Development of instructional systems for teaching an electricity and magnetism course for engineers

M. J. Marr  
School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332

E. W. Thomas  
School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332

M. R. Benne and A. Thomas  
School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332

R. M. Hume  
OMED, Educational Services, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 8 September 1998; accepted 31 December 1998)

Techniques for the prediction, measurement, and improvement of student performance were examined in an introductory physics course required for engineering majors. The contributions of this study include (1) the application of a statistical technique for predicting performance, (2) a computer program for training basic problem-solving skills, and (3) evidence for the value of training with complex homework problems. The prediction of performance was calculated using the method of discriminant analysis and data from the student’s academic record. Specifically, this method predicted the chance of a satisfactory grade or a risk of failure using the student’s grade point average (GPA) at entry to the course and grades in certain preceding technical courses. The technique was successful in predicting outcome of the course for over 70% of the students and provided a baseline of anticipated performance against which the results of an intervention could be measured. Improvement of performance resulted from two intervention techniques that modified the students’ out-of-class assignments. The first intervention was Precision Teaching: a modification of homework exercises designed to improve basic skills in problem solving. Evaluation of this intervention indicated that class performance improved substantially; the number of students failing the course dropped to about one-half of that predicted by the discriminant analysis technique and overall class performance improved by almost one letter grade. The second intervention was based on the use of complex, multi-step homework problems designed to discourage “matching” of problems to operational formulae in the text. Evaluation of this intervention indicated that student performance on specific parts of the curriculum improved by 10%–20%. Both of these techniques resulted in significant improvement in performance largely on the final examinations. © 1999 American Association of Physics Teachers.

I. INTRODUCTION

Approximately 86,000 students enter engineering programs as freshmen each year. According to Astin6 only 47% actually graduate as engineers; most of the 53% who leave engineering will eventually graduate in some other discipline. At our own institution, Georgia Tech, the numbers are similar to the national statistics. We observe that most of the attrition occurs in the student’s first two years, prior to taking their major discipline courses. Attrition occurs when the students are studying pre-engineering mathematics and science courses.

The present article describes a programmatic effort to analyze student and instructor behavior in a particular physics course which is part of the pre-engineering science sequence. The goal of the effort was to enhance student performance. We chose to examine the second-quarter sophomore physics course in electricity and magnetism because this is the most difficult of the three-quarter physics sequence. Engineering majors must make a “C” in this course in order to progress to the courses of their major. In practice some 30% of all students make a grade of “D,” “F,” or “W” (withdraw).

Part of the difficulty of the course resides in erroneous notions about basic concepts such as voltage, current, and electrical energy brought into the course by students. Many other abstract concepts such as electric or magnetic fields, potential, charge distributions, capacitance, etc., are new to most students. While the previous course in mechanics deals largely with forces and moving objects accessible to everybody in everyday life, electricity and magnetism are phenomena generally accessible only through complex measuring devices. Thus establishing intuitions about such concepts is a difficult task when only traditional materials are available. As Arons3 points out: “…it is all too easy to lose sight of how highly abstract this part of physics really is and how frustratingly difficult it turns out to be for many students.” The poor performance in this course makes a significant contribution to the overall attrition from the engineering program. While our study was focused on only this single course, we hope the lessons learned are applicable to all parts of the introductory physics sequence, and beyond.

The efforts to be described reflect a five-year collaboration of physicists, psychologists, and computer scientists. The principal goal was to devise methods to enhance students’ performance measurably in electricity and magnetism. We recognized early that the conduct of a course and the students’ interaction with the material represented a complex system operating within the context of numerous constraints.
Particularly in a major research institution, factors such as class size, faculty time, teaching assistant contributions, use of a common text, available class time, and budget are all limiting. Yet such constraints reflect a common reality within which instructional innovations must generally be developed and evaluated.

Our approach was directed in two interconnected areas: (1) systematically evaluating student performance including determination of those variables that predict students at risk; and (2) the development of effective problem solving skills. The first is an essential task if one is to properly apply and evaluate instructional innovations. Any program of instruction, no matter how well meaning, elaborate, or seemingly valid, is suspect, if not useless, until one demonstrates clearly that it is effective under practical conditions. Evaluation of instructional interventions are difficult because of the multitude of relevant variables that can affect performance. Our techniques for assessing performance included a strong component related to prediction of performance. Prediction is very helpful in identifying student groups who might be targets for an intervention.

Our second focus was on the development of effective problem solving skills that are the basis for achieving success in physics. As will be subsequently detailed, problem solving in this domain is a very complex behavior that required task analyses to tease apart component skills. Training was then reconstructed to develop the appropriate component skills.

We first review the course design at our institution. We then discuss techniques for the prediction of performance. We specifically show that the technique of discriminant analysis is effective in 70% of all cases in identifying whether a student’s ultimate grade will be unsatisfactory or satisfactory. Finally, we describe two interventions for training problem solving which led to a reduction of unsatisfactory grades and to an overall improvement in class performance of approximately one letter grade.

II. THE COURSE DESIGN

The course at Georgia Tech on which the present studies were based is typical of most large engineering programs in the USA. All engineering majors at Georgia Tech are required to take three five-credit-hour courses in calculus-based physics. These courses cover the elements of classical mechanics, electricity and magnetism optics, and certain topics in modern physics. The courses are taught in large class sizes, typically 150 students, with four lecture periods per week and a one-credit-hour laboratory. The second course of the three courses devoted to introductory physics is mechanics, electricity and magnetism optics, and certain topics in modern physics. The courses are taught in large class sizes, typically 150 students, with four lecture periods per week and a one-credit-hour laboratory. The second course topic of the three courses devoted to introductory physics is the electricity and magnetism component. For this particular course we generally allot only three one-hour class periods per week to lectures by the faculty instructor and in the fourth lecture period the class is broken up into smaller groups (25–35 students) for instruction by graduate teaching assistants. These recitation sessions generally focus on review of homework and test problems. Of the 150 students per class, students majoring in the sciences are a relatively small fraction of the total. The class is essentially all engineering majors. For all of the engineering majors the physics courses are foundational and their successful completion is a prerequisite to entering and performing well in subsequent work. The topic of electricity and magnetism is obviously basic for electrical engineering students. The course is taught from well-established standard texts such as Serway or Halliday, Resnick, and Walker. These two texts have almost identical coverage, subject order, and level of difficulty.

This course had never undergone systematic evaluation, most likely because its format is so common among universities. However, it became apparent that at least two objectives must be met. The first objective was to determine what the predictors were of passing or failing the class. The second objective was to determine what intervention would promote passing the class. Predictors must be identified prior to intervention design and evaluation because intervention design and evaluation require predictor information.

III. PREDICTION OF STUDENT PERFORMANCE

A technique for predicting performance outcome may be used in a number of important ways in intervention design and evaluation. First, the method is a means of selecting matched groups of students for use as controls and intervention test cohorts. Alternatively, it may be used as a means of measuring the differences in such groups. Second, one can conduct a within-class assessment by predicting at entrance what the distribution of performance measures should be without intervention and then comparing with the actual distribution achieved. Such an in-class assessment may obviate the need for separate test and control groups. Third, the prediction of outcomes allows us to identify students at risk and possibly target interventions specifically to assist them. Students predicted to perform well might show little effect of an instructional intervention simply because of ceiling effects. Nevertheless, in assessing the effect of such an intervention on an entire class, these high-performing students may wash out any positive effects shown in those predicted to do poorly. Put in another way, if an instructional intervention is truly effective, it should shift the lower end of the distribution toward the upper end. Finally, the variables that enter into the prediction may include those that could, with advantage, be improved by an intervention before the student ever enters the class. An example of this might be improving success in a prerequisite math course.

