1

Gifted Students' Individual Differences in Distance-Learning Computer-Based Physics

With Eric Cope, Education Program for Gifted Youth, Stanford University

1.1 Introduction

The student performance results from the EPGY physics courses presented in this paper extend the results of the companion papers by Stillinger and Suppes (1994) and Cope and Suppes (1999) on the EPGY algebra and calculus curricula, respectively. Here we examine the range of progress rates, error frequencies, response latencies, and total time spent by students in the two courses Physics C: Mechanics (P051) and Physics C: Electricity and Magnetism (P053). As may be inferred from their titles, these courses were designed specifically to prepare students for the Advanced Placement (AP) examinations of the same names given by the College Board. They were produced in 1994 shortly after EPGY completed its calculus and algebra courses, to which the physics courses bear a similar instructional model and style of presentation. To date, over 150 students have completed these courses successfully; from these, 142 records were used to form the sample population of this study. As in the previous papers, we are interested in determining how much variation in performance is present among a population of highly gifted students. In addition, we wish to find the amount of dependence present among various measures of performance in our sample, as a

Individual Differences in Online Learning: Gifted and Other Diverse Populations.
Patrick Suppes.

means of determining whether any given measure largely determines students' overall rankings.

1.2 The EPGY Physics C Curriculum

The curricular design of the physics courses was largely informed by the requirements of the Advanced Placement exams. Mechanics covers the fundamentals of motion, gravitation, the concepts of work, energy, and momentum, as well as rotational and oscillatory motion. Electricity and Magnetism (E&M) includes material on electric fields and electrostatics, magnetic fields and magnetic induction, DC and AC circuits, and Maxwell's equations. Both courses carry knowledge of calculus as a prerequisite, and students are required to know the contents of the Mechanics course before starting E&M. The course material is largely presented via audiovisual lectures played from a CD-ROM, with follow-up exercises done on the computer as well as from a companion textbook. Students communicate weekly with EPGY instructors via email and phone; recently students have also met with instructors in "virtual office hours" using online videoconferencing software. Students ordinarily take exams at the end of each chapter of material covered from the text, as well as exams on larger parts of the courses that form conceptual units. All the results of this work are reported via email to the instructors, who are able to review the student's performance in detail and make suggestions for review. It is these emailed records that form the basis of the analysis presented in this report. At the end of the course, students take a timed final exam (ordinarily three hours in length), and usually students then go on to take the Advanced Placement examinations in their local school district. Further descriptions of the course model can be found in Cope & Suppes (1999) and on the EPGY web site, http://epgy.stanford.edu .

1.3 Student Selection

This study was performed on all of EPGY's extant records on the physics courses, which date back to 1996 and include students who completed the courses by May 2000, when this study was performed. As the courses have remained largely intact throughout this period, with only minor modifications made to the course content and structure, it is appropriate to compare student performance across this long time interval. The students who registered for the courses came to EPGY independently without being selected for eligibility in advance. Admission to EPGY is granted to students if they demonstrate mathematical ability that places them within the top fifteen percent of students na-

TABLE 1 Breakdown of total student numbers in each course by test scores, age, and sex. Ages are calculated at the time students began the courses. Note that the ages of 6 Mechanics and 3 E&M students were not reported.

Student Demographics

	Mechanics	E & M
Total	92	50
Number scoring in upper 3% of math ability	89	49
Male	77	43
Female	15	7
Average Age	16.1	16.6
Standard Deviation	1.6	1.2
Maximum Age	18.1	18.4
Minimum Age	10.0	12.2

tionally as measured by commonly recognized measures of aptitude or achievement, such as the SAT, PSAT, ACT, etc. Students are also admitted if they have scored a 5 on any of the Physics or Mathematics AP exams.

In all, the records of 92 students from Mechanics and 50 students from E&M were studied in this report, from a total of 107 students who completed Mechanics successfully and 52 who completed E&M.[1] Of these, 35 students took both the Mechanics and E&M course, which comprises a large percentage of the total E&M sample population. The large majority (over 95%) of the participants studied scored within the top three percent of mathematics ability, with the remainder scoring in the top 15%. There was also wide variation in age: the youngest students were younger than 12, and the oldest over 17 years of age at the point when they began the courses. A large majority of the students were male. Table 1 shows the breakdown of students by test score category, sex, and age.

1.4 Main Results of Individual Differences

Prior findings by EPGY (Stillinger & Suppes 1994, Cope & Suppes 1999) have indicated that several performance measures and characteristics vary widely among students in the high-ability range. We believe

[1] Some of the records were corrupted or unavailable; see below.

that in large part it is the self-paced nature of these courses which allows these differences to come forth, as students are able to work on the courses as frequently as they like, at any time of day, week, or year. Furthermore, students working in EPGY courses are not subjected to the social pressures that typically arise in a classroom setting. Students working in a computer-based curriculum perform very differently from each other. Among the students we studied, a ratio of approximately 2 to 1 was found between the third and first quartiles for many performance measures of interest, including rate and acceleration of progress, error rates, and question response times. This factor-of-two difference in performance shows that even within a group of students that is narrowly categorized as being among the top two or three percent in ability, there is wide variation in performance. Furthermore, we find that in general these performance measures are not highly correlated, indicating that overall student performance cannot be characterized according to a single measure or ability parameter.

In the sections that follow, our primary interest is to determine the degrees of variation among students, as well as to investigate the dimensionality of performance characteristics. However, it is also our concern to analyze the effectiveness of the curriculum, as well as to test the validity of certain commonly employed models of individual student progress through computer-based courses.

