental handbook contains detailed instruction on how to use the Minnesota Problem-solving Strategy and several examples of how problems are solved using the Strategy. The students were encouraged to solve their homework using the strategy.

EPS Discussion Sessions

The discussion sessions lead by the TAs assigned to the EPS section were another place where the Minnesota Problem-solving Strategy was reinforced. In their cooperative groups, students worked together to solve the assigned problem using the Minnesota Problem-solving Strategy. The TAs were instructed to coach the students both in their physics knowledge and in the use of the strategy.

EPS Grading

The final reinforcement of the Minnesota Problem-solving Strategy in the EPS course is evident from the grading scheme. The students were told that their "problem solutions will be graded based on how well [they] communicate in writing a logical and organized problem solving process grounded in the correct assessment of the physical situation" (Heller, 1995). The TAs for the course were similarly instructed along these lines including specific instruction based upon the problem being graded. So, while the five-step Minnesota Problem-solving Strategy was not explicitly required, adopting it would be to the student's benefit.

## Specific details of setting in the TRD section

In contrast to the EPS section, another lecturer (BL) had a different style he used within the general course structure discussed earlier. Recognizing the atypical aspects of this course structure, he admitted that his team is "not being totally traditional" (B. Lysak, personal communication, September, 1995), however, he did not emphasize a problem solving strategy in his section. Given this distinction, this section is referred to as the TRD section.

The extent of the TRD instructors problem solving instruction was to discuss with the students the following six points, parroted from the problem-solving strategy included in the textbook (Fishbane, et al., 1993, p 35):

> (1) Read the problem carefully; (2) draw relevant pictures and diagrams, (3) identify known quantities, and which you need to find; (4) plan the solution; (5) execute the plan; and (6) check units, reasonableness, etc.

Even though these six points were presented, no action was taken to support them. The students were not grading for the strategy, nor were these six-points consistently modeled in lecture, nor were there any supplemental materials prepared for these six-points (although the students in TRD section were informed of the *Competent Problem Solver* (Heller & Heller, 1995)).

TRD Lecture

The TRD instructor's style also manifested itself in the course structure in other ways, such as the lecture. Unlike the EPS lecture, the TRD instructor did not explicitly set goals for the lecture in his syllabus, but the implicit goals seemed to have been: (1) present the concepts in their mathematical framework, (2) support these conceptual frameworks with demonstrations, and (3) work example problems. For example, the concept of conservation of momentum was introduced by defining vector momentum as mass multiplied by velocity ($\mathbf{p}=m\mathbf{v}$). With this definition, the time derivative of momentum was shown to be net force. Finally, if there was no net force applied, momentum will be conserved. The utility of momentum conservation was then shown by doing a problem from the textbook (Fishbane, et. al, 1993). When demonstrations were performed, the students were typically kept as passive observers, although occasionally predictions were solicited. In addition, the students were infrequently (one or twice a quarter) asked to solve a textbook problem in lecturer and to turn in their solution.

The TRD instructor did not make any additional noteworthy changes to the rest of the course structure. The TRD TAs used context-rich problems in their discussion sessions and encouraged the students to work in cooperative groups. Essentially the principle difference between the sections was the inclusion of the Minnesota Problem-Solving Strategy into the EPS course.

## Research Design

Given the overwhelming similarities and limited differences between the EPS and TRD lecture sections (summarized in Table 3.1), these two sections would appear, at first glance, to be ideally suited for a quasi-experimental study of the effect of explicit problem-solving instruction in physics. However, there is one unavoidable confounding influence in such a study - the Instructor Effect. This section explores the implications of this effect on the research design of this study.

Table 3.1

Similarities and Differences Between the EPS and TRD Courses.

| Similarity in classes | Differences in classes |
|---|---|
| • Same laboratories which account for the same percentage of the students grade.<br>• Context-rich problems in discussion sections which account for the same percentage of the students grade.<br>• Common final every quarter which account for the same percentage of the students grade.<br>• Cooperative groups in labs and discussion sessions.<br>• Both courses have the same published grading scale.<br>• Three lectures per week.<br>• Meet in teams of TAs to plan course.<br>• Instructors meet weekly to discuss courses.<br>• Both instructors occasionally engage the students in lecture.<br>• Same optional homework sets.<br>• Homework problems show up on exams as context-rich problems.<br>• TAs went through the same 10-day orientation.<br>• Both teams let TAs write group problems.<br>• Same textbook used.<br>• Both the EPS and TRD lecturers are respected full professors. | • EPS class consistently modeled a problem-solving strategy.<br>• The EPS instructor used overheads and had these available for the students in the university book store.<br>• The EPS students were strongly encouraged to buy the *Competent Problem Solver*<br>• The EPS lecturer collected note cards from students with snippets of concepts. The TRD instructor had students work an entire problem to be turned in for credit.<br>• The EPS class is larger: ~250 vs. ~150 students in the TRD class.<br>• The EPS instructor reminded students of concept placement in course. |

Instructor Effect

The Instructor Effect acknowledges the important role the teacher plays in any classroom. Measuring and controlling this effect has proven nearly impossible. As Ornstein (1991) summarizes, "once we start to analyze the various relationships among the variables, we come to numerous interactions and still other new and untested interactions, which in turn can be analyzed. This process is endless." (p.71) The Instructor Effect is too complex to be neatly measured. Rather than trying to quantify the Instructor Effect, most researchers design their experiments around it by involving multiple instructors who provide both control and experimental classes (see Huffman, 1995) or doing a cross-treatment design (see Ward & Sweller, 1990). In these designs, any Instructor Effect will impact both the control and experimental classes. The effect will then hopefully be accounted for by the statistical analysis.

Regrettably, at a large research university it is impossible to set up such an experiment. Most professors teach only one class a term, which prevents having a single instructor teach both a control and an experimental class. Institutional memory (test-files, etc.) prevent an instructor from providing the control and experimental classes in successive years. In addition, professional development to learn and embrace new experimental teaching techniques takes precious time, which at a research university where teaching amounts to at best a third of a professor's time, is not likely to be encouraged or rewarded. It is also impossible to turn on and off problem-solving instruction during a class. Given all these factors, it was impossible to statistically control for the Instructor Effect in a university setting.

Case-study

Fortunately, all was not lost. It was still possible to use the EPS and TRD sections to learn about the development of student problem-solving ability; it was just not possible to compare the two classes statistically in an experimental design. Instead, each class was treated as an independent case study. As case studies, each class will be discussed in this dissertation on its own merits. The two cases will be compared when necessary to interpret the results within a case, and this comparison will be done within the limitations of the research design.

For these two case studies, the case unit will be defined as a matched cohort of students. The case unit is not the students themselves, nor is it either lecture section. The purpose of the match cohort is to follow an equivalent set of students through each course. The matching is done to strengthen each case by removing exceptional or extraneous students who might misrepresent themselves in the data. The specifics of how the matched cohort was selected will be discussed later in this chapter.

With these two matched cohorts of students, one in each section, the study can examine how student's problem solving ability develops. As was discussed in Chapter 2, most of the research of students problem-solving ability have examined it through a single lesson, with typically about 6 weeks of instructional time. To avoid this potential shortfall, the cohorts in this study were followed for an entire academic year.

This time constraint posed a special attrition problem for the cohorts at a large research university where the students self-select their lecture section enrollment time. Since the introductory course is divide into three academic quarters, roughly a third of the students did not re-enroll with the same lecturer. Appendix A provides a complete accounting of this migration. Students who did not remain with the same lecturer (either EPS or TRD) are not elligled to be in the cohorts.