We expected that the performance of a student in this course would be in some way related to the performance of the student on previous academic activities. For each student we had at hand a substantial body of records of academic achievement. Performance records exist on individual courses taken at the Institute as well as the cumulative result to date expressed through the GPA (grade point average). Generally, records of performance before entry also exist such as the SAT (Scholastic Aptitude Test) and high school GPA. We compared such records with each student’s performance in this class. We repeated this comparison for student groups from several years past and taught by a variety of instructors. Through a procedure known as discriminant analysis (see below) we were able to predict from these records, with 70% success, whether a student would make an acceptable grade (“A,” “B,” or “C”) or an unacceptable grade (“D,” “F,” or “W”). Only three variables in a student’s incoming record were significant in this prediction and all related to very recent performance at the Institute. We also examined whether the results of an easily administered test of learning style exhibited by students, such as the Myers–Briggs Indicator, might lead to an improvement in the prediction of outcome beyond the three academic variables.
A. Academic factors as predictors

Many statistical approaches are available to predict classroom performance. Probably the most often used measure is a simple Pearson product-moment correlation. While the correlation coefficient is well known and easily interpretable, it is less than ideal when used in the current context. The correlation coefficient is insufficient in that it provides little information about the relative effects of multiple predictors on a criterion, and offers no simple means of combining the effects of multiple measures into one single measure of predictive ability. One method often employed to deal with this issue is some form of multiple regression. This technique predicts performance based on some linear combination of predictors according to a model of the following type:

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n + b.$$  

In this case $Y$ is classroom performance, the $n$ $X_i$'s are predictors such as SAT scores and grades in related classes, the $a_i$'s are weights associated with each predictor, and $b$ is a constant. The key benefit of multiple regression over simple correlation is that the model provides a way to combine multiple predictors into a single estimate of performance. Likewise, the weights (if standardized) associated with each predictor account for the relative importance associated with each potential predictor. Moreover, it is possible to compute a measure, $r^2$, which gives a single estimate of the total predictive power of the entire set of predictors. Unfortunately, the interpretation of this measure has lead to some confusion among scientists, administrators, and educators.

Despite multiple regression's continued widespread use in educational interventions, there is a much more troublesome problem than properly interpreting an $r^2$. The model is designed and developed based on the fundamental assumption that both the predictors and the criterion are continuous in nature. However, most variables used in educational interventions are probably not best considered as continuous. Variables such as sex, race, and minority status are obviously not continuous. Even variables such as SAT and class size that might at first appear continuous are probably better characterized as discrete. At Georgia Tech SAT's are highly restricted so that SAT math scores, for example, may only range from about 600 to 800. Given the five-point intervals of the SAT, this results in only 41 distinct scores. Likewise, class size certainly does not approach infinity, and in all likelihood any effects of class size follow a step model instead of a linear one. On the performance side, almost all variables that we attempt to predict in educational interventions are categorical (i.e., class grade, number passing, etc.). Given that research has shown that the assumption of continuous data is not very robust, this assumption's continued use in educational interventions where categorical data are the rule is highly questionable. For this reason the current research turned to a more appropriate method of analysis that is not dependent on continuous data.

Discriminant analysis was developed explicitly for use with categorical dependent variables, and research has demonstrated that this procedure is not deleteriously affected by categorical independent variables, as is multiple regression. Instead of producing an estimated $Y$ on a continuous scale, discriminant analysis predicts categorical membership on the dependent variable. Although the mathematics involved is too involved for presentation here, the easiest way to conceptualize discriminant analysis is to assume for a moment that we know each student's category membership on the dependent variable. Then we could think of computing a mean on the independent variable(s) for each category of the criterion. For example, the category memberships may be comprised of students who achieved different letter grades in a course. The independent variables might be such factors as SAT, grades from previous courses, etc. The magnitude of difference between the means for the various categories of the dependent variable tells us how discriminating the independent variable is. Based on backward probabilities discriminant analysis simultaneously analyzes all of the mean differences for the potential predictors and determines which variables are the most discriminating. The key concept is breaking the criterion group into separate identifiable units (such as whether the person passed or failed). No assumption of continuity is required.

In practice, discriminant analysis can be used much like multiple regression. Given a list of potential predictors, discriminant analysis can determine which variables are the most effective in predicting performance. Just as in regression, the relationships among the predictors is taken into account. That is, the analysis looks at independent contributions of each variable in an equation and produces an equation with the most parsimonious collection of variables accounting for the greatest predictive power. One major advantage of discriminant analysis over regression is that its measure of predictive ability is in terms of a more recognizable unit, namely category membership on the dependent variable. By looking at actual group membership one can compute the percentage of correct classifications for the dependent variable. In fact, if several types of misclassifications occur one can go as far as to determine the number of each type of misclassification. This measure of percent correct classification seems inherently more meaningful than percent variance accounted for (i.e., $r^2$). Percent correct classification is more easily translatable to scientists, educators, and administrators who are not statisticians. Considering this and the fact that the dependent variables in the current studies were clearly categorical in nature, we decided to use discriminant analysis as our statistical model.

In order to predict student performance in electromagnetism we began by collecting historical data for one professor's electromagnetism class over a five-year period. These data included performance in the class as well as various potential predictors such as sex, race, gender, major, SATs, high school and college GPA, diagnostic measures, and grades in most mathematics and science classes taken at Georgia Tech. Data were obtained for over 1600 students. The goal was to be able to predict student performance using data available before students began their course in electromagnetism. A preliminary analysis showed that almost all academic variables had univariate relationships with performance in electromagnetism. While this fact is informative in its own right, it is not very useful in terms of prediction since these variables do not contribute unique variance. Discriminant analysis, such as regression, allows one to determine which variables make independent contributions in predicting a dependent variable. In the present work that dependent variable is the classification of whether the student is "at risk" of making an unsatisfactory grade (i.e., will make a letter grade of D or F in our context) or whether the student is "not at risk" of an unsatisfactory grade (i.e., will make a letter grade of A, B, or C in our context). The procedure first selects the variable which correlates most strongly with the course outcome, then selects from the remainder a second
variable which most improves the classification. This is repeated until inclusion of any of the remaining variables does not improve the accuracy of the classification. When $N$ variables are under consideration the procedure seeks to define a surface in $N$-dimensional space which best separates the student outcomes into the two classifications. Applying the procedure to the current data, it was found that three potential predictors made significant independent contributions: (1) overall GPA; (2) grade in the calculus course dealing with the differential calculus of functions and curvilinear motion; and (3) grade in the physics course previous to electromagnetism, particle dynamics. We can derive a discriminating score $D$ in terms of the three contributing variables of the form

$$D = -4.823 + 1.196(\text{overall } gpa) + 0.359 \times (\text{physics}) + 0.232 \times (\text{math}). \quad (1)$$

The value $D = 0$ defines the surface which best separates the students into the two classification. When the student’s record leads to a positive value of $D$, then that student is classified as being “not at risk” and when the score is negative the student is classified “at risk.” For example, a student with a “B” (numerical value of 3) in each of these three grade categories has a predicted discriminant score of $+0.538$ which classifies them as “not at risk.” A student with a “C” (numerical value of 2) grade in each of these three categories has a predicted discriminant score of $-1.249$ which classifies them as being “at risk.” Interestingly, a student with a “C” in each of these categories is considered by their major school as qualified to undertake the course but the analysis predicts that their result in the course is likely to be unsatisfactory. It is clear that overall GPA is the most important variable in arriving at a classification. The reader is cautioned that the numerical parameters of the equation are valid only for the experience of the program under study and that the parameters for other programs may be different.