1.4.1 Computer-time Trajectories

Our first performance measures of interest characterize students' progress through the courses as may be plotted by a graph of their position in the course as a function of time. These plots, known as "trajectories" allow us to identify overall trends in student progress, particularly rates of progress and rates of acceleration through the courses. Reasoning from basic information-theoretic principles has led the second author to posit a power-function model to characterize rates of progress through a course.[2] Such a model has the form

$$y(t) = bt^k + c + \epsilon(t),$$

where t represents total hours of (computer) time spent working on the course, $y(t)$ represents the student's position in the course, and b, k, and c are parameters to be fitted to a given student's trajectory, and $\epsilon(t)$ is a mean-zero noise process. "Position in the course" simply

[2] This model has received a great deal of empirical justification; cf. Macken, van den Heuvel, Suppes & Suppes, 1976; Malone, Suppes, Macken, Zanotti & Kanerva, 1979; Suppes, Fletcher, & Zanotti, 1975, 1976; Suppes, Macken & Zanotti, 1978; Suppes & Zanotti, 1996 for the formal basis of the model.

means the percentage of computer-based course exercises completed; note that this weights all exercises equally. No particular distribution is assumed for the noise process.[3]

Of the fitted parameters, the most important in determining the overall shape of the trajectory is the power term k, which quantifies the student's rate of acceleration. Most fitted values of k from prior studies have been less than one, as we expect students' rates of progress to diminish as they progress to unfamiliar and generally more difficult material. The value of b approximates the average rate of progress throughout the course.[4] The parameter c is set to equal the student's starting position in the course; usually this was zero, unless the student either had not reported the first few course exercises or they were deliberately placed at a different starting point by the course instructor.

Data Collection

EPGY relies on its students to submit reports of their progress regularly. While course instructors try to ensure completeness of the student records, naturally it is not always possible to achieve this goal. Thus many of the available records had to be rejected because data were missing. If the student's trajectory was missing a consecutive block of exercises exceeding 8% of the total number of course exercises, the record was discarded; other student records were corrupted or missing altogether. Note that students do not necessarily complete all the exercises in the course, especially if they skip introductory material, and they may occasionally repeat exercises. Repeated segments were subsequently removed from the student record. In all, 70 trajectories for Mechanics students and 31 trajectories for E&M students were found acceptable for analysis. Table 2 gives the numbers of exercises in each course, as well as some statistics on the number of exercises that students completed.

Exercise completion times were recorded to the accuracy of a second. Only the time between when students were first presented with questions to when they answered was recorded. Students are normally given two chances to answer questions correctly: if the first answer is wrong, the computer displays information on the type of answer format expected, and students are then allowed to retype their answers. Unfortunately, the data-collection system on the EPGY software only

[3] The requirement of zero mean is unnecessary given that c is an arbitrary constant; the mean may be taken to be whatever the value of c may be.

[4] The average rate of progress based on our model may be calculated as bT^{k-1}, where T is the overall time spent in the course. As fitted values of k are generally near 1, we see that the value of this term will be near b.

TABLE 2 Average, minimum, and maximum number of questions answered by each student in each course.

Number of Course Exercises

	Mechanics	E&M
Per Course:	640	450
Per Student:		
Average	566	394
S.D.	78	65
Maximum	681	450
Minimum	322	211

records the total amount of time spent on each question, regardless of how many attempts the student made to answer correctly. Because it is also impossible to tell if a student is attending to the question during this entire time interval, the measurement includes any possible time off-task, such as getting up from the computer to answer the telephone, take a break, etc. We rejected any data sets where it appeared that an inordinate amount of time was taken on any single question.[5] Time spent watching lectures, taking tests, doing homework, and other course-related activities besides the computer-based exercises, is not included in these trajectories.

Table 3 shows the distribution, mean, and standard deviation of the completion times in hours for the students studied in this sample. This table shows some remarkable results regarding how quickly students could work through the material. Two students completed the exercises in the E&M course in less than an hour, although it should be noted that the records for these students included only around half the course exercises. Still, the students had average response times of 6 and 12 seconds per correct answer, which is even more impressive considering that over 90% of their answers were correct.[6]

[5] The threshold was 20 minutes for any given question.

[6] Indeed, these completion times are almost too low to be believed. We hesitated to include them at all on the grounds that it was most likely some sort of measurement error; however, since the data come from several separate reports to the same processor, and that it is nearly impossible to systematically falsify the record from the students' end, we have included them. It may have been the case, however, that the students were able to preview the questions (and answers) before they answered them for the record, as the EPGY software allows students to view the course in "browse" mode. It is unlikely that this feature was abused on a widespread scale by most students taking the course, however.

TABLE 3 Completion times for each course, in hours of computer time.

Total Computer Time to Complete Courses (Hours)

	Mechanics	E&M
Average	11.1	4.8
S.D.	5.9	3.0
Maximum	26.3	12.6
3rd Qu.	14.4	6.3
Median	10.2	3.8
1st Qu.	6.3	2.8
Minimum	2.7	0.6

Turning to the distributional statistics displayed in Table 3, we see a wide degree of variation; the ratio between the third and first quartile is again above 2, and the standard deviation is more than half the mean value in both courses. Of course, the amount of computer time spent in each course is very much related to the number of questions answered, which as we have already seen does vary among students. We shall eliminate this effect by considering average latencies in a later section; for now we shall focus on examining the fitted trajectory parameter values. One effect of the generally low values of the completion times is that the fitted values of k will be very sensitive to any outlying observations, because for small time values around one, there is little variation in t^k for a wide range of k values. Thus we must be careful to ensure that the fits are good.