Role of the Researcher

In any valid interpretive study, the bias of the researcher must be made explicit. In the interpretative research paradigm, bias is not a bad word. The interpretative paradigm recognizes that the interpretation of data plays a central role in any science. Furthermore, any interpretation made by humans is inherently biased by the humans making the interpretation. The role of science is to decide how to handle the effects of bias. In a science like physics, the researcher bias is minimized by replication studies. If the physics community can replicate the results, then the original interpretation is considered valid. The history of physics is full of examples of replication challenging the interpretation of results. The claim of cold fusion in the late 1980s is the quintessential modern example.

In education research, replication of research is not easy to do. It is impossible to gather the same students and expose them to the same study under the same conditions. The solution to the interpretation problem in education research is full disclosure of the methods, author's bias, and results. The intent is for any knowledgeable reader to be able to make the same claims based upon the same data. If this condition can be met, then there is validity in the interpretative research paradigm.

In this study, the author had a vested interest in the Minnesota Problem-solving Strategy, both intellectually, personally and financially. The literature review in Chapter 2 solidified the author's intellectual paradigm around the necessity of physics problem-solving instruction. An enormous effort had been put forth personally to create the cooperative group problem-solving environment through the creation of the *Competent Problem Solver*, time spent in TA orientation, and mentoring TAs during the year. Given this near Herculean task, the author has a clear commitment to the EPS section's success. In addition, the author has had his graduate school career partially funded by grants awarded to his advisor and the EPS instructor. All of these biases toward the EPS class suggest the need for full disclosure of the researcher's role.

The role of the researcher during the academic term was to be a non-participant observer of both the TRD and EPS lectures. The researcher used his lecture notes to devise quiz (and one exam) problems which would be accessible to both classes. The researcher then copied all relevant quiz and exams problems for both cohorts. After the year was completed, the researcher selected the cohorts described in more detail in a later section of this chapter. Once the cohorts were selected, all the cohort students' problems were coded in the order they were pulled from the files. This was rarely alphabetical by student's name and the EPS cohort was coded first about half of the time. No attempt was made to mask which cohort was which since the solutions' length and student use of the Minnesota Problem-solving Strategy made such masking attempts artificial and useless. Once the coding was completed, the various analyses reported in this dissertation (and many false starts) were completed.

It must be re-emphasized that bias in any study is inevitable. But once this fact is accepted, the bias can be used to strengthen results. In this study this was tried in two ways. First, when decisions needed to be made, a conscious attempt was made to bias the study toward the TRD cohort. For example, when the content was selected for quiz problems written by the researcher, topics were chosen which were accessible to students in both classes. However, occasionally the TRD students had more experience using the topics. It will also be seen that the choice of problem-solving skills and the selection of the cohort were also biased toward the TRD cohort. The second manner in which the bias was used to strengthen this study was in the order of the analysis. Whenever possible, the EPS cohort was analyzed first and the TRD cohort analyzed in exactly the same fashion. Since there was a bias toward the EPS cohort, the tendency would be to create the best presentation for the EPS cohort. Once this was created, the TRD cohort was analyzed in exactly the same fashion. This way the TRD cohort was subjected to the same best presentation. Both manners of strengthening this study show how bias can be used in a positive fashion.

## Instruments

The case study design in the interpretative paradigms of this study requires several different types of measurements to be made. Some of these measurements will be used to select the two cohorts of students. Other measurements will be made to verify the validity of the results. And there is the principle measurement of the study; coding the students' solutions for problem-solving skill. Naturally there will be some overlap in these measurements and their uses. This section of this study will describe the instruments used to make the measurements. The following section will provide a description of the data analysis procedures. Even though the measurements served different principle purposes, all of the measurements themselves and their analysis were identical for both cohorts.

### Matching Instruments

Nearly every instrument used in this study was used to help select the two matched cohorts of students from each of the two lecture classes.  This section will describe the two instruments principly used for matching students.  Every attempt was made to create two evenly matched teams of students based on data taken at the start of the study.  This a priori approach was used because any post hoc measurements might be biased by the instruction the students received.

### Maryland Physics Expectation Survey

The Maryland Physics Expectation (MPEX) survey (Redish, Saul, & Stienberg, 1998) is a 34-item Likert-scale (agree-disagree) survey that probes student attitudes, beliefs and assumptions about physics.  The items on the survey were chosen after a literature review and discussions with physics instructors.  The MPEX survey was validated through student interviews and by comparisons with expert physicists.  Even though the MPEX survey was a relatively new instrument, it provided a means of assessing student expectations of physics efficiently given the large number of students in the course involved. The MPEX can be found in Appendix E.

### Demographic Survey

In order to understand some of the characteristics and background of the students in the course, a demographic survey was included with the MPEX.  This survey asked about the students' math background, physics background, college major, and hours employed.  The survey was closed-ended, so the students needed to choose from the available answers.

## Verification Instruments

With any interpretive study, the researcher looks for any additional evidence that the interpretations are credible. The best evidence comes from radically different measurements of related concepts. In education research this concept of gathering supporting evidence is called triangulation. In this study there were a few occasions to gather triangulation data and they are described below.

Force Concept Inventory

The Force Concept Inventory (FCI) is a 29-question multiple-choice test specifically designed to measure students' conceptual knowledge of mechanics (Hestenes, Wells & Swackhammer, 1992). The test items were developed through extensive interviews that produced not only reliable questions but also detractor items based on common incorrect answers. The FCI is included in Appendix E. With this format, the FCI authors' claim that a score of 70% correct implies that the student is conceptualizing physics in a manner consistent with Newton's Laws.

When the FCI was first developed, teachers began using it to compare their classes to everyone else. For some it was a revelation (Mazur, 1997). The extent of its use is best illustrated by Richard Hake who tallied over 6000 FCI data points to demonstrate the effectiveness of active engagement instructional approaches (Hake, 1998). However, there has been some debate over the validity of the FCI. The first published objection to the FCI came from Huffman and Heller (1995) who ran a factor-analysis on the FCI and found only one factor. They concluded that the FCI seems to measure something related to a force concept, but not necessarily a complete force concept. They urged caution when using the test. The correctness of using a factor analysis has since been discussed (Huffman & Heller, 1995; Hestenes & Halloun, 1995), but the floodgates were opened.  Several other researchers began to report problems with the FCI (for example Sabella & Stienburg, 1997).  The FCI has become akin to chicken-soup as a cold remedy; the evidence doesn't support it, but everyone continues to use it. Given the aforementioned limitations, only the total score of the FCI will be used.

In-house Concept Tests

Another obvious shortcoming of the FCI is that it only measures a force concept, while a typical introductory physics course covers far more physics content. Unfortunately, there are currently very few well-developed or nationally norm-referenced tests to cover the remainder of the physics concepts.  Given this vacuum, the creation of the remaining concept tests was a task for the Physics Education Research Group at the University of Minnesota and the lecturers of each section.  Two in-house tests were created one for each of the last two academic quarters of the course.  The questions on the first in-house test focused on torque (both static and dynamic), simple harmonic motion and waves.  The concepts covered by the second in-house test included electricity and magnetism.  Both tests can be found in Appendix E.

Course Grades

Within each of the two instructor's courses, overall grades were assigned to the students for their work completed in the classes.  Both instructors used the same two-choice algorithm to compute the student's grade from their exams, quizzes, recitation, lab and lecture attendance. The instructors automatically chose the best possible choice for their students.  The two different grade choices are shown in Table 3.2.