The discriminant analysis procedure also examined other variables, including grades in other courses, high School GPA and SAT (Scholastic Aptitude Test) to determine whether their inclusion, in addition to the three parameters already identified, would significantly improve the fraction of students classified correctly. No other parameters provide a significant improvement. The criterion actually used in discriminant analysis is related to a detailed statistical analysis of significance. From a practical point of view the inclusion of any additional variables from that available data set changes the success of the classification by less than 1%.

The next question to be addressed was how successful these three variables were at predicting classroom performance. The discriminant equation was applied to the group under study to classify students into the categories of “at risk” (predicted grade of D or F) and “not at risk” (predicted grade of A, B, or C). Table I shows the results of this analysis. Some 48% (319/661) of those predicted to be “at risk” actually made a D or an F in the class, while 86% (828/961) of those predicted to be “not at risk” actually made a C or better in the class. It is not surprising that the majority of the cases “wrongly classified” have discriminant scores close to zero and therefore represent cases close to the surface of separation between the two classifications. While these numbers may not seem impressive, there are several important aspects to keep in mind. First, there were many more individuals predicted to pass than were predicted to fail. Because a higher predictive accuracy was obtained for those in the passing group the overall correct prediction is 71% (1147/1622) of all students. Second, this 71% is a significant ($p<0.05$) improvement over the base rate where the student’s performance is predicted only on the basis of the typical overall distribution, without any additional information. On average, about 70% of the class gets a “C” or better, leaving 30% to get a “D” or “F.” If one were simply to flip a weighted coin (70/30) to predict performance for each student on entering the course one would succeed, on average, 58% in predicting passing or failing, compared with the obtained 71%. In terms of the current sample that represents an increase of 206 correct predictions. Because each of these correct predictions represents a student who may or may not receive a potentially valuable intervention based on our analysis, the significance of increased accuracy is evident. Recall that such helpful prediction only required three measures readily available from the Registrar. Third, a significant number of the misclassifications that resulted occurred for “C” students. Given that the main goal is to predict students at risk the error of predicting someone at risk when they, in fact, make a C in the class had few associated costs. In fact, considering that the material in electromagnetism serves as a building block for the subsequent curriculum, marginally acceptable performance may be little better than failing the course. Moreover, a failing grade and the subsequent retake may provide the student with another opportunity to master the material. Based on these two facts and the overall predictive accuracy of over 70% the current endeavor appears reasonably successful. However, with any exploratory data analysis approach it is quite possible that the results of a particular study may capitalize on chance. It is thus necessary to cross-validate results obtained with the discriminant analysis on a second sample. A related concern was that the current results were based on only one professor’s classes. Perhaps instructors with different teaching styles or points of emphasis might necessitate a different set of variables being predictive of classroom performance. In fact, the predictive power of the previous analysis would be of little use if it were not generalizable to other instructors. In order to answer both of these questions another sample was obtained. The second sample consisted of four separate course sections taught by four different instructors. Data were obtained for a total of 534 students. The predictive parameters obtained from analysis of the first (larger) group

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>173</td>
</tr>
<tr>
<td>D</td>
<td>279</td>
</tr>
<tr>
<td>C</td>
<td>494</td>
</tr>
<tr>
<td>B</td>
<td>442</td>
</tr>
<tr>
<td>A</td>
<td>234</td>
</tr>
<tr>
<td>Total Students</td>
<td>1622</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actual Grade Awarded</th>
<th>Preliminary Classification of Outcome; Number of Students in Each Letter Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Classify at Risk (Predict F or D)</td>
</tr>
<tr>
<td></td>
<td>Classify Not at Risk (Predict C, B, or A)</td>
</tr>
<tr>
<td>D</td>
<td>175</td>
</tr>
<tr>
<td>C</td>
<td>203</td>
</tr>
<tr>
<td>B</td>
<td>114</td>
</tr>
<tr>
<td>A</td>
<td>25</td>
</tr>
</tbody>
</table>

Table I. Discriminant analysis prediction for the reference group of 1622 students showing the number of students awarded each letter grade at the end of the course. Also shown is the number of students in each letter grade category who were assigned predicted classifications of At Risk or Not At Risk.

792 Am. J. Phys., Vol. 67, No. 9, September 1999 Marr et al. 792
were used to predict performance of the students in the four new classes, and the predictions were compared with the results actually obtained. The results are presented in Table II. The results clearly demonstrate the robustness of the original discriminant analysis. Success in identifying students not at risk was 84% while success in identifying students at risk remained around 50%. Overall correct predictions were obtained for 70% of the students (373/534). Once again, the large number of the incorrect predictions was for "C" students. This analysis and that conducted in subsequent quarters with other samples showed that the results of the original analysis remained stable. Predictive accuracy of around 70% was found regardless of instructor, quarter, or year. Thus by using information about a student’s overall GPA, their performance in a specific calculus class and their grade in particle dynamics we were consistently able to predict whether or not they were at risk in the electromagnetism course. Furthermore, the consistency of the predictive power of these three variables allowed us to evaluate subsequent interventions.

B. Nonacademic factors as predictors

Academic variables contributed to the successful prediction of 70% of the students in the electricity and magnetism course, but the possibility existed that nonacademic factors could also contribute to students’ likelihood of passing or failing the course. Consider that at Georgia Tech the average SAT of entering freshmen exceeds 1300. From an academic perspective, virtually all of these students might be expected to do well in mathematics, science, and engineering, yet many do not succeed. Even if they have excellent grades, some abandon these fields for other pursuits. Changes in interest patterns, motivational issues, personality factors, attitudes, and learning style incompatibilities are just some of the possible reasons suggested for academic shifts and failures. Learning style in particular has received considerable attention as a possible predictor of general academic performance.13–18

We examined whether learning style could be used as a predictor of performance in the course under study. We selected the Myers–Briggs Type Indicator (MBTI) as the instrument for organizing the students’ information-seeking and decision-making preferences because it has been extensively tested for reliability17,18 and it has been used in other engineering programs.19,20 The MBTI is a questionnaire based on Jung’s theory of psychological types.21 The responses to the questionnaire are used to classify the respondent’s type as Extrovert or Introvert (E or I), Sensing or Intuitive (S or N), Thinking or Feeling (T or F), and Judging or Perceiving (J or P). No particular classification is superior to another, although in a given situation the preferences of one type might match the demands of a situation better than that of a different type.22 We administered the MBTI questionnaire to a large class in the electricity and magnetism course with the objective of determining whether students’ performance in this course could be related to their MBTI classification.

First, we compared those students who passed the class (marks of A, B, C) with those who failed or withdrew from the class (marks of D, F, W) according to their preference on each of the four scales. We found no significant difference between these groups (Table III). That is, the learning styles of the students who succeeded and the students who failed were not significantly different.