Fitting the Curves

Once the student data had been collected, an iterative nonlinear least squares fitting procedure was used to find the best estimates of b and k (c was taken to be the starting position in the course). The procedure was first to find an initial estimate of the parameter values by performing a standard linear regression on the logarithm of the data, using the linear model

$$\ln(y(t) - c) = k \ln t + \ln b + \epsilon(t).$$

Once the initial values of k and b were established, a binary line search procedure was used to minimize the sum of the squared error terms

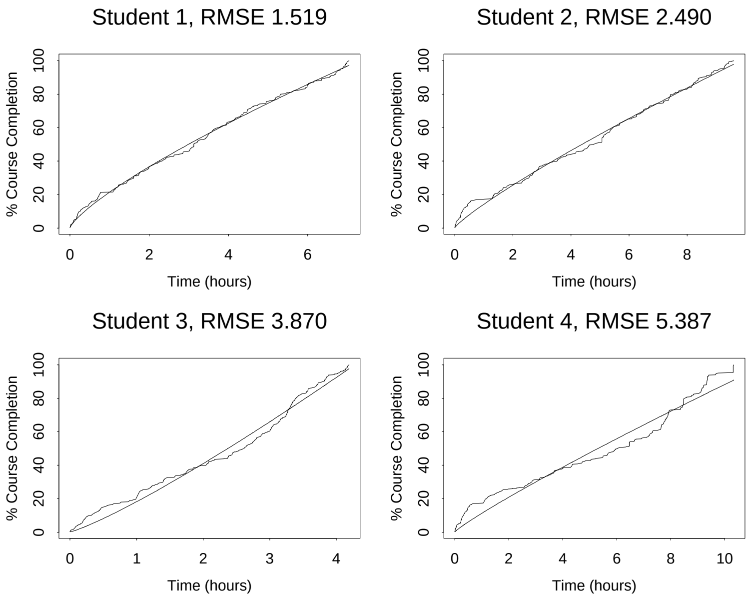


FIGURE 1 Sample student trajectories with varying degrees of root mean square error (RMSE). Jagged lines represent actual student trajectory data; smooth lines represent fitted curves. Fitted trajectories with RMSE less than 2 tend to be very good fits; those with RMSE between 2 and 4 moderately good; and those over 4 are poor fits.

over k and b, until the first-order minimization condition

$$\left(\frac{\partial}{\partial k} \;\; \frac{\partial}{\partial b}\right)\left[\sum_{j=0}^{n}(y_j - bt_j^k - c)^2\right] = (0 \;\; 0)$$

was fulfilled to within five digits of accuracy for k.

In general, the fitted curves seemed to match the overall student progress trends well, given that power curves are limited to being either strictly convex or strictly concave with no points of inflection. To assess the overall goodness-of-fit, a root mean square error (RMSE) statistic was recorded. As can be seen in Figure 1, trajectories with an RMSE of 2 or less seem to fit very well; trajectories where the RMSE was between 2 and 4 fit moderately well; and fitted trajectories whose RMSE was over 4 had poor fits. Figure 2 shows histograms of the RMSE values for both Mechanics and E&M. These histograms indicate that most of the fitted trajectories were very close to the actual progress curves.

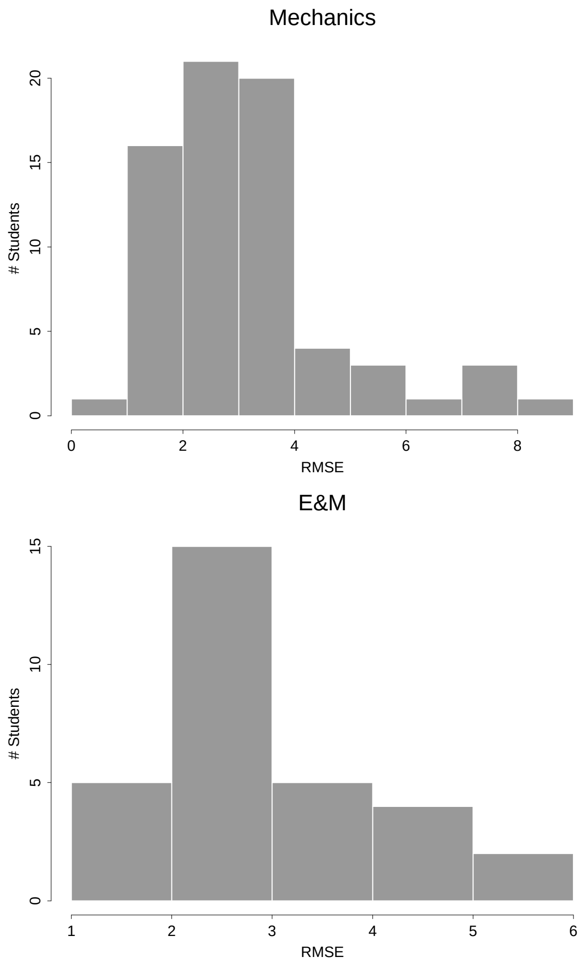


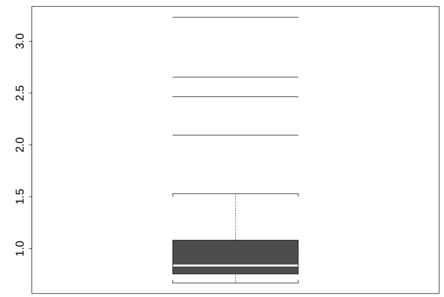
FIGURE 2 Histograms of RMSE values for fitted trajectories for each course. Most fitted trajectories show an acceptable level of fit, indicating that the model generally does a good job in characterizing the rate and acceleration of students' progress through the courses.

TABLE 4. Fitted trajectory parameter values for b and k for Mechanics.