Table 3.2

Two methods to determine the final course grade

| Choice1 | | Choice 2 | |
|---|---|---|---|
| 3 Quizes | 60% | 2 Best Quizes | 40% |
| Final Exam | 20% | Final Exam | 40% |
| Laboratory | 15% | Laboratory | 15% |
| In-lecture | 5% | In-lecture | 5% |

For both instructors, the same grade divisions were published to assign the students' overall grades. These are shown in Table 3.3. However, closer inspection of the grade books showed that these were not rigidly followed. The actual grade divisions are shown in Table 3.4. The differences seen in Table 3.4 between each class and from the published guidelines were not done with malice. Rather the intent was to be certain all the lecture sections gave proportionally about the same grades.

Table 3.3

Published Percent of Total Possible Points Necessary to Receive Each Grade.

| | Course | | |
|---|---|---|---|
| Grade | T1 | T2 | T3 |
| A | 80-100 | 83-100 | 81-100 |
| B | 70-79 | 70-82 | 70-80 |
| C | 50-69 | 50-69 | 50-69 |
| D | 40-49 | 40-49 | 40-49 |
| F | 0-39 | 0-39 | 0-39 |

Table 3.4

Actual Percent of Total Possible Points Necessary to Receive Each Grade.

| | EPS Course | | | TRD Course | | |
|---|---|---|---|---|---|---|
| Grade | T1 | T2 | T3 | T1 | T2 | T3 |
| A | 79.5 | 79.6 | 78.5 | 80 | 81.3 | 78.4 |
| B | 69.5 | 69.5 | 68.5 | 70 | 69.9 | 64.8 |
| C | 50 | 49.5 | 50 | 50 | 49.3 | 40 |
| D | 40 | 41 | 47 | 40 | 39 | 38 |

There were many differences in the grading between the two courses. Since the graduate student teaching assistants (TAs) were different people for each lecturer, the grading done by each TA was different. More importantly, each course emphasized different goals in the grading. As was already mentioned, the EPS class emphasized clear presentation of the material. In the EPS class, the explicit problem-solving strategy was reinforced by grading for a logical presentation. In the TRD class, the emphasis was placed on getting toward the right answer. However, the overall course grades still provided a measurement of the quality of the students, provided the similarities and differences in grading were properly accounted.

## Open-ended Problems

The primary data interpreted for each case were the student's solution to open-ended problems given on the quizzes and exams. To aid in this interpretation, each student in both classes completed identical open-end problems throughout the academic year. As described earlier, the problems the students solved were more than typical end-of-the-chapter problems (Maloney, 1994), they were context-rich problems. The student solutions to the quiz problems were graded by each team's TAs, photocopied, and returned to the students. The final exam solutions were also photocopied, but they were not returned to the students consistent with the policy of the physics department.

The problems for the quizzes were created based upon material presented by both lecturers to their respective students. Both lecturers reviewed the supplied quiz problems and a consensus about the wording was reached before giving the quiz problems to the students. The final exam problems, which are the same for all sections of the course, were created by the lecturers of each course section. The quiz problems were originally designed to be less difficult in the hopes of allowing for more detailed and thoughtful solutions completed by more students. This was not always the case. No such limiting constraint was placed upon the final exam questions. All the problems used in this study are included in Appendix F.

## Overview of Data Analysis Procedures

The purpose of the study is to examine the development of student problem-solving ability within each case. To make the interpretations for each case, the instruments needed to be used in tandem. Even though the open-ended problems were the central artifact being examined; the concept tests, MPEX, demographics, and student grades will all be used to establish the internal validity of the study as well as to select the two matched cohorts. Rarely will any single instrument be used alone.

The next two sections examine how the matching and verification instruments will be analyzed. These sections are very straight-forward. The following sections describe the steps taken to analyze and interpret the open-end problems. Discussed first are the choice of problem-solving skills, how those skills were coded, and how the codes were used. The next section examines the role of problem difficulty on student solutions. Finally there is a discussion of how the codes and problem difficulties are combined to give problem-solving skill development graphs.

## FCI and In-house Concepts Tests

The FCI was given during the first week of class in the laboratories to every student in both lecture sections. The FCI was given again as part of the common final exam. This pretest/posttest design is typically done by the Department of Physics as part of a regular research and evaluation program. The second in-house test emphasizing rigid-body mechanics was given as part of the final. There was no pretest for this test. Similarly, the last in-house test emphasizing topics from electricity and magnetism was only given as a posttest. The students were awarded one point for each correct answer on each test and percent of correct answers could be readily determined.

MPEX and Demographic Survey

Both the MPEX and the demographics survey were given at regular intervals throughout the academic year. The demographics survey and pre-MPEX (the pre and post MPEX differ only in the tense of the belief statements) were given at the start of every academic term. The post-MPEX was given at the end of the last academic term. Items on the MPEX are judged to be favorable if they agree with the beliefs of expert physicist interviewed during the construction of the MPEX. The analysis of the MPEX was done using the calibrations developed by the University of Maryland Physics Education Research Group. However, rather than the favorable/unfavorable graphical analysis used by the Maryland Group, standard descriptive statistics were generated where higher scores denote more expert-like beliefs.

Coding Rubric (Open-ended problem analysis)

Measuring students' problem solving ability is a non-trivial task. Many researchers of classroom physics problem-solving use student grades as such a measure (deJong, & Ferguson-Hessler, 1986). While others do a controlled grading where they have several expert physicists grade identical solutions, compare their respective scores, then reach consensus on differences (Mestre et al., 1993). Both methods fall short of measuring problem-solving ability. Instead what they measure is the correctness of a solution. A student who chose to apply the wrong physics concepts, yet applied them correctly, typically received a low grade. Conversely, a student who managed to reach a correct solution by manipulating all of the given information in a haphazard manner would probably receive a high grade, yet failed to display a desirable problem-solving ability. This is not meant to imply that grades are without merit, only that grades measure something slightly different than problem-solving ability. For this study, a more sophisticated technique needed to be applied.

To develop a technique to measure problem-solving ability, the first step is to decide what is meant by "physics problem-solving ability." The expert-novice problem solving literature provides a starting point. From this literature reviewed in Chapter 2, many relevant behaviors are evident in expert solutions:

(1) experts have better organized knowledge (Chi, et al., 1981);

(2) expert perform an initial qualitative analysis of a problem (Larkin, 1979);

(3) experts use their initial analysis to create a domain specific representation and diagrams (Larkin & Reif, 1979);

(4) experts work from general principles to the desired goal (Larkin, 1980);

(5) good problem-solvers plan their solution before starting it (Finegold & Mass, 1985);

(6) expert problem solvers evaluate their work (Reif & Heller, 1982).

In addition to this list, physics instructors could add their own pieces to the definition of good physics problem solving:

(7) good solutions are clear and orderly;

(8) good solutions cite examples from lecture or lab;

(9) good solutions do not have any math errors;

(10) good students manipulate their equations algebraically before substituting in numbers;

(11) good problem solvers start with fundamental principles and not specific equations;

(12) experts frequently form alternative arguments or representations to double-check their results.

This is a daunting list of skills. Not only would it be difficult to assess all of these skills, any measurement that did would be bias against the TRD cohort. These students were not required to evaluate their solutions, or be planfull, or refrain from substituting numbers, all skills for which the EPS cohort was graded. Rather a fair subset of skills needed to be selected and applied to both cohorts. These skills are assessed by a problem-solving ability coding rubric used in early studies (Blue, 1997), but refined for this study.