Second, we examined whether adding the students’ MBTI classification to the academic data used for discriminant analysis could improve the accuracy of that prediction. In this particular class the discriminant analysis that included only academic information correctly categorized 75% of the students on a pass or fail performance. A second discriminant analysis that included the MBTI categories did not improve that prediction; it remained at 75%. The conclusion was that prediction via the discriminant analysis technique remained dominated by academic factors: incoming GPA, previous particle dynamics physics grade, and previous calculus grade.

While these analyses have led to our conclusion that MBTI categories do not aid in the prediction of success or failure in this course, the class composition itself may be predictable. We organized our data according to the 16 personality types that result from all possible combinations of the four scales: INTJ, ENFP, ESTJ, etc. The results of the MBTI classification are presented in matrix form (Table IV). Take as an example the top left-hand box; this shows that 16.75% of the class is categorized as ISTJ (Introvert rather than Extrovert, Sensing rather than Intuitive, Thinking rather than Feeling, and Judging rather than Perceiving). It is remarkable how similar the distribution of Myers–Briggs types are between this electricity and magnetism class of 191 stu-
students and that shown by McCaulley et al. for 3784 freshman engineering students from eight different schools. McCaulley et al. found significant differences among the MBTI distribution of high school graduates, all college graduates, and of engineering freshmen. Thus the decision to attend college and the decision to study engineering may be related to the MBTI categories, but, as our study suggests, performance within the electricity and magnetism component of the pre-engineering program is not significantly related to the MBTI categories.

The MBTI did not organize nonacademic variables so they improved our prediction of success and failure, yet it must be the case that nonacademic variables contribute to shifting or failure of students. For example, a study of colleges and universities in Colorado found that students who left the engineering program were not academically different from those who stayed in engineering. In fact, investigators have found that such groups differ on their reasons for choosing engineering as a major and the extent to which they found the engineering coursework to be interesting relative to other fields. In these studies, those who left engineering found other fields more interesting. Somewhat bewildering, Besterfield-Sacre et al. found that the students who left the engineering program actually had a more favorable impression of the engineering profession and the subject matter than those who stayed in the program. Pickering and Calliotte also detected a significant relationship between nonacademic variables and retention, but the relationship was very complex, with many different factors accounting for a student’s likelihood to leave a program. Finally, Woods and Crowe revealed retention predictors that seemed to straddle nonacademic and academic issues. The predictors involved the students’ knowledge of their role in the university classroom. The investigators found students who dropped out were not aware of what was expected of them nor what to expect of their instructors. They were also unaware of the importance of good note taking. These fundamental issues contributed to students’ success or failure within the class. In summary, the evidence presented here indicates that complex nonacademic factors need to be identified and addressed. While the prediction of student performance can be improved, our use of discriminant analysis established sufficient information. The baseline information allowed us to compare actual outcomes to initial predictions after the introduction of interventions. As our efforts switched from prediction to intervention design and evaluation, we identified problem solving as the most critical skill for student success.

IV. INTERVENTION DESIGN AND EVALUATION: THE DEVELOPMENT OF PROBLEM SOLVING SKILLS

Intervention design should focus on those factors that determine success or risk of failure in a course. As just described, students entering this course with weaknesses in math and physics are most likely to fail. The commonality between these courses is problem-solving skill. Because problem solving is a skill, it can be trained; thus our intervention was designed with a training perspective. Our efforts included (1) skill analysis, (2) basic skill training for improved accuracy and speed, and (3) complex skill training for improved problem analysis and knowledge application. Both the basic skill training intervention and the complex skill training intervention were evaluated and found to be effective.
A. Skill analysis

Successful problem solving in physics, as in any other area, is a skill founded on extensive experience of a certain kind. The goal of an introductory course in physics is, however, not to generate sophisticated and truly expert problem solvers. Rather it is to have students learn the basic concepts so that these concepts might be applied in relatively simple situations and serve as entries into more advanced work. Problems serve both to teach principles and to test one’s understanding of those principles. Effective problem solving reinforces contact with the material and thus helps maintain academic performance when more difficult material is encountered. Fancy and exciting demonstrations, and more colorful, contemporary texts may have their place in maintaining student interest, but as emphasized by McDermott, effective acquisition of physics skills will only emerge through active participation in doing physics, not simply looking at it or listening to it.

1. A model of skill analysis

The most frequent practical objective of a course is that the students demonstrate successful problem solving at an intermediate level as defined by acceptable quiz and exam performance using a standard set of problems. Relative to their skills at entry to the course, they acquire some expertise in the solving of problems. During the last 25 years there has been a growing focus on the nature of expertise. This research has identified some significant differences between expert and novice performers and these differences can be used to guide the development of training and educational programs. A major concern of this work has been to characterize the different skill repertoires that people have within a domain and the roles that these different types of knowledge play in solving problems.

In general, the knowledge and skills possessed by experts can be characterized by three types of information: declarative, procedural, and conceptual. Declarative knowledge refers to facts at hand; for example, a picometer is 10^-12 m. Procedural knowledge, on the other hand, refers to algorithms or rules for conducting a procedure, for example, the rule for calculating a vector cross product. Conceptual knowledge refers to the knowledge about the interrelationships among problems, how they are similar or different. This knowledge reflects an understanding of principles that may be applied in a number of different contexts.

According to some models, skill acquisition in any domain goes through a series of stages. First, one acquires declarative and basic procedural knowledge about a domain. This knowledge is comprised of facts and of simple rules or algorithms, which are treated as facts. With practice, the rules become “proceduralized” and “compiled,” a process called automaticity. Once the procedural knowledge has become automatic, it is rapidly accessed from memory and requires relatively little cognitive effort. As people learn more about a domain they begin to acquire conceptual knowledge. Conceptual knowledge refers to higher levels of cognitive functioning within a domain. For example, one initially develops knowledge about specific algorithms or rules that can be used to solve specific problems. After exposure to a large number and variety of problems, students begin to develop a conceptual repertoire of problems sometimes referred to as a “schema.” This repertoire comprises knowledge of the domain that is less dependent on specific examples or problems and instead is organized around relationships within the domain (see, e.g., Ross). The advantage of this repertoire is that it promotes transfer of performance from specific learned instances to novel problems.

To summarize this work, experts and novices differ in three general ways. First, experts have more declarative knowledge about a domain. This includes terminology, and a large number of facts. Second, experts have more procedural knowledge. That is, they are more skilled with rules and algorithms that can be quickly and easily applied to solve a variety of problems. Finally experts have a greater conceptual understanding of the domain. This type of knowledge promotes flexibility and creativity in problem solving. Our approach to training was to seek to address all these differences that exist between novices and experts. We applied this model of skill development to our course and attempted to determine how our students were weak in gaining one or more of these three types of information.

2. An analysis of student’s problem-solving skill

Ideally problems should be solved by first stating the basic laws and facts (i.e., declarative information) followed by manipulation of such information in the form of symbols to yield an equation (i.e., procedural knowledge) from which a numerical answer may be determined. We refer to this final equation as the “operational” equation. The substitution of appropriate numbers into this equation gives the correct answer to the problem.