Fitted Parameter Values

Mechanics

	k	b
Average	1.00	12.8
S.D.	0.46	6.7
Max.	3.23	35.4
3rd Qu.	1.07	16.0
Median	0.84	12.1
1st Qu.	0.76	7.7
Min.	0.67	1.6



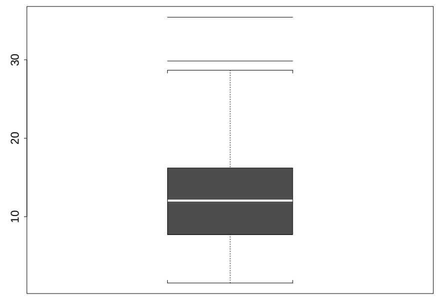
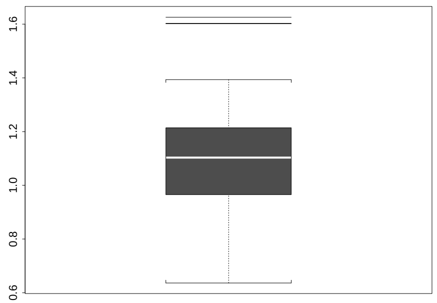
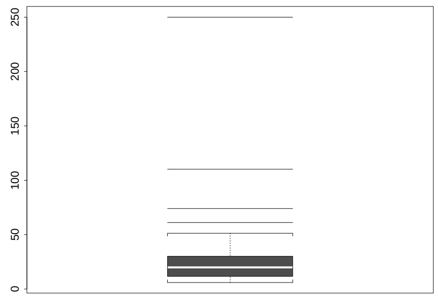


FIGURE 5. Fitted trajectory parameter values for b and k for E&M. The average and standard deviation of b was calculated without the two maximum values removed (250 and 110). These values came from the students who had extremely low completion times, and they greatly affected the values of the mean and the average.

E & M

	k	b
Average	1.13	22.8
S.D.	0.23	16.4
Max.	1.63	250.0
3rd Qu.	1.21	28.2
Median	1.10	20.0
1st Qu.	0.98	11.8
Min.	0.64	6.1





Fitted Parameter Values

The results of the parameter fittings are displayed in Tables 4 and 5, along with boxplots displaying the distributions. Most of the values of k for Mechanics showed sub-linear ($k < 1$) progress, whereas most of the values of k for E&M were greater than one. In the case of four Mechanics students, values of k exceeding 2 were achieved, which represents a very high rate of acceleration through the course material. In light of our earlier comments regarding our expectations for values of k, this latter result comes as something of a surprise. However, much of this may be attributable to the variable nature of the average length of time spent per question in different sections of the course. In the case of E&M, near the beginning of the course students are often asked to perform integrations which are not conceptually difficult but are time-consuming. Thus it takes students longer on average to get through the first few exercises than the last exercises, which is reflected in the unusually high values of the acceleration parameter k. If the exercises had been weighted to reflect this uneven character, then perhaps more normal fitted values would have resulted. However, it is beyond the scope and intent of this analysis to provide such a weighting; we are concerned here with the range of these values more than their actual magnitude.

Given that k is an exponent, the fact that most observed k values fall within a spread of width 0.31 in Mechanics and 0.23 in E&M constitutes a significant range. However, since most values of k are near 1 (as evidenced by the medians), and the completion times are as small as they are, we should instead view b as the parameter that better characterizes student performance. Values of b varied wildly, with extremely large outlying values in the case of E&M, where values of 250 and 110 were attained. These values corresponded to the two students discussed above with extremely low course completion times. We removed these scores when calculating non-robust statistics such as the mean and standard deviation. Again, we see that the standard deviation is more than half the mean value, and the first and third quartile differ by a factor of two in both courses.[7]

However, this seeming wide spread in parameter values is lessened slightly by an overall negative correlation between fitted values of b and k. For Mechanics students, the correlation coefficient between b and k was $r = -0.44$, and for E&M (again, deleting the outlying values), $r = -0.43$. These values are very close, and indicate that large b values were generally matched by smaller k values. This indicates that the

[7]This remains true even if we omit the two outliers in E&M.

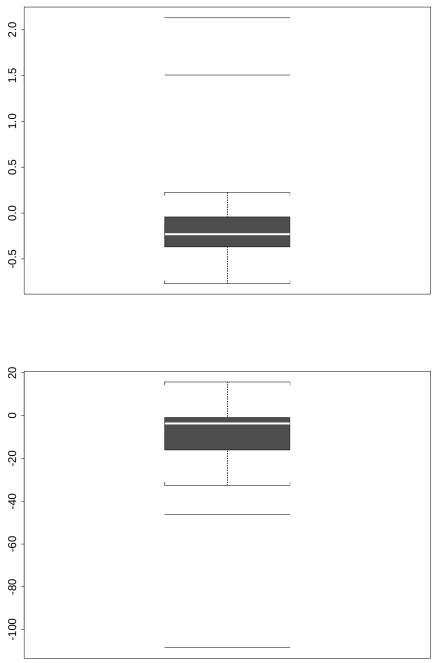


FIGURE 3 Boxplot of differences in trajectory parameter values. The plots show E&M parameter values subtracted from Mechanics parameter values.

overall variation in progress rates is not as great as the combination of ranges of b and k might suggest.

We may also compare values of k and b across courses for students who took both Mechanics and E&M. Eighteen students had valid trajectory data for both courses. Subtracting the parameter values for these students for E&M from those for Mechanics, we see that there is often a large difference in the fitted values; boxplots of the differences are shown in Figure 3. Values of k and b generally increase as students go into E&M. The correlation coefficients between parameter values across both courses is not high, and show a slight *negative* correlation (For k, $r = -0.122$, and for b, $r = -0.300$). We thus cannot conclude that the fitted parameter values are consistent across the different courses.