The problem solving ability coding rubric used in this study has four dimensions. These dimensions and sub-codes are in Tables 3.5 through 3.8. The first dimension was General Approach (GA). This dimension assessed the student's initial qualitative approach. It is in this dimension that any conceptual error the student made will be reported. The second dimension is Specific Application of Physics (SAP). This dimension is the assessment of the student domain-specific knowledge. A student's SAP is dependent upon their GA, so even if the concepts applied are not wholly appropriate for a successful problem solution; the application of those concepts are still judged. The third dimension is Logical Progression (LP). This dimension codes a student's cohesiveness of the solution. It also measures whether a students works forward or backwards (Larkin, 1980). The final dimension of the coding rubric is Appropriate Mathematics (AM). This dimension accounts for a student's level of mathematical skill as applied to the specific problem. It is essentially a judge of a student's ability to transfer the mathematics they learned in the context of a math class to the new context of a physics class.  Each of the dimensions will be coded and reported separately.

Table 3.5
<u>General Approach</u>

0 Nothing written
1 Physics approach is inappropriate. Successful solution is not possible
2 Physics approach is appropriate, but the manner of its application indicates a fundamental misunderstanding.
3 Physics approach is appropriate, but a wrong assertion is made as a serious misinterpretation of given information.
4 Physics approach is appropriate, but neglects one or more other principles necessary for the solution.
5 Physics approach is appropriate and all necessary principles included, but errors are evident.
6 Physics approach is appropriate and all necessary principles included without any conceptual errors.

Table 3.6
<u>Specific Application of Physics</u>

0 Nothing written.
1 Difficult to assess (GA#2).
2 Solution does not proceed past basic statement of concepts.
3 Vector/scalar confusion, or specific equations are incomplete, or confusion resolving vectors into components.
4 Wrong variable substitution: Poor variable definition.
4.5 Wrong variable substitution: Difficulty in translating to a mathematical representation.
5 Careless use of coordinate system without a coordinate system defined.
5.5 Careless use of coordinate system with a coordinate system defined.
6 Careless substitution of given information.
7 Specific equations do not exhibit clear inconsistencies with the general approach, but hard to tell due to poor communication.
7.5 Specific equations do not exhibit clear inconsistencies with the general approach and the solution is clear.

Table 3.7
<u>Logical Progression</u>

0 Nothing written.
1 Not applicable - one step problem.
2 The use of equations appears haphazard and the solution unsuccessful. Student may not know how to combine equations.
3 Solution is somewhat logical, but frequent unnecessary steps are made. Student may abandon earlier physics claims to reach answer.
4 Solution is logical, but unfinished. Student may stop to avoid abandoning earlier physics claims.
5 Solution meanders successfully toward answer.
6 Solution progresses from goal to answer.
7 Solution progresses from general principles to answer.

Table 3.8
<u>Appropriate Mathematics</u>

0 Nothing written
1 Solution terminates for no apparent reason
2 When an obstacle happens, “math magic” or other unjustified relationships occurs
3 When an obstacle happens, solution stops.
4 Solution violates rules of algebra, arithmetic, or calculus
5 Serious math errors
6 Mathematics is correct, but numbers substituted at each step
7 Mathematics is correct, but numbers substituted at last step.

In spite of the efforts to assess a fair sub-set of skills, a clever reader might suspect that the coding rubric will still favor the students in the EPS section. This is not a concern. The coding rubric can be used to judge solutions from either class fairly. Blue (1997), in her doctoral thesis demonstrated that a similar coding rubric can be used for any problem solution regardless of problem-solving instruction. In addition, Appendix G shows six student solutions and the problem-solving skill codes they received. Also, since this study is examining the development of problem-solving ability, any systematic error reflecting this supposed bias from the coding rubric's theoretical foundation will affect every solution within a lecture section equally. If this bias exists, it will not impact the longitudinal results within a section.

Problem Difficulty (Open-ended problem analysis)

Many different people wrote the open-ended problems used on the exams and quizzes. These problems were very straight-forward and simple, or difficult and convoluted, or anywhere in-between. Given this variability, the problem difficulty needed to be measured prior to determining the development of students problem-solving ability. Without this difficulty determination, an easy problem might artificially inflate a student problem-solving ability score much like measuring an expert's problem-solving ability using a novice problem fails to elicit the expert's full ability. Conversely, students faced with an extremely difficult problem might not even be able to start a solution regardless of their true problem-solving ability. Any such misinterpretation of an individual's problem-solving ability, while not a significant problem for one solution, is detrimental when looking for development across many sequential problems. There are two possible determinations of the problem difficulty, one made before giving the problems to the students and the other after the students have completed their solutions to the problem.

The a priori determination is based on knowing what characteristics of context-rich problems make them more difficult. This work was started by Heller & Hollabaugh (1992) but has since been enhanced. Originally there were only a handful of recognizable characteristics, now there are 21 traits which make problems more difficult. If a problem has all twenty-one, it will be nearly impossible for a student to solve, while a problem with only one of these traits is an easy problem. For individual quiz or exam problems, there should be about five of these traits.

All of the traits that make a context-rich problem difficult have face-validity; they are fairly common-sense. Additionally, many of these traits are supported by research that highlights known conceptual difficulties as well as problem-solving pitfalls students have. The traits themselves are grouped into three major categories, each with two or three subcategories. While these categories are not mutually exclusive, they are helpful in deciding the difficulty traits of a problem. The three categories are Approach to the Problem, Analysis of the Problem, and Mathematical Solution. The traits in each category are described below and summarized in Table 3.9.

Approach to the Problem

The traits in the Approach category are grouped together because they all affect how a student decides which concepts, principles and laws to apply to a problem. In a traditional textbook problem this is often given to the students either by a direct statement, such as "the carts have an inelastic collision" or merely by placing the problem at the end of the chapter under a subheading such as “Inelastic Collisions.” Without such cues, the following seven problem traits can make it more difficult for students to decide how to approach a problem.

1. *Problem statement lacks standard cues:* Novice problem solvers often decide on an approach from the "cues" in a problem statement. The two difficulty traits in this subcategory thwart this tendency.

*A. No explicit target variable.* The unknown variable of the problem is not explicitly stated. Problems with this difficulty trait typically include statements such as: "Will this plan (design) work?" or "Should you fight this traffic ticket in court." Novice problem solvers often use the explicit statement of the desired variable (e.g., find v) as a cue to the concepts and principles they should apply to the problem (Chi, Feltovich, & Glaser, 1981). Research has shown that a problem without an explicit statement of the goal is very difficult for the students to solve. (Vollmeyer, Burns, & Holyoak, 1996).

*B. Unfamiliar context.* The context of the problem is very unfamiliar to the students. If the students have no experience with a context, such as particle accelerators, they have difficulty creating a mental translation from the problem statement to their understanding of what the problem is asking. This translation is critical for any successful problem solution (Larkin, et al., 1980).

*2. Solution requires agility in using principles.* As novices in physics, most students are not initially very adept at using the fundamental principles they just learned, or in making connections between principles. Problem statements which require the students to appreciate a concept's complexity will be more difficult for them. The next three difficulty traits are examples of how problem statements can tax the students' fluency with principles.