We found students do not solve problems using such a process. We conducted interviews with students and found that many students look at an assigned problem, then search the relevant chapter for a similar example. They pull the operational equation from the example and use numbers from their assigned problem to derive an answer. Interviews also revealed that their strategy for test preparation is to commit as many of these operational equations to memory as possible. Thus the student treats the operational equation as declarative knowledge—that is to say, as a piece of information to be remembered. At the declarative level, however, these routines typically lack any referents, and certainly cannot be generalized for application in a variety of situations.

Further study confirmed the reliance of students on recall of operational equations. We compared students’ solutions on quizzes and exams to the “ideal” solution. The types of problems used in the study were familiar to the students. They had been seen as worked examples in the text; they had been assigned in similar form as homework; and they were completed in a quiz situation. Thus these were very standard problem types on which the student had experience and should have quickly understood the word descriptions.

For students who ultimately acquired a course grade of “A,” only 20% provided an adequately labeled diagram of the problem and only 50% provided some indication of the fundamental law or definition involved with the problem. However, over 70% of these “A” students provided the correct operational equation and all of these gained the correct answer to the problem. Among students who ultimately gained a “C” in the course, the frequency of providing adequate figures was 10% with only 30% correctly quoting fundamental information. Yet 65% quoted the correct operational equation and thereby acquired the correct answer. What about students who did not acquire a correct answer to a problem? Among students who eventually earned an “A” or “B” in the course, fully 20% of them started by writing an operational equation, which was incorrect! These analyses
confirmed the results of student interviews. Many students start a problem by quoting an operational formula, but with no statement of basic principles or even a figure. The ultimate success of the students depended on whether they remembered the appropriate formula for the scenario of the problem and whether they accurately performed the subsequent mathematical manipulations.

By conducting a skill analysis on problem solving, we recognized two major flaws in the existing training interventions. First, students were learning through their homework to perform problems by “matching” them to solutions or operational equations from the text. In a test situation, the students have no text to use for reference and the success of their solutions is poor. Second, students were also reinforced for memorizing operational equations for exams. The reason was that the students were required to work several problems on an exam very quickly (e.g., five in less than 50 min), but were not trained to work quickly. They discovered that recall of operational equations increased their chances of obtaining an answer (albeit sometimes wrong) for all of the problems on the exam. We determined student performance could be improved if explicit training focused on learning the three types of knowledge required for skill acquisition and if students were trained on speed in addition to accuracy.

B. Basic skills development through precision teaching

The analysis of skill acquisition has provided a number of potentially helpful approaches to intervention design. As mentioned previously, one of the more salient features of a skilled performance is the character of automaticity. At least certain components of a higher-level skill are executed rapidly and without apparent effort. Automaticity emerges from extensive practice under “consistent” conditions, that is, where similar inputs call for similar outputs. The operations of ordinary arithmetic are a paradigm case. Multiplying eight times nine always gives the same result, and we acquire an extensive repertoire of exemplars of these kinds of operations, so that when given a certain combination, we simply emit the correct response immediately and effortlessly. We can then be said to be “fluent” in basic arithmetic. While fluency in a skill may be shaped through an extensive history, this does not mean that, for a given performance, it is necessary to impose thousands upon thousands of trials for the performance to be acquired. Fluency is typically thought to emerge from simple drill. Drill is actually basic to the acquisition of any significant skill, but in addition to being dull, by itself it is inefficient. The basic problem is that, with simple drill, fluency is not explicitly shaped. Through a set of procedures called precision teaching, both accuracy and rate of execution are reinforced with the explicit goal of so many correct items per unit time.

Precision teaching, a concept pioneered by Lindsey and based on the work of B. F. Skinner, is a technique for the enhancement of basic skills and achieving fluency. Students undertake a large group of relatively simple problems and their success is measured by the number of correct responses in a set time interval, that is, rate of corrects. Fluency of correct responding, or the rate of correct responses, is a critical feature of the intervention. With repeated exposure to different problem sets, the rate of correct responses typically increases significantly. The student should actively record their rates. Research shows that basic skills are enhanced, performance on standard courses is improved, and the improvement is retained for extended periods. Fluency gained through precision teaching has been shown to enhance attentional behavior, recall of subject material, and performance on associated, but more complex skills.

Precision teaching is a proven methodology that has been shown effective in various educational contexts. The most frequently cited example is the Great Falls PT Project where investigators found that 79% of precision taught groups performed better on post-test measures than did comparable control groups receiving traditional instruction. In comparison with state averages on the Iowa Basic Skills Test, students who received precision teaching out-performed others in the state by 40 percentile points in math and 20 percentile points in reading. Moreover, Beneke has shown that precision teaching can be successful at the college level. With an intervention that lasted less than 5 min a day, Beneke was able to increase his students’ reading rate by 49% with a concomitant 75% increase in recall of material.

Fluency of a basic skill allows for the development of more elaborate performances, in part, because the latter do not require the additional efforts of the more basic skills. Thus higher levels of performance on more complex tasks are founded upon lower level, and more fluent repertoires. With respect to electricity and magnetism, what this means is that if the student has not reached a certain level of fluency in skills, such as algebraic manipulation or vector analysis, then considerable time and misdirected energies may be devoted to these basic operations and leaving little or nothing for dealing with the problem at hand. For example, a physical principle expressed as the integral of a dot product of vectors cannot be understood without being fluent in the mathematical structure of the principle. No application is possible even in the simplest situations. Moreover, in general, students do not prepare themselves to execute problems under contingencies similar to those of an in-class test or exam. A test typically involves five problems that must be completed within less than 50 min. If these problems are of reasonable difficulty, then there is little time to spend pondering what the integral of the sine function is or how to find components of a vector.

Our precision teaching methods targeted those areas of difficulty as identified by a task analysis of the skill requirements for solving standard problems in the text and on quizzes and exams. The specific types of items developed targeted all three components of expertise, that is, declarative, procedural, and conceptual knowledge. The program consisted of four components: units, mathematical operations, intuitive or conceptual problems, and one- or two-step basic physics problems.

1. Design of precision teaching exercises

Our exercise modules each consisted of four components. The first component tested the student’s recall of units as reflected in the particular topic under study. For example, under the topic of magnetism, one of the simpler items might be $T \frac{m^2}{A}$ where the correct response would be “Henry.” The second component was a set of routine mathematical problems related to the mathematical operations to be employed in the topic. Examples would be vector products and vector sums for a topic area involving magnetic fields. Third was a set of multiple-step, noncalculational, conceptual problems (we called these “intuitive” problems). For example, the student had to identify the directions of the $E$ field at various points in space around a pair of charges.
Each of these first three components had to be performed within 5 min. The third and largest component was a set of simple one- or two-step physics problems. Each of these involved a simple word description leading to the numerical manipulation based on a fundamental law or equation that the student ought to remember. An example might be to calculate the force between two charges separated by a distance. This problem was followed by further applications of the same law such as calculation of distance when force and charge was given and calculation of charge when force and distance was given. These physics problems, generally 10–15 in number, were to be performed within 20 min. The number of problems and their difficulty were chosen so that an “expert” (in practice, a faculty member or a graduate teaching assistant) could correctly perform all the problems within two-thirds of the time allotted to the student. Five similar modules were written for each chapter of the course. The course curriculum was based on treating one chapter per week so that the five modules also represented a week of exercises.