Model Validation and Course Assessment

As was mentioned above, the RMSE values of the fitted trajectory curves was for the most part quite good from the standpoint of tracking the general progress of the student. However, the model indicates that the residual differences between the actual trajectory data and the fitted curve should form a noise process that is uncorrelated over time. In general, this was not found to be the case. Applying a battery of tests for randomness, including a portmanteau test, runs test, and a sign test, a lack of randomness was consistently found among the residuals. Plots of the autocorrelation and partial autocorrelation functions generally indicated that the residuals had an autoregressive structure of order 1. An autoregressive model of the form

$$r(t_i) = ar(t_{i-1}) + \epsilon(t_i)$$

was fitted to the trajectory residuals, where $r(t_i)$ indicates the residual at the ith sample time t_i, a is an estimated parameter, and ϵ is again a zero-mean noise process. For both courses, the fitted values of a were very close to one (Mechanics: $\mu_a = 0.989, \sigma_a = 0.007$; E&M: $\mu_a = 0.978, \sigma_a = 0.012$). This indicates that there is an extremely high correlation between a residual value and the prior residual value. When this effect is subtracted from the residuals, the partial autocorrelation disappears and the remaining process resembles pure noise.[8]

This deviation from the theoretical model indicates that we should change the overall model slightly to compensate for the autocorrelation of the residual process. Indeed, if we treat a as being roughly 1, we arrive at the following model:

$$\frac{y(t) - y(t-\Delta)}{\Delta} = \frac{b(t^k - (t-\Delta)^k)}{\Delta} + \epsilon(t),$$

where $y(t)$ again is course position as a function of time, and Δ is an arbitrary unit of time difference. Taking limits as Δ approaches zero, we can abstract the above model to

$$y'(t) = bkt^{k-1} + \epsilon(t).$$

Given that both b and k are unrestricted variables, we may combine b and k to simply write

$$y'(t) = bt^k + \epsilon(t),$$

i.e., we now fit a power curve to the rate of student progress rather than to the position in the course. Note that we have now abandoned the constant c, as the constant term disappears from the derivative.

[8] The residuals from these processes do not appear to be normally distributed, however.

While we do not present this data here, subsequent fits of the student trajectories to this new model yielded an uncorrelated set of residuals, and the fitted values remained close to what one would expect based on the fitted values of the original model.

Rather than use the data to validate the model, however, we may alternatively treat the original model as normative and assess the relative difficulty of different sections of the course based on whether students are progressing at the rate that the best-fit trajectory curves would indicate. To this end, we may invert the trajectory curves to instead view time as the independent variable, and take averages of the residual values between the actual data and the fitted trajectories at several points in the course. That is, we simply take the residuals from the original trajectories in the "horizontal" direction.

Figure 4 shows the result of averaging these residual values across all students in each course. We can clearly see sections of the course where student's progress lines consistently deviate above or below the fitted trajectory line. However, what is more important is areas where student's rate of progress, rather than course position, is greater or less than that of the fitted trajectory line, which are indicated by areas where the plotted points show a distinct and sustained upward or downward trend. If we take the fitted trajectory line as indicative of a student's learning patterns, then areas where a student's rate of progress is greater than what would be indicated by their fitted trajectory points to an area of the course that is easier than may be appropriate for that student. Thus areas with a upward trend in the plotted points show harder sections of the course, and conversely for downward trends.

Thus we see in this anaylsis that even though the trajectory fits were globally good, in the sense that the overall RMSE values were small, more localized analysis reveals some problems. Not only are local values correlated in individual trajectories, but also across the trajectories of various students. This latter analysis indicates areas where the course material may become too easy or too difficult relative to the surrounding course material. However, as we already saw in the case of E&M, there is a difference between questions being conceptually difficult and their simply requiring a lot of time to solve; this distinction must be kept in mind if any revisions are to occur on account of these data.

1.4.2 Calendar Time to Completion

In self-paced courses there will naturally be variation in the amount of calendar time students take to complete their courses. A surprisingly high degree of variation was observed among the students in our sample. The distributions of the amount of time in days are displayed in Table 6

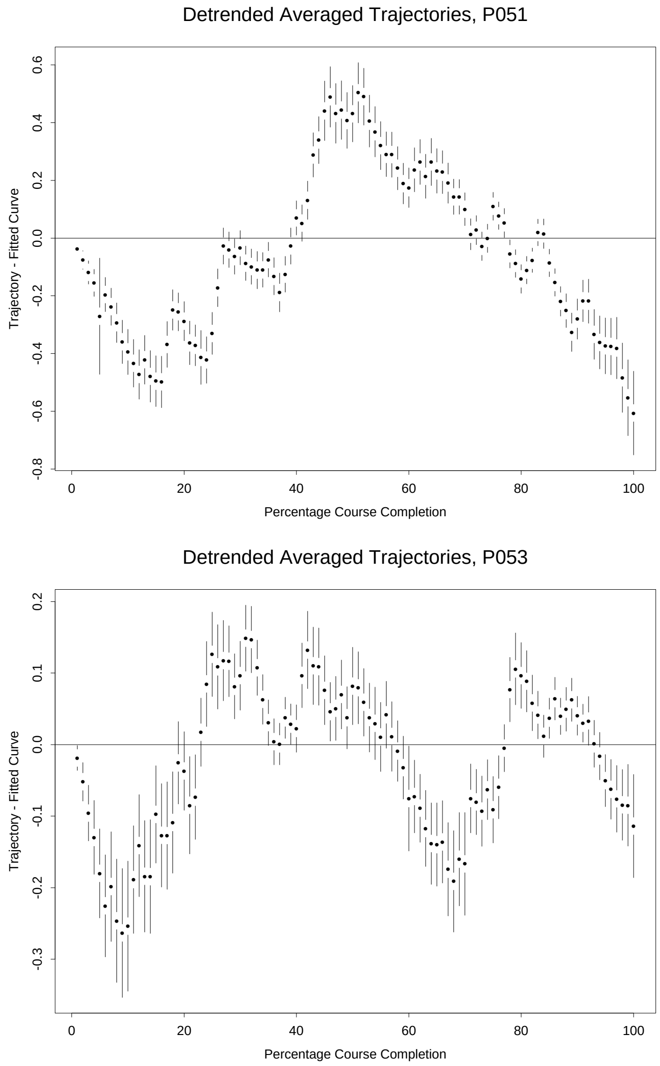


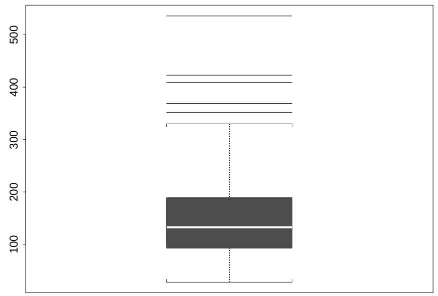
FIGURE 4 Plot of the average "horizontal" residuals of the best-fit trajectory curves for Mechanics and E&M students. Error bars indicate a 95% confidence interval for each point.