*A. Very abstract principles.* The central concept required to solve the problem is an abstraction of another abstract concept. Most college freshmen tend to be concrete thinkers (King, 1986), yet many of the more interesting topics in physics tend to be abstract concepts, such as fields or the angular momentum vector. While these concepts are abstract, their presence alone does not warrant a difficulty trait. Rather, when the problem requires the students to use an abstract concept that is based upon another abstract concept, then this difficulty trait is present. An example would be magnetic flux or the direction of a changing angular momentum vector.

*B. Choice of useful principles.* The problem has more than one possible set of useful fundamental principles that could be applied for a correct solution. For example, consider a problem with a box sliding down a ramp. Typically either Newton's Laws of Motion or the conservation of energy will lead to a solution, but deciding which principles to use can be difficult for students. In this particular example, researchers have shown that novice problem solvers will often group these problems as "incline plane" problems instead of by the physics principle needed to solve the problem (Chi, Feltovich, & Glaser, 1981).

*C. Two fundamental principles.* The correct solution requires students to use two or more fundamental physical principles. Examples include pairings such as Newton's Laws and kinematics, conservation of energy and momentum, conservation of energy and kinematics, or linear kinematics and torque. Combining what the students learned several weeks ago with a current principle is difficult for the novice student who perceives physics as a set of incoherent topics (Hammer, 1994).

*3. Non-standard Application of Concepts and Principles.* Students typically learn new concepts or principles by solving problems that require only a simple, straightforward application of the concept or principle. For example, students initially learn Coulomb's Law by solving problems that require the determination of the total force on a charge located at known distances from other charges. The two difficulty traits in this subcategory require students to generalize their problem-solving knowledge to atypical situations or combinations beyond the standard situations.

*A. Atypical situation.* The setting, constraints, or complexity is unusual compared with standard application problems. That is, the problem combines objects or interactions that are not normally put together, such as angular momentum and simple harmonic motion. This trait challenges the students' naïve pattern-matching problem-solving strategies (Chi, Bassok, Lewis, Reimann, & Glaser, 1989).

*B. Unusual target variable.* The problem involves an unusual target variable. The problem asks students to solve for an unknown variable which is usually supplied in standard application problems. This trait also challenges the students' naïve pattern-matching problem-solving strategies (Chi, et. al., 1989).

Analysis of the Problem

The difficulty traits in the Analysis category tax the novice problem solving strategy of plugging numbers into formulas without taking sufficient time and care to analyze the problem. Problem analysis is the translation of the written problem statement into a complete physics description of the problem. It includes a determination of which physics concepts apply to which objects or time intervals, specification of coordinate axes, physics diagrams (e.g., a vector momentum diagram), specification of variables (including subscripts), and the determination of special conditions, constraints, and boundary conditions (e.g., $a_1 = a_2 =$ constant). The next nine traits are all examples of how problems that require a careful and complete qualitative analysis are more difficult for students to solve.

*4. Excess or Missing Information.* Typical textbook problems give exactly the information necessary to solve the problem. Consequently, some students use these values in helping them decide which "formulas" they need to solve the problem. Excess or missing information in a problem thwarts this naive strategy and requires students to analyze the problem situation to decide how to proceed.

*A. Excess numerical data.* The problem statement includes more data than is needed to solve the problem. For example, the inclusion of both the static and kinetic coefficients of friction in a problem requires students to decide which frictional force is applicable to the situation. Excess information challenges the naive student problem-solving strategy of trying to use every number in the problem statement. When more information is given in a problem, some students will force this information into the solution.

*B. Numbers must be supplied.* The problem requires students to either remember a common number, such as the boiling temperature of water, or to estimate a number, such as the height of a woman. When students are required to generate their own values in the midst of other given values, then the problem has become more difficult.

*C. Simplifying Assumptions.* The problem requires students to generate simplifying assumptions to eliminate an unknown variable or term in an equation. *All* problems require students to use their common sense knowledge of how the world works (e.g., boats move through water and not through the air!). Typically assumptions, such as frictionless surfaces or massless strings, are explicitly made for the students in class or in textbooks. Therefore, asking students to make their own simplifying assumptions is a new and difficult task. Problems that require students to make their own simplifying assumptions are more difficult to solve. To be included as a difficulty trait, the simplifying assumption must be uncommon, such as ignoring a small frictional effect when it is not obvious to do so. Simplifying assumptions recognized by this difficulty trait are used to denote only uncommon assumptions of essentially two classes: neglect and ignore. The first class is instances where the students must *neglect* a quantity, such as neglecting the mass of a flea when compared to the mass of a dog. The next class of assumptions involve *ignoring* effects that cannot be easily expressed mathematically, such as how a yo-yo’s string changes its moment of inertia.

*5. Seemingly missing information.* The problem requires students to generate a mathematical expression from their analysis of the problem. This expression might be derived from their understanding of how real systems work or from a careful diagram. The creativity involved in overcoming this class of obstacles is not normally encountered in textbook problems nor is it usually taught. There are three traits in this subcategory.

*A. Vague mathematical statement.* The problem statement introduces a vague, new mathematical statement. For example, if the problem statement tells the students that "A is proportional to B," then the students must not only translate the written statement into a mathematical expression, but then know where and how to use it.

*B. Special Conditions or Constraints.* The problem requires students to generate a mathematical expression from their analysis of the special conditions or constraints of the problem.

*C. Necessary relationships from diagrams.* The problem solution requires using the geometry/trigonometry of the physical situation to generate a mathematical expression to eliminate an unknown. This characteristics adds difficulty to a problem since it emphasizes the necessity of a diagram, a skill many novice problem-solvers lack (Larkin & Reif, 1979).

6. *Additional Complexity.* The problems in this subcategory require students to be especially careful in their analysis and variable definitions. The more "pieces" students have to keep track of, the more difficult the problem.

*A. More than two subparts.* The problem solution requires students decompose the problem into more than two sub-parts. Two or more sub-parts can arise because there are more than two interacting objects or more than two important time intervals. Changing systems of interest for students can be hard. Also novice problem-solvers will often lose sight of the problem goal through numerous subparts.

*B. More than 4 terms per equation*. The problem involves five or more non-zero terms in a principle equation. After a principle required in a problem has been identified by a student, it still needs to be applied. Problems which have five or more terms for the students to include in their equation not only tax the student's procedural knowledge base (Reif & Heller, 1982), but push the limits of their short-term memory (Miller, 1957).

*C. Two directions (vector components).* The problem requires students to apply the same principle (e.g., forces, conservation of momentum) in two directions. This requires both the decomposition of the physics principle and the careful subscripting of variables. Moreover, understanding and exploiting the richness of vectors is a daunting task for most college freshmen. Some students are still tripped up taking vector components even after weeks of using vectors.

Mathematical Solution

Mathematical difficulty is the last category of traits. There are some simple mathematical hurdles a teacher can put into a problem that prevents some students from reaching a final answer. Some of these are included in the last five traits.

*7. Algebraic solution.* A strictly algebraic solution is challenging for many novice problem-solvers. There are three problem types that can require algebraic solutions.

*A. No numbers.* The problem statement does not use any numbers. Expert problem-solvers routinely solve problems without substituting in numbers until the very end. They understand the richness symbolic expressions give them, especially in checking their work (Larkin & Reif, 1979). Students, however, do not appreciate this richness. Nearly every experienced teacher has heard their students groan when faced with a numberless problem. Many students use numbers as place holders to help them remember which variables are known and which are unknown (Wenger, 1987). Therefore if a problem is written without numbers, it is more difficult for the students.