The students were instructed to perform all five of the modules on five successive days during the week allotted to cover the relevant physics material. In a paper-and-pencil version of the program, students were told to cease work on a section at the end of the allocated time and move to the next part of the module. The number of correct responses in the allocated time period was recorded and this figure tabulated as each module was performed. The intent of the tabulation procedure was that the student could observe his or her own improvement with repeated performance of the exercise.

We used the material in a number of delivery modes that differed primarily in the way that we sought to insure compliance with the rules. In all the modes described below, four of the modules were to be performed as “homework” on the students own time and the fifth was to be performed under supervision. The intent of the fifth supervised session was to maintain contact with the students and to provide assistance with problem types that gave the students difficulty. By comparing the scores attained in this supervised session with scores attained on unsupervised work “at-home” we could also monitor whether the “home work” was being performed in an honest fashion.

### 2. Implementation and evaluation of precision teaching in classes

For one set of experiments, participation in the precision teaching program was made mandatory and a small part of the class grade based on performance of the work. Materials were handled in two different ways. One mode was to prepare all the material as paper copies, instruct the students to perform four of the modules in their own time, and to self-grade (answers being handed out). The fifth module was performed under supervision and graded by a graduate teaching assistant as an audit of performance. The disadvantage of the “take-home” delivery in the form of paper copies is that there was no guarantee of student compliance with the rules. Interviews with students revealed that a significant number (58%) did not properly follow instructions to perform the modules on a daily basis; rather there was a tendency to do all the modules on the day before the class audit.

A second mode was to administer the materials on computer clusters where the machine recorded the day of performance, restricted the available time to the intended period, and recorded the number of correct responses. The four preliminary modules were performed at the student’s convenience and the fifth was again performed, on computer platforms, under supervision. Use of the computer delivery mode forced compliance with the rules regarding frequency of access and time limitations. Student performance in the course was tested with standard quizzes based on problems of a type found at the end of the chapter of standard texts. Comparisons were made with a control group of students who did not participate in the precision teaching exercises, but instead were given a 1-h tutorial program based on standard text problems.

To compare performance we calculated the mean quiz scores for all students in a particular range of incoming GPA. This study was conducted prior to our discriminant analysis procedure, but we had previously shown that GPA was an effective predictor of performance in this class. Figure 1 shows performance on the final examination (scored out of 200 points) for students in each of five GPA intervals (1.5–2.0, 2.1–2.5, etc.). The cohort undertaking computer-delivered precision teaching scored higher than the other cohorts did for four of the five GPA intervals. The exception was the highest incoming GPA interval (3.6–4.0) which we would presume to represent the most able students. In this category the students with a conventional form of instruction exhibited the highest performance.

For students who undertook the precision teaching on paper, where compliance with instructions was not insured, the final exam performance was generally higher than that of students taught on a conventional basis, but the improvement was not quite as high as under the computer-delivered regime. We note that in the lowest GPA sector (1.5–2.0) student performance under the paper-and-pencil regime was poorest of all the groups; we suspect that these students had poor study discipline. We carried out the same comparisons with the periodic tests performed during the course of the quarter. Figure 2 shows the results (for a total of three tests added together and with a maximum possible score of 150). On performance as measured by the periodic tests the three groups were not significantly different.

We conclude that the precision teaching activity results in a substantial improvement in long-term performance for most students in the class as reflected in the final examina-
tion scores, with the greatest benefit derived from regular work on the activity. Precision teaching had no significant impact on the shorter-term performance as measured by periodic tests. Students entering the class with very high GPAs, expected to be the best performers, gained the least from these exercises. This conclusion is in accordance with our earlier findings when testing a preliminary version of the precision teaching program based only on the physics problem component. In that version there were no units, mathematics and intuitive components to the modules.

In a further test we asked students to volunteer for participation in the precision teaching with no inducement other than the instructor’s word that their average performance was likely to be improved. Approximately 60% of students initially volunteered to participate in the training (n = 80 out of 140). During the quarter approximately 50% of the student volunteers quit attending the group session and discontinued work on the materials at home. The analyses reported are based on the quiz means for the students who completed the materials for the weeks preceding each quiz. For the course grades, any student who completed 60% of the materials was considered as part of the precision teaching group. We compared performance of those students considered to be participating in the precision teaching program with those students in the same class who did not volunteer for the program.

We checked the overall grade point averages and grade in the previous physics class (particle dynamics) and ensured that there was no significant difference between the two groups (those who were in the program and those who had not volunteered). We first analyzed performance for the two groups for each of the five class quizzes. In contrast with our earlier study, as shown in Fig. 3, the students in the precision teaching group (mean = 86.64) performed better on all five quizzes than did the nonprecision teaching students (mean = 75.40). Individual analyses of course grade distributions revealed that the difference was significant for each quiz (p < 0.05). We also analyzed course grade distributions for the two groups. Course grades were determined on a basis of quiz averages (for the best four out of the five quizzes), laboratory grade, and final exam. Figure 4 shows a higher proportion of students in the precision teaching group made course grades of A’s and B’s and fewer of the precision teaching students made C’s, D’s, or F’s. Based on a chi-squared analysis, these differences were significant at the p < 0.05 level.

Interviews with students who had dropped from the precision teaching program showed that often they felt that their final grade (whether good or bad) had become predictable and that further participation in the program would make no difference. We sampled student opinions about the precision teaching program in the quarter where it was voluntary. Some 58% of respondents (n = 61) stated that they had not properly followed the instructions on timing and, by exactly the same figure, stated that they would have taken the program instructions more seriously if there had been a weekly supervised test (graded for credit) based on the modules. Students found the required performance criteria (problems
correct in the allocated time interval) to be reasonable and that the physics problem component was the most valuable (3.8 on a basis of 5.0) and the math least valuable (2.6 on a basis of 5.0).

The overall results of the precision teaching intervention show that it can make a substantial difference to class performance if it is properly used. Performance improves among almost all students and is most marked on the final examination. The great difficulty is to persuade the students to undertake the work and to fully comply with the guidelines. Precision teaching was an intervention designed to improve student performance in the course. Ideally one would hope that students would see this as a useful tool for improving grades and would undertake the work voluntarily and with enthusiasm. The reader will not be surprised that this is generally a false assumption. As with all instructional interventions that call for intensive and active learning on the part of the student outside of class, contingencies must be established to provide clear and significant incentives for their appropriate use.

The Precision Teaching program along with a system for management of scoring and timing is available from the authors as a CD or can be downloaded from a web site (http://blackbox.psych.gatech.edu).

C. Skill development through complex problem exercises

Precision teaching addressed primarily the development of fluency in underlying skills. The exercises themselves do not represent the level of detail and complexity which one expects the student to master. The acquisition of skills in handling more elaborate problems may be developed further through the assignment of problems where the student works at their own pace and, if sufficiently diligent, develops the required level of understanding.

Inappropriate problem choice, design, and sequencing can be a major impediment to student success in a physics course. Our assessment of instructor behavior indicates that all too often little thought goes into selection of problems; or, equivalently, bad problems are chosen because they are “interesting” to the instructor. The most interesting problems for the instructor may not be the most appropriate for the student who is struggling to grasp basic concepts. Little effort may be made by instructors toward understanding: (1) the declarative and procedural knowledge of an assigned problem, (2) how basic principles are being tested and integrated, or (3) what might be a proper sequence of problems for training declarative, procedural and conceptual knowledge. We suspect that most instructors assign problems sets that are chosen from the many examples supplied by the textbook author at the end of the chapter and that their choice is made without sufficient attention to learning objectives.