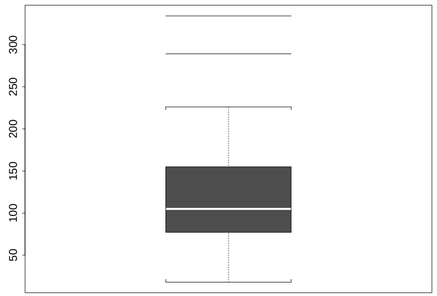
along with corresponding boxplots. In both courses, we see again that the ratio of the third to first quartile is approximately two, and the standard deviations are greater than half the average values. There

Table 6 Completion times for each course, in days. Again, minimum and maximum time differ by an order of magnitude, and a factor of two difference is evident between first and third quartile values for each course.

Course Completion Times (Days)

	Mechanics	E&M
Average	156	120
St.Dev.	88	66
Maximum	536	334
3rd Qu.	188	153
Median	133	105
1st Qu.	93	77
Minimum	28	18





September 23, 2009

is an especially wide range for Mechanics, as one student took more than a year to complete the course and another finished in less than 30 days. In addition, one student completed E&M in less than 20 days. We should note that unlike many other EPGY courses, and particularly the calculus courses studied in a previous paper, the physics courses are not billed on a per-time basis; course fees are rather paid in full before the course is begun. This may account for the wider spread in the distribution of calendar time to completion than in previously studied courses, as there is no financial incentive for early completion. Also, these statistics do not account for student time off from the course for vacations, sickness, etc., which may also inflate the completion times slightly.

Interestingly, there is little correlation between calendar times to completion and the amount of computer time required to answer all the questions (For Mechanics, $r = 0.112$, for E&M, $r = -0.080$). This shows that students who take longer to answer questions are about as likely to complete the course in the same number of days as those who answer questions quickly.

1.4.3 Response Latencies

We now turn to completion times of individual exercises, which again were measured as the number of seconds between the initial presentation of the exercise and the student's answer. Because students were given a second chance to answer if their initial response was incorrect, we analyze response times for correct and incorrect responses separately. We should note however that responses were counted as correct in the case where the first answer given by the student was wrong but was corrected on the second attempt. The distributional statistics of the average latencies for both correct and incorrect answers are listed in Table 7. As expected, latencies for incorrect answers are more than twice as long as those for correct answers, with the average time to answer Mechanics questions being longer than in E&M. For both incorrect and correct answers in each course, the ratio of the 75th to the 25th percentile is again approximately two and the standard error is approximately half the mean value. Also, the standard deviation of recorded latencies for each student is very large, indicating that there is a wide spread of latencies across all questions of the course. Again, some very low response times arise:

Table 7. Distributions of latencies per student in each course, in seconds. "S.D., student" refers to the average standard deviation of response latencies for each student. "S.D., course" refers to the standard deviation of the average latency values for each student.

Student Average Latencies, seconds

	Mechanics		E & M	
	Correct	Incorrect	Correct	Incorrect
Average	50.9	147.9	31.1	120.3
S.D., student	92.9	174.5	60.2	151.7
S.D., course	25.9	67.6	16.6	72.8
Maximum	142.0	320.6	79.6	269.0
3rd Qu.	63.5	178.9	37.9	187.2
Median	45.1	134.8	26.2	103.5
1st Qu.	34.0	109.3	19.7	60.9
Minimum	13.8	26.5	6.5	31.8

There were high correlations of average response times for students who had taken both courses (for correct answers, $r = 0.88$, and for incorrect answers, $r = 0.89$). There were also high correlations between student response latencies for correct and incorrect answers (for Mechanics, $r = 0.87$, for E&M, $r = 0.66$), but no large correlations with percentage correct on exercises or age for both courses.

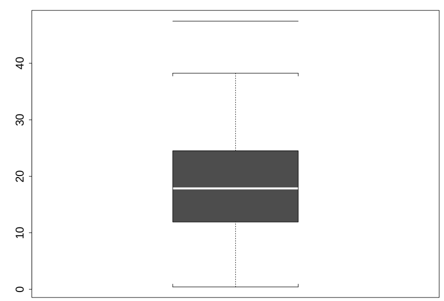
1.4.4 Error Rates

We next consider the percentage of errors students made on computer-based exercises, tests, and exams. Table 8 shows the distributional statistics of the percentage of computer-based exercises missed by each student, with associated boxplots. Table 9 shows the distributions of students' exam scores, both on the final and the chapter exams. Note that students are required to receive a score of 85% on each chapter test; until students reach that mark, they may continue to resubmit their exam scores. Table 9 thus shows not only the average maximum score earned by students on these tests, but the more telling statistic of the average scores on their first attempts.

TABLE 8. Student error rates for computer-based exercises, expressed as percentages.

Error Rates,
Computer-Based Exercises

	Mechanics	E&M
Average	18.7	12.7
St.Dev.	9.7	8.3
Maximum	47.5	34.2
3rd Qu.	24.5	17.4
Median	17.8	10.3
1st Qu.	12.0	6.7
Minimum	0.4	2.1



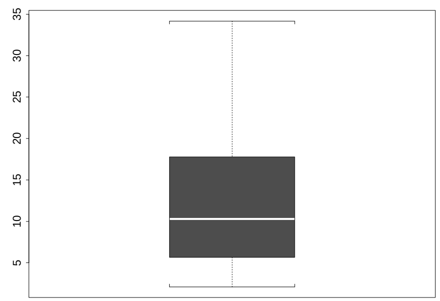


Table 9. Student error rates in online and offline tests (with both first attempt scores and maximum scores for any test reported), as well as final exams. All scores are expressed as percentages.