*B. Unknown cancels.* Any problem in which an unknown variable, such as a mass, ultimately factors out of the final solution is more difficult. The students must not only decide how to solve the problem without all the cues they expect, but keep track of all the variables.

*C. Simultaneous equations.* The problem requires simultaneous equations for a solution. Simultaneous equations are hard for the students not only because of the algebra involved (a skill many students lack), but because there are at least two unknowns in each equation and they need to keep track of these variables. A typical circuit-analysis problem best illustrates this trait.

*8. Targets math difficulties.* The problems in this subcategory require students to use mathematics that is known to be problematic.

*A. Calculus or vector algebra.* The problem requires calculus, or vector cross products for a correct solution. Most students are still learning these skills in their math courses and have not learned how to transfer these skills from their math class to their physics class.

*B. Lengthy or detailed algebra.* A successful solution to the problem is not possible without working through lengthy algebraic calculations. While these calculations are typically not difficult, they do require care.

Using the twenty-one problem difficulty traits (summarized in Table 3.9) required the researcher to solve each problem, then tally each trait evident. For the first stage of tallying, the researcher assumed that the students had familiarity, but not mastery of the relevant concepts. The actual familiarity the students had with each problem was factored in later. In addition, this tallying was not necessarily binary. If a problem was set in an extremely unfamiliar context, then the researcher awarded doubled the weight of this one characteristic, although this double-weighting was kept to a minimum. The converse was also true. If a problem required the use of vectors in an almost trivial setting, then the weight of this one characteristic was halved. These tally were added together to produce the initial difficult ranks for each problem statement.

The next stage of rating the difficulty was for the researcher to adjust the initial difficulty ranks to account for the familiarity the students should have with the concepts central to the problem. For example, the concept of energy conservation can be hard to learn, so a problem given during the first week of instruction in this concept would be more difficult than if the same problem was

Table 3.9:
Summary of the 21 difficulty traits.

## Approach

### 1 Cues Lacking

A. No explicit target variable.
B. Unfamiliar context.

### 2 Agility with Principles

A. Choice of useful principles.
B. Two general principles.
C. Very abstract principles.

### 3 Non-standard Application

A. Atypical situation.
B. Unusual target variable.

## Analysis of Problem

### 4 Excess or Missing Information

A. Excess numerical data.
B. Numbers must be supplied.
C. Simplifying assumptions.

### 5 Seemingly Missing Information

A. Vague statement.
B. Special conditions or constraints.
C. Diagrams.

### 6 Additional Complexity

A. More than two subparts.
B. Five or more terms per equation.
C. Two directions (vector components).

## Mathematical Solution

### 7 Algebra Required

A. No numbers.
B. Unknown(s) cancel.
C. Simultaneous equations.

### 8 Targets Math Difficulties

A. Calculus or vector algebra.
B. Lengthy or Detailed Algebra.

given later in the course. However, if the students have not used a set of concepts for several weeks, the students may lose familiarity with those concepts. The course syllabi and lecture observation notes were used to determine when concepts were introduced and if their use was continuous. If necessary the ranks were adjusted by one. These familiarity adjusted difficulty ranks are shown in Table 4.5.

Another method of determining the difficulty of a problem was to give the problem to the students and assess the damage. A very easy problem results in the majority of the class getting it correct. A very difficult problem stumps most of the students. In order to avoid any confounding influence from the partial credit assigned to each problem, only an absolutely right solution should be counted as correct and subsequently used to validate the difficulty rankings. Chapter four demonstrates that the a priori problem difficulty rankings were related to student performance.

Development Graphs (Open-ended problem analysis)

The final step in analyzing the students' solutions to the open-ended problems was to combine the coding rubric and difficulty ranks to produce development graphs. Development graphs can be created for individuals, for the cohorts, or sub-groups of the cohorts. These graphs form the heart of the analysis in this dissertation. This section will describe their creation.

Horizontal axes

Figure 3.1 shows the General Approach development graph for the EPS cohort. The horizontal axes is a date axes. There are six data points in the development graph and they map onto the date axes. The first data point is the average of the third and the fourth quizzes of the first term (T1-Q) shown on the date of the third quiz (November 17). The first two quizzes of the term were not included in this study. The second data point (December 5) is the average of all six problems given on the first term final exam (T1-F). The third data point is the average of the three second-term quizzes (T2-Q) and is the date of the second quiz (February 9). The fourth data point (March 14) is the average of all six problems from the second term exam (T2-F). The fifth data point is the average of the three third term quizzes (T3-Q) and is the date of the second quiz (May 3). The last data point (June 4) is the average of all six problems of the third term final exam (T3-F).

The quizzes were grouped together to smooth out fluctuations from one problem. Recall that only one problem was analyzed per quiz. As will be shown later in this study, a single problem in not necessarily an accurate measure of student skills at that instant in time. The quizzes are therefore combined and shown on the date of the middle quiz.

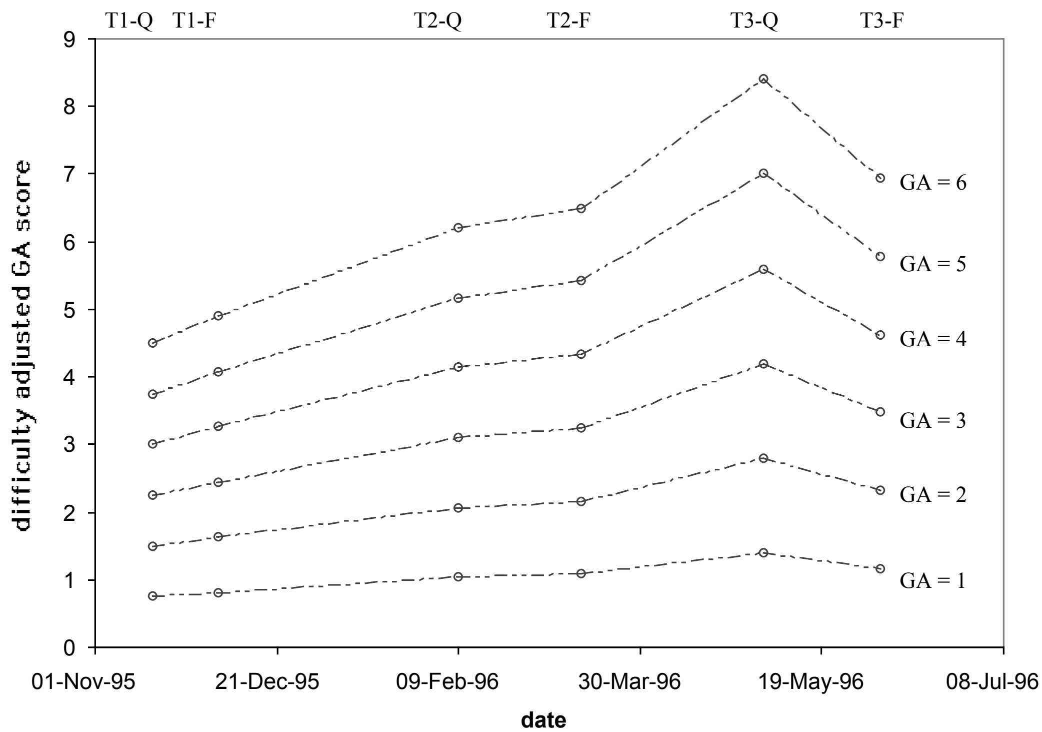


Figure 3.1

Development graph template for the General Approach skill showing the difficulty bands for the EPS cohort.