1. Design of complex problem exercises

Problem choice should, first of all, be based primarily on course objectives. For example, two objectives should be that students could extend principles to elementary designs of functional devices and be able to interpret common physical situations. Problem choice should also focus on the development of the declarative, procedural, and conceptual skills mentioned earlier. These goals require that problems be assigned in sufficient numbers for extensive practice and possess a significant range of difficulty. Moreover, the problems need to train not only mathematical skills applied to basic laws, but conceptual, qualitative, integrative, and intuitive skills as well. While such skills may seem complex, they are of the same level of complexity typically required on quizzes and exams. These objectives were the basis of the second major approach to performance enhancement.

In an attempt to improve overall student performance, we designed a specialized set of homework problems to exercise complex skills. These homework problems required students to understand and use principles effectively because the problems could not be solved merely by the recall of an operational formula or by the matching of homework assignment to in-text worked examples. Each of the problems was couched in an applied scenario to make them attractive to the students, who were overwhelmingly engineering majors. The problems involved a series of steps where fundamental concepts were applied successively. In some cases we used problems where the student had to estimate an answer for a real-life situation where conditions were not exactly those of the idealized examples found within the text. We wanted to promote the notion that relatively simple physical principles can lead to substantial understanding of a complex applied situation, and thus encourage students to see the relevance of his or her physics course.

We can best illustrate the approach through examples of problems used. First consider deflection of a moving charged particle in an electrostatic field. This can be formulated as in terms of the operation of an electrostatically deflected CRT or as an ink jet printer configuration. The standard in-text scenario is to calculate the deflection of the charged particle at the exit from a parallel plate electrostatic field region. The problem is often worked out in the text and a similar problem is available in the end-of-the-chapter problem sets from which homework is conventionally assigned. Analysis of student responses, both in reviewing homework and in reviewing quiz answers, showed that most students simply quoted a formula like $1/2(Eq/m)(x/v)^2$ and substituted numbers. To replace this procedure we posed a problem that represents a fairly complete attempt to design a CRT where the required deflection on a distant screen and the energy of impact on the screen is specified and geometry given. The student’s task was to calculate the necessary fields in the electron gun and the deflection plate system. The question required the student to consider acceleration parallel to an electric field (in the electron gun), perpendicular to the electric field (in the deflection plate region), and then the subsequent motion in a field-free region before impact. This problem was later revisited with a similar question where the deflection is by a magnetic field (a problem which is not amenable to a precise algebraic solution). The student was thus informed as to the reasons why magnetic deflection is invariably used in wide screen CRT displays. A second example was to have the student consider designing a heart pacemaker with a simple LC circuit as the mechanism for providing the clock frequency. The student finds that the size of the components is too large to fit conveniently into the human body. A third example was to calculate the power loss and voltage drops in a design of the city’s subway system assuming power is supplied only to one end of the “live rail” and when the train is at the far end of the track system. This particular question set the student into seeking out information on the length of their local subway system tracks and worrying about the area and shape of the live rail.

The homework assignments related to a particular text chapter consisted of about five standard end-of-the-chapter
problems plus two or three extended multi-step problems of the kind just described. The end-of-the-chapter problems were retained to give the student experience manipulating data in simple scenarios. Homework was graded with about half the points being given for the applied multi-step problem. Assignment of complicated multi-step problems did have some impact on the instructional support given to the class. Student demand for assistance from staff in a voluntary tutorial or ‘‘help’’ session increased markedly with most of the questions being related to the applied multi-step problems. A further complexity was that the recitation instructors were asked to spend a large part of the recitation period time in the explanation of these problems.

2. Implementation and evaluation of complex problem exercises

To test the impact of this revised homework regime we compared students’ responses on quiz questions with responses from another class taught by the same instructor and from the same text, but where the homework assignments were entirely drawn from the standard text. The quiz questions were similar to the text problems. Thus students were, in part, practicing their skills on complex multi-step problems, but they were tested on their ability to solve standard text problems.

In Table V we show the performance comparison between the two student cohorts. One group practiced only on end-of-the-chapter problems; the other practiced on a combination of these problems and the applied multi-step complex problems. We have, for convenience, shown results on individual problem types. For most quiz problems the scores by students trained on the applied multi-step problems was substantially higher than those scores for students trained on text problems only. The performance improvement averaged 15%. Note, for example, the performance on an electrostatic deflection problem to calculate deflection at the exit from a pair of deflection plates. Sixty percent of the students trained only by working this type of problem worked it correctly, while 90% of the students trained on a more complex variant (calculation of the field required to get a particular deflection on a distant screen) worked it correctly; a quite significant increase. We also made the same types of comparisons on periodic quizzes given during the quarter. In this case the scores of the two cohorts were essentially the same. Apparently the improvement is in the longer term, just as the precision teaching results indicated. Performance as measured on a final exam was significantly enhanced; there was no significant improvement in short-term performance.

We also examined whether the beneficial result of this intervention was related to overall abilities of the student. Our discriminant analysis technique (see Sec. III(A) above) showed that the most important factor in prediction of performance was the entering GPA of the student. This variable we took as a simple measure of potential performance. Figure 5 shows the student scores on the final examination as a function of their incoming GPA. The cohort trained on applied multi-step problems had a mean performance superior to the standard group at all student GPAs. Thus the intervention is of value to all students in the class and does not favor one particular ability group or another.

Finally, we examined the overall performance of the class in terms of final letter grade using the discriminant analysis technique. For simplicity we normalized these data to 100 students. Analysis of the class incoming statistical profile indicated that 54 students were ‘‘strong’’ students who we would predict should get a satisfactory grade of ‘‘A,’’ ‘‘B,’’ or ‘‘C’’ and 46 students were ‘‘weak’’ and in danger of attaining a grade of ‘‘D,’’ ‘‘F,’’ or ‘‘W.’’ Bear in mind that these predictions were tentative and not all students would have some impact on the instructional support given to the class. Student demand for assistance from staff in a voluntary tutorial or ‘‘help’’ session increased markedly with most of the questions being related to the applied multi-step problems. A further complexity was that the recitation instructors were asked to spend a large part of the recitation period time in the explanation of these problems.

<table>
<thead>
<tr>
<th>Topic</th>
<th>% Correct ‘‘Simple training’’</th>
<th>% Correct ‘‘Complex training’’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric field</td>
<td>78</td>
<td>89</td>
</tr>
<tr>
<td>Potential 2 charges</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>Electron Acceleration</td>
<td>70</td>
<td>65</td>
</tr>
<tr>
<td>Deflection E field</td>
<td>60</td>
<td>90</td>
</tr>
<tr>
<td>dc circuit</td>
<td>60</td>
<td>90</td>
</tr>
<tr>
<td>dc circuit qualitative</td>
<td>47</td>
<td>62</td>
</tr>
<tr>
<td>Deflection B field</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>B field + wire</td>
<td>32</td>
<td>74</td>
</tr>
<tr>
<td>B field 2 wires</td>
<td>81</td>
<td>83</td>
</tr>
<tr>
<td>Induced emf</td>
<td>18</td>
<td>35</td>
</tr>
<tr>
<td>LR circuit</td>
<td>78</td>
<td>72</td>
</tr>
<tr>
<td>ac circuit</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>54</td>
<td>69</td>
</tr>
</tbody>
</table>

Table VI. Predicted classifications and observed course outcomes for students working the complex homework problems compared with those working the simple (back of the chapter) problems. The data are normalized to 100 students in both cases. The classification using discriminant analysis predicted that 46 students would attain a ‘‘D’’ or less and 54 students would attain a ‘‘C’’ or better.