Student Error Rates, Tests and Final Exams

Mechanics

	First Attempts	Max Scores	Final Exams
Average	15.0	3.2	17.7
S.D.	9.5	3.6	11.6
Max.	37.8	16.7	61.0
3rd Qu.	20.7	4.7	23.0
Median	13.2	2.1	16.5
1st Qu.	7.7	0.5	10.0
Min.	0.3	0.0	0.0

E&M

	First Attempts	Max Scores	Final Exams
Average	12.3	5.5	17.0
S.D.	9.0	6.1	13.6
Max.	39.0	25.5	68.0
3rd Qu.	15.4	6.8	25.0
Median	10.3	3.8	15.0
1st Qu.	6.4	0.5	8.5
Min.	0.0	0.0	0.0

A positive correlation existed in both courses between students' first attempt test scores, computer-based exercise scores, and final exam scores; see Table 10 for a listing. A strong positive correlation was found between the test scores of students who took both courses (for first test attempt scores, $r = 0.87$, while for final exam scores, $r = 0.59$). Finally, no significant correlations between error rates and computer or calendar time to completion, or fitted trajectory parameters, was observed, nor were these rates significantly correlated with student age.

1.5 Dimensionality of Student Performance

In the preceding analyses we have looked at several measures of performance and mentioned in passing the relationships between these

Table 10. Correlation coefficients (r-values) for tests, final exams, and computer-based exercise scores. All show a positive correlation.

Correlations of Test and Exercise Error Rates

Mechanics

	First Test Attempt	Final Exams
Comp. Exercises	0.66	0.36
First Test Att.		0.34

E&M

	First Test Attempt	Final Exams
Comp. Exercises	0.55	0.47
First Test Att.		0.30

measures in terms of the sample correlation coefficient r. We now look more closely at these relationships to determine to what extent it is possible to rank students according to their overall performance using these measures. In particular, we wish to determine if only a few measures are sufficient to characterize student performance. That is, given that we know a student's ranking with respect to only one or two select measures, may we safely conclude that that student will achieve a similar rank with respect to the other measures? To this end, we construct a partial ordering of student performance according to subsets of performance scales. Among the measures we consider are error rates on computer-based exercises, tests, and final exams; average latencies on correct responses; fitted parameter values k and b; and, optionally, age, where we rate the student more highly the younger they are. If, given some or all of these measures, only a small group of students are undominated with respect to the partial ordering, then we may consider the various performance measures to be highly dependent in character, possibly leading to the conclusion that a single performance measure or latent variable might suitably capture the student ability. On the other hand, if the set of undominated students is large, then it seems more likely that the measures are independent of each other, and to characterize student ability requires knowing a set of measure values.

We also characterize the partial orderings in terms of the length of their largest *dominance chain*, which we define as an ordered list of students where each one dominates the next one in the list on all selected performance measures. Naturally, if we only consider one performance measure, then the longest dominance chain will have the same length as the size of the sample, assuming that none of the measure values are missing in our sample. As we add more measures, however, the lengths of the longest chain will decrease. If the longest chain becomes very small relative to the sample size, then there is little hierarchy evident in the partial ordering, so we may conclude that these measures do not stratify the student sample into many "levels of dominance". A highly stratified sample will generally have both a small undominated set and a large number of dominance levels. Thus by using these two features of the partial ordering we hope to characterize the degree of independence between performance measures.

The results of the partial orderings are summarized in Tables 11 and 12, where the maximum undominated set sizes and minimum number of dominance levels are listed for every size of subset of the nine performance measures considered. We have listed both the maximum set sizes where age was included and removed from the parameter set. Thus, for Mechanics, the combination of three parameters yielding the largest undominated set when age was not included was first exam average, final exam score, and percentage error on computer-based tests. In computing the partial orderings, any missing values were treated as actually being very poor scores. This was necessary as transitivity of the partial ordering can be violated if values are missing. Any student missing values in all categories under consideration was not included in the partial ordering.

We see in these tables that the undominated set sizes grow to 46% of the total student sample size in Mechanics and to 54% in E&M (where we do not including age among the performance measures); thus a very large percentage of students are not dominated in six measures. Also, the length of the dominance chains gets very small as more measures are included, to the point where only 3 dominance levels are present. In addition, there are significant increases in the undominated set without attendant decreases in the length of the maximal dominance chain up to the inclusion of about 5 parameters in both courses. The hierarchy of performance is thus very flat, and we might estimate the number of parameters relevant to characterizing performance to be around 5.

We have considered age as a performance measure separate from the other parameters to better view the effect of this factor on the other performance measures. If older students had a significant advantage over younger students, we should see a large increase in the undominated set sizes. Naturally, we do see some increase in the sizes of the sets, as we are adding another degree of freedom into the sets of possible order-

Table 11. Undominated sets and set sizes for partial orderings of Mechanicsstudents. The codes used for performance measures are listed in the above table.