## Vertical Axes

The vertical axes of Figure 3.1 is the difficulty adjusted average score of the cohort. An average score was used at the cohort level. For the individual graphs (shown in Appendix B) the median score was used. This distinction is necessary given the smaller number of solutions and the fluctuations seen in the individual data.

In this dissertation a score was created from the code assigned to each student's solution to a problem by multiplying the numerical code by the difficulty rank for each problem. This number was normalized by dividing it by five which was near the average difficulty rank for both cohorts. Therefore the higher the score, the better the student's performance.

Another complication added to the development graphs was unfinished student work and how it was handled. For the individual development graphs there was no adjustment made for this phenomena. If the students had incomplete solutions, this was a reflection of their problem-solving skill. At the cohort level, it was another matter. What this thesis is principally interested in is student development. A blank solution does not show development only that the solution is unfinished. While this is important data, including the blank solutions in with completed solutions biases the development curve downward. Therefore on several development graphs, the blank solutions will be left out of the average score and the number of blank solutions will be reported in an accompanying table.

Skill Bands

The dashed lines on Figure 3.1 are the skill bands. The skill bands represent each of the specific problem-solving skill codes. The skill bands enable interpretation by allowing the student's score to be compare to the codes. For example, if the student's average scores fall near the GA = 2 skill band, then it can assumed that on average the student has serious misconception errors.

Each skill band is computed by using the same code for each problem, transforming the code into a score, and then averaging the score according to their date. In effect it is as if the student solution was always the same score. For example Figure 3.1 is a graph of the General Approach skill and this skill has seven codes ranging from zero to six hence why there are seven skill bands. The skill code is listed on the right hand side of the development graphs.

Since the skill scores are computed from an average, error bars should be on each point. However, the error bars reach to the midpoint of next highest line whose error bars overlaps those of the line beneath them. It is difficult to distinguish the error bars from each other. Regardless of this visual impediment, it is possible to think of these lines as skill bands covering the space of the graph. These skill bands allow a direct correspondence back to the problem-solving skill codes to allow for a more complete description of student performance.

The growth of the skill bands over time is a direct result of the problem difficulty ranks. As shown in Table 4.5, the difficulty ranks increase during the year with the most difficult problem on average occurring on the third term quizzes. The shape of the skill bands graphically shows this.

## Matched Cohort

The first application of the aforementioned instruments was to find the matched cohort of students. Recall that the purpose of the matched cohort was to create and follow two evenly matched teams of students through each of the two instructor's courses. The hope is that these teams would simulate what would happen if the same set of students could have experienced both forms of instruction for the first time. The students were matched on the following characteristics: sex, FCI score, problem-solving Specific Application of the Physics and Logical Progression skills for the first problem (T1-Q3), physics experience, math background, MPEX score, age, and hours employed. Each of these matching parameters had a different level of significance. This section of the study is dedicated to presenting the logic used to match the students for each cohort. The cohorts are presented in Tables 3.7 and 3.8.

The students were Boolean matched for sex. This simply means that a man was never matched with a woman and vice versa. Blue (1997) made a compelling case for expecting sex differences between males and females on problem-solving ability, and even though her study did not find a significant difference, this author would be foolhardy to ignore this obvious matching parameter.

The next three parameters (FCI, Specific Application of the Physics and Logical Progression) were weighted equally at triple the base weight for a successful match. To judge the student's conceptual understanding of mechanics, their pretest score on the FCI was used. To match, two student's scores must be within a point and half of each other on the raw FCI score.

To judge the student's initial problem-solving ability for Specific Application of the Physics and Logical Progression, the third quiz of the first academic term was used. To match, the students must within a code of each other. The problem used was T1-Q3 in Appendix F. Even though the third quiz occurred six weeks into the term (quarter), it provided the best chance to get this assessment. Huffman (1994) reports that after six weeks of instruction there was no significant difference between students who had instruction emphasizing the Minnesota Problem-solving Strategy and the textbook strategy. Based on Huffman's study, six weeks is as good as the first week. Also, the sixth week allows those students who were predisposed to change their problem-solving strategy to be identified. By waiting until this event occurs, a better assessment of the development can be obtained without a confounding spike near the second quiz. The spike would occur at the second quiz because the students wouldn't change until after their score on the first quiz demonstrated a need to improve.

The next three parameters were weighted equally at twice the base weight for a successful match. To prevent a student who had never taken a physics course from being matched with a three-time veteran of the introductory college course, the student's physics backgrounds were matched. Unlike sex, a mismatch on this measure will not prevent a student from being matched, but it will reduce their overall match score. Similarly the student's mathematics backgrounds were matched. To match, the students must be with an academic quarter of the same math class. So a student who had taken the third term of college calculus could be matched with a student who was just finishing their second term of college calculus. Finally, the students were matched on their initial, first-quarter responses to the MPEX. With the MPEX, two students could be different by up to 0.5 on their average score (out of a possible 5).

The last two parameters were awarded the base weight for a successful match. If the students were within a year of being the same age, it was considered a match. Some students did not report an age, so a "?" is used as place marker. For the hours employed, if the students were within 10 hours per week of being the same, it was considered a match.

After all of the parameters have been considered the fewer students from TRD section were matched with the larger sample of EPS students. The maximum possible score is 17. The minimum score used was ten. The total matching scores are presented in Table 3.9. The matching was made trying to maximize the number of students in the cohorts.  Compromises were made in total match score to allow for more students into the study.

There were 39 students in the TRD class who stayed with the same instructor for all three terms.  Four of these students either did not consent to be in the study or did not complete all the matching instruments.  This left 25 men and 10 women.  The EPS cohort had 76 students who stayed with the same instructor for all three terms.  Only 68 of them completed the necessary instruments and consented to be in the study.  15 of these students were women and the remaining 53 were men.

There were 25 matches for the TRD students found in the EPS cohort.  There were five men of mixed ability who could not be matched.  There were five women who generally had FCI pretest scores too high to find a match among the women in the EPS cohort.  These students are shown in Tables 3.7 and 3.8.