<p>| Preliminary Classification of Outcome; Number of Students in Each Letter Grade |
|---------------------------|-------------------|-------------------|</p>
<table>
<thead>
<tr>
<th>Actual Outcome</th>
<th>Classify at Risk (Predict F or D)</th>
<th>Classify Not at Risk (Predict C, B, or A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>With ‘‘Complex’’ Homework</td>
<td>Inadequate</td>
<td>15</td>
</tr>
<tr>
<td>Adequate</td>
<td>31</td>
<td>52</td>
</tr>
<tr>
<td>With ‘‘Simple’’ Homework</td>
<td>Inadequate</td>
<td>23</td>
</tr>
<tr>
<td>Adequate</td>
<td>23</td>
<td>43</td>
</tr>
</tbody>
</table>
attend the result predicted. As indicated in Table VI for a
class taught using the standard text problem sets the 46
“weak” students are, based on experience, likely to divide
up into 23 who actually fail to get a satisfactory grade and 23
who will actually attain a satisfactory grade. For the class
taught using the applied multi-step homework problems the
number of weak students actually attaining a satisfactory
grade increased from the anticipated 23 to 31; a statistically
significant change. Among the students classified as
“strong,” the number who in fact attained an unsatisfactory
grade was reduced from the anticipated 11 under a standard
regime to 2 for the class taught with applied multi-step prob-
lems. Again this indicated that the improvement in class per-
formance occurs at all levels, among strong as well as among
the weak students.

The general conclusion is that the use of complex multi-
step homework problems does give rise to a significant in-
crease in student performance. We conclude that homework
assignments chosen from the end-of-the-chapter problem
sets are not optimal for the development of student perform-
ance. We suggest the reason is that textbook problems can
often be solved by “matching” them to worked examples
and that use of such problems does not train the student to
understand or manipulate the material, but rather trains the
student to match specific problems to possible specific solu-
tion strategies.

We recommend that homework assignments include a sig-
nificant number of multi-step problems. These problems will
need to be created by the instructor. We find that the subject
matter needs to be periodically refreshed in order to avoid
students copying solutions from classes in previous quarters.
Copies of the problem sets used in these tests can be ob-
tained from the authors.

V. CONCLUSIONS

A common anecdotal view of basic courses such as chem-
istry, calculus, and physics is that they should serve the role
of “filters” in allowing only the most dedicated and talented
students through the system into their major courses and be-
ond to graduation. This has led to a “boot camp” tradition
that has contributed to many students’ disillusionment with
pursuing careers in science and technology. Poor student per-
formance tends to be attributed solely to nonremediable ex-
ternal forces, for example, student incompetence, lack of
preparation, laziness, no motivation, etc., without question-
ning present methods of instruction, or devising contingencies
to deal with ineffective academic performances, whatever
their source. This research effort has attempted to address
what we saw as the inadequate performance of basically
good students in a core course in science and engineering
education. Our approach was two-fold: understanding what
variables at input predicted final performance and designing
training interventions to improve the sort of complex prob-
lem solving essential to success in the course.

The discriminant analysis technique is an effective method
in classifying whether students are at risk of failing to make
an adequate grade or are likely to succeed. In our samples,
success in prediction was about 70%. The study shows that
the principal indicator of likely performance in the electricity
and magnetism class is the overall GPA that the student has
acquired at the point of entry. Of lesser weight, but still
significant, is the performance in the calculus class and the
physics class that immediately preceded the class under

study. It is not surprising that these three indicators all rep-
resent immediate past performance and are biased towards
technical classes that are closely related to the course that is
the subject of our study. Performance indicators such as SAT
and high school GPA which represent work a considerable
time before entry to the class were not significantly useful in
prediction. Learning styles as measured by the Myers–
Briggs Inventory did not contribute significantly to predict-
ing course performance.

The technique of Precision Teaching was an effective
method for improving the basic skills of students and its use
resulted in a performance improvement, over the whole
class, of close to a letter grade. The use of complex multi-
step problems as part of the student homework regime also
resulted in a performance improvement of close to one letter
grade over the whole class. The improved performance in
both cases was manifested in reduced rates of failure as well
as improved grades overall. The improvement, which was
largely manifested on the final exam, was found across a
wide spectrum of student abilities. The impact of these inter-
ventions was primarily on the long-term performance of the
students.

ACKNOWLEDGMENTS

Supported in part by NSF Grant No. DUE-9455470,
SUCCEED Engineering Coalition, CEISMC, and EduTech.

1 P. Lashley, “Enrollments ’96: Inching Downward,” ASEE Prism, Sep-
pp. 27–30.
4 D. D. Carlsen and T. Andre, “Use of a Microcomputer Simulation and
Conceptual Change Text to Overcome Student Preconceptions About
Electrical Circuits,” Journal of Computer-Based Instruction 19, 105–109
5 R. A. Serway, Physics for Scientists and Engineers (Saunders, Philadel-
6 R. R. Schmeck, F. Ribich, and N. Ramanaiah, “Development of a self-
7 L. V. Hedges, “The Meta-Analysis of Test Validity Studies: Some New
Approaches,” in Test Validity, ed. by H. Wainer and H. I. Braun
8 M. S. Lewis-Beck, Applied Regression: An Introduction (Sage, Newbury
Park, CA, 1980).
10 R. D. Cook and S. Weisberg, Residuals and Influence in Regression
Assoc. 1399–1422 (1968).
12 F. M. Lord and M. R. Novick, Statistical Theories of Mental Test Scores
(Addison-Wesley, Reading, MA, 1968).
13 R. R. Schmeck, F. Ribich, and N. Ramanaiah, “Development of a self-
report inventory for assessing individual differences in learning pro-
14 B. G. Ferrell, “Factor analytic comparison of four learning-styles instru-
differences in learning strategies research,” Individual Differences in Cog-
16 J. M. Cornwell and P. A. Manfredo, “Kolb’s learning style theory revis-
17 I. B. Myers and M. H. McCaulley, A Guide to the Development and Use of
the Myers Briggs Type Indicator (Consulting Psychologist’s, Palo Alto,
CA, 1987).
18 D. A. Johnson, “Test–retest reliabilities of the Myers–Briggs type indi-
19 M. H. McCaulley, G. P. MacDaid, and R. Walsh, “Myers–Briggs type


21 C. G. Jung, Psychological Types (Harcourt, Brace, New York, 1926).


THE ANOMALOUS ZEEMAN EFFECT

This was the beginning of [Pauli’s] occupation with the irregularities induced by a magnetic field in the atomic spectra, the so-called anomalous Zeeman effect. Pauli comments on this activity as follows: ‘‘A colleague who met me strolling rather aimlessly in the beautiful streets of Copenhagen said to me in a friendly manner, ‘you look very unhappy’; whereupon I answered fiercely, ‘How can one look happy when he is thinking about the anomalous Zeeman effect?’ ’’