Mechanics

Number of Measures	Largest Size of Undominated Set	Measures Included	Least Number of Dominance Levels	Measures Included
2	9	1E, PE	9	K, B
3	24	1E, F, PE	5	K, B, 1E
4	33	B, 1E, F PE	5	K, B, 1E F
5	39	K, B, 1E F, PE	4	K, B, 1E F, PE
6	40	K, B, 1E, F, PE, AC	4	K, B, 1E F, PE,AC

Mechanics, Including Age

Number of Measures	Largest Size of Undominated Set	Measures Included	Least Number Of Dominance Levels	Measures Included
2	9	A, PE	9	K, B
3	24	1E, F, PE	5	K, B, 1E
4	33	B, 1E, F PE	4	K, B, A 1E
5	42	K, B, A 1E, PE	3	B, A, 1E F, PE
6	48	K, B, A, 1E, F, PE	3	K, B, A 1E, F, PE
7	49	K, B, A, 1E, F, PE AC	3	K, B, A, 1E, F, PE AC

Code	Name
K	Fitted Parameter k
B	Fitted Parameter b
A	Age of Student
1E	First Exam Average
F	Final Exam
PE	Percent Error (Computer-based Exercises)
AC	Average Latency, Correct Resonses

ings; but none of these increases are very dramatic. However, because the undominated sets are very large already, it is difficult to conclude from this that younger students possess any significant advantage over older students.

Table 12. Undominated sets and set sizes for partial orderings of E&M students. The codes used for performance measures are listed in the above table.

E&M

Number of Measures	Largest Size of Undominated Set	Measures Included	Least Number Of Dominance Levels	Measures Included
2	11	K, F	6	B, F
3	17	K, 1E, F	4	K, B, 1E
4	24	K, B, 1E F	4	K, B, 1E F
5	27	K, B, 1E F, PE	3	K, B, 1E F, PE
6	27	K, B, 1E, F, PE, AC	3	K, B, 1E F, PE,AC

E&M, Including Age

Number of Measures	Largest Size of Undominated Set	Measures Included	Least Number Of Dominance Levels	Measures Included
2	11	K, F	6	B, F
3	20	A, F, AC	4	K, B, 1E
4	26	K, A, 1E F	3	K, B, A F
5	32	K, A, 1E F, PE	3	K, B, A, 1E, F
6	34	K, B, A, 1E, F, PE	3	K, B, A 1E, F, PE
7	34	K, B, A, 1E, F, PE AC	3	K, B, A, 1E, F, PE AC

Code	Name
K	Fitted Parameter k
B	Fitted Parameter b
A	Age of Student
1E	First Exam Average
F	Final Exam
PE	Percent Error (Computer-based Exercises)
AC	Average Latency, Correct Resonses

Table 13. Ratios of 75th to 25th percentile for many performance measures studied in this report. Most are around 2; recall that parameter k is an exponent so its relatively small ratio actually represents a large difference in itself, at least for the Mechanics course.

Ratios of Performance Measures
75th to 25th Percentile

	Mechanics	E&M
Computer Time	2.3	2.3
Parameter k	1.4	1.2
Parameter b	2.1	2.4
Calendar Time	2.0	2.0
Error Rate	2.0	2.6
First Test Att.	2.7	2.4
Final Exam	2.3	2.7
Latency, Corr.	1.9	1.9
Latency, Incorr.	1.6	3.1

1.6 Conclusions

Our findings reinforce the conclusions drawn in earlier papers that student performance in computer-based courses varies greatly at the high end of student ability. Most of the results again show that the 75th and 25th percentile of performance measures varies by a factor of two. Table 13 recaps these ratio values. We list parameter k in this table even though the ratio is much less than 2 since it is an exponent parameter, so even small deviations result in large differences in the resulting trajectories.

What is more, these performance measures are in large degree independent of each other, indicating that it may require several of these measures to identify overall student performance. The partial orderings computed in the previous section indicate that several students are not dominated in all categories by other students, which suggests that we may not be able to select only one or two measures by which to rank students. Furthermore, it indicates that students possess various abilities in different measures that contribute together to their success in the courses.

The data indicate that a modified power function which eliminates the autocorrelation in the residuals between the actual trajectories and the best-fit power curves may be more appropriate. When power curves are fit to the average velocity of a student over sections of the course, the residual values show no autocorrelation. This new model is consistent with the reasoning that underlay the prior model. The high order of autocorrelation in the model indicates that entire segments of the course are perhaps easier or harder than may be appropriate, as was indicated by the graphs of the residuals taken over percentages of course completion. However, the high degree of correlation found in these residuals indicates that large sections of the course may be too easy or too difficult; the graphs of these residuals presented above may now be used to update the courses in future revisions of the software.

1.7 References

[1] Cope, E., and P. Suppes. (1999). Gifted Students' Individual Differences in Computer-Based Calculus and Linear Algebra. Education Program for Gifted Youth Technical Report, Stanford, CA.

[2] Macken, E., R. van den Heuvel, P. Suppes, and T. Suppes. (1976). *Home-Based Education: Needs and Technological Opportunities.* National Institute of Education, Washington, DC.

[3] Malone, T. W., P. Suppes, E. Macken, M. Zanotti, and L. Kanerva. (1979). Projecting Student Trajectories in a Computer-Assisted Instruction Curriculum, *Journal of Educational Psychology* 71, 74-84.

[4] Stillinger, C., and P. Suppes. (1994). Gifted Students' Individual Differences in Computer-Based Algebra and Precalculus Courses. Education Program for Gifted Youth Technical Report, Stanford, CA.

[5] Suppes, P., J. D. Fletcher, and M. Zanotti. (1975). Performance Models of American Indian Students on Computer-Assisted Instruction in Elementary Mathematics, *Instructional Science* 4, 303-313.

[6] Suppes, P., J. D. Fletcher, and M. Zanotti. (1976). Models of Individual Trajectories in Computer-Assisted Instruction for Deaf Students, *Journal of Educational Psychology* 2, 117-127.

[7] Suppes, P., E. Macken, and M. Zanotti. (1978). The Role of Global Psychological Models in Instructional Technology, in *Advances in Instructional Psychology* (R. Glaser, ed.), Lawrence Erlbaum, Hillsdale, New Jersey, 229-259.

[8] Suppes, P., and M. Zanotti. (1996). Mastery Learning of Elementary Mathematics: Theory and Data, in *Foundations of*

Probability with Applications, (P. Suppes and M. Zanotti, eds.), Cambridge University Press, 149-188.