Table 3.7

Student Cohort from EPS section

| Student | SEX | FCI | SAP | LP | PHYS | MATH | MPEX | AGE | WORK |
|---|---|---|---|---|---|---|---|---|---|
| EPS01 | F | 11 | 7.5 | 5 | HS | 5 | 3.1 | ? | 0 |
| EPS02 | F | 5 | 7.5 | 4 | HS | 5 | 3.5 | 21 | 3 |
| EPS03 | M | 14 | 7.5 | 7 | HS | 5 | 3.5 | 18 | 1 |
| EPS04 | M | 24 | 1 | 6 | HS & U | 5 | 2.7 | ? | 0 |
| EPS05 | M | 24 | 7.5 | 7 | HS | 4 | 3.6 | 18 | 0 |
| EPS06 | M | 12 | 2 | 0 | HS | 4 | 3.2 | 18 | 2 |
| EPS07 | M | 13 | 3 | 6 | HS | 7 | 3.2 | 18 | 0 |
| EPS08 | F | 8 | 7.5 | 3 | HS | 6 | 3.1 | 18 | 0 |
| EPS09 | M | 25 | 7.5 | 7 | HS | 5 | 3.9 | 22 | 0 |
| EPS10 | M | 12 | 4.5 | 5 | HS | 5 | 3.4 | 18 | 0 |
| EPS11 | M | 15 | 2 | 0 | HS | 5 | 2.7 | 19 | 3 |
| EPS12 | M | 17 | 3 | 7 | HS | 3 | 3.6 | 18 | 0 |
| EPS13 | M | 14 | 7.5 | 7 | HS | 7 | 3.1 | 18 | 0 |
| EPS14 | F | 11 | 7.5 | 7 | HS | 5 | 3.5 | 19 | 1 |
| EPS15 | M | 13 | 5.5 | 4 | HS | 7 | 3.7 | 18 | 0 |
| EPS16 | M | 16 | 7a | 4 | HS | 7 | 3.3 | 19 | 1 |
| EPS17 | M | 13 | 6 | 7 | HS | 4 | 3.1 | 18 | 0 |
| EPS18 | M | 17 | 6 | 6 | HS | 6 | 3.3 | 18 | 0 |
| EPS19 | F | 11 | 7.5 | 3 | HS | 6 | 3.4 | 20 | 1 |
| EPS20 | M | 17 | 7.5 | 7 | HS | 3 | 3.2 | 18 | 0 |
| EPS21 | M | 13 | 7.5 | 7 | HS | 6 | 2.8 | ? | 0 |
| EPS22 | M | 10 | 7.5 | 7 | HS | 6 | 3.0 | ? | 0 |
| EPS23 | M | 13 | 7.5 | 4 | HS | 5 | 3.2 | ? | 0 |
| EPS24 | M | 10 | 7.5 | 3 | HS | 6 | 3.0 | ? | 2 |
| EPS25 | M | 11 | 3 | 6 | HS | 5 | 2.4 | 20 | 0 |

Phys Code: HS = high school physics only;
HS & U = high school physics and a university physics course.

Math Codes: 3 = trigonometry; 4 = first quarter calculus:
5 = second quarter calculus; 6 = third quarter calculus;
7 = other.

Work Codes: 0 = none; 1 = 1-10 hours per week; 2 = 11-20 hours per week;
3 = 21-30 hours per week; 4 = 30 or more hours per week.

Table 3.8

Student Cohort from TRD section

| Student | SEX | FCI | SAP | LP | PHYS | MATH | MPEX | AGE | WORK |
|---|---|---|---|---|---|---|---|---|---|
| TRD01 | F | 13 | 7.5 | 4 | HS | 4 | 3.2 | 18 | 0 |
| TRD02 | F | 6 | 7 | 3 | HS | 5 | 3.2 | ? | 4 |
| TRD03 | M | 14 | 6 | 6 | HS | 4 | 3.4 | 19 | 0 |
| TRD04 | M | 23 | 1 | 5 | HS | 6 | 3.3 | 18 | 1 |
| TRD05 | M | 25 | 6 | 6 | HS | 5 | 3.4 | ? | 0 |
| TRD06 | M | 13 | 1 | 2 | HS | 3 | 3.2 | 18 | 0 |
| TRD07 | M | 13 | 3 | 7 | HS | 6 | 3.3 | ? | 2 |
| TRD08 | F | 8 | 6 | 4 | HS | 5 | 4.0 | ? | 2 |
| TRD09 | M | 27 | 3 | 6 | HS | 5 | 3.4 | 18 | 0 |
| TRD10 | M | 11 | 5 | 4 | HS | 3 | 2.9 | 18 | 0 |
| TRD11 | M | 15 | 0 | 0 | HS | 3 | 2.9 | ? | 0 |
| TRD12 | M | 19 | 3 | 7 | HS | 4 | 3.2 | 18 | 0 |
| TRD13 | M | 17 | 7 | 7 | HS | 7 | 3.0 | ? | 0 |
| TRD14 | F | 11 | 7 | 7 | HS | 5 | 3.3 | 19 | 2 |
| TRD15 | M | 14 | 5 | 3 | HS | 7 | 3.4 | ? | 1 |
| TRD16 | M | 13 | 7 | 7 | HS | 7 | 3.6 | ? | 0 |
| TRD17 | M | 12 | 7 | 7 | HS | 4 | 2.9 | 18 | 1 |
| TRD18 | M | 15 | 7 | 5 | HS | 4 | 3.5 | 25 | 2 |
| TRD19 | F | 14 | 7.5 | 4 | HS | 6 | 3.0 | ? | 2 |
| TRD20 | M | 19 | 7 | 7 | HS | 3 | 3.1 | ? | 2 |
| TRD21 | M | 15 | 7 | 7 | HS | 6 | 2.5 | 22 | 3 |
| TRD22 | M | 10 | 7.5 | 3 | HS | 6 | 3.4 | 17 | 0 |
| TRD23 | M | 14 | 4 | 3 | HS | 4 | 3.1 | 20 | 2 |
| TRD24 | M | 8 | 7.5 | 3 | HS | 5 | 3.1 | 19 | 2 |
| TRD25 | M | 11 | 4 | 3 | HS | 3 | 2.9 | ? | 2 |

Phys Code: HS = high school physics only;
HS & U = high school physics and a university physics course.

Math Codes: 3 = trigonometry; 4 = first quarter calculus:
5 = second quarter calculus; 6 = third quarter calculus;
7 = other.

Work Codes: 0 = none; 1 = 1-10 hours per week; 2 = 11-20 hours per week;
3 = 21-30 hours per week; 4 = 30 or more hours per week.

Table 3.9

Match Score for Cohorts

| Student | SEX | FCI | SAP | LP | PHYS | MATH | MPEX | AGE | WORK | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|
| 01 | * | 3 | 3 | 3 | 2 | 2 | 2 | | 1 | 16 |
| 02 | * | 3 | 3 | 3 | 2 | 2 | 2 | | 1 | 16 |
| 03 | * | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 17 |
| 04 | * | 3 | 3 | 3 | | 2 | | | 1 | 12 |
| 05 | * | 3 | 3 | 3 | 2 | 2 | 2 | | 1 | 16 |
| 06 | * | 3 | 3 | | 2 | 2 | 2 | 1 | | 13 |
| 07 | * | 3 | 3 | 3 | 2 | 2 | 2 | | | 15 |
| 08 | * | 3 | 3 | 3 | 2 | 2 | | | | 13 |
| 09 | * | 3 | | 3 | 2 | 2 | 2 | | 1 | 13 |
| 10 | * | 3 | 3 | 3 | 2 | | 2 | 1 | 1 | 15 |
| 11 | * | 3 | | 3 | 2 | | 2 | | | 10 |
| 12 | * | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 17 |
| 13 | * | 3 | 3 | 3 | 2 | 2 | 2 | | 1 | 16 |
| 14 | * | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 17 |
| 15 | * | 3 | 3 | 3 | 2 | 2 | 2 | | 1 | 16 |
| 16 | * | 3 | 3 | | 2 | 2 | 2 | | 1 | 13 |
| 17 | * | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 17 |
| 18 | * | 3 | 3 | 3 | 2 | | 2 | | | 13 |
| 19 | * | 3 | 3 | 3 | 2 | 2 | 2 | | 1 | 16 |
| 20 | * | 3 | 3 | 3 | 2 | 2 | 2 | | | 15 |
| 21 | * | 3 | 3 | 3 | 2 | 2 | 2 | | | 15 |
| 22 | * | 3 | 3 | | 2 | 2 | 2 | | 1 | 13 |
| 23 | * | 3 | | 3 | 2 | 2 | 2 | | | 12 |
| 24 | * | 3 | 3 | 3 | 2 | 2 | 2 | | 1 | 16 |
| 25 | * | 3 | 3 | | 2 | | 2 | | | 10 |