Gifted Students' Individual Differences in Computer-Based Algebra and Precalculus Courses

Constance Stillinger and Patrick Suppes

Stanford University

Abstract

In this article we summarize data from mathematically gifted middle school students working through home-computer-based first-year algebra, second-year algebra, and precalculus courses. Although these students represented only the extreme upper end of the full range of mathematical ability, they nevertheless displayed considerable individual variation on all observed outcomes with little relation to PSAT score, sex or age. About equal numbers of students accelerated and decelerated through each course. Any particular student's probability of ending a computer session at any point was largely independent of the number of instructional units already worked in that session; however, the value of that probability differed dramatically across students. A gamma model fit the observed distribution of latencies fairly well. Implications are discussed.

1 Gifted Students' Individual Differences in Computer-Based Algebra and Precalculus Courses

The Education Program for Gifted Youth (EPGY) is the descendant of a long tradition of research on computer aided instruction in mathematics and other disciplines at Stanford University (see e. g. Atkinson & Hansen, 1966; Suppes, 1981; Suppes, Jerman & Brian, 1968; Suppes & Morningstar, 1970; Suppes & Morningstar, 1972).

Like its predecessors, EPGY is dedicated to implementing traditional curricula with current computing technologies while collecting data relevant to both program evaluation and theories of student learning and performance. With facilities for communicating with tutors via electronic mail or telephone and for transmitting detailed progress information via electronic email, students are able to participate from anywhere that the appropriate voice and data connections can be established. Our focus as we have developed the EPGY course materials has been on mathematically gifted students, who frequently receive insufficient opportunities or encouragement to accelerate as dramatically as they could. Though gifted students share a common tendency to score at the top of relevant tests of ability, the findings in this article will show that once placed in an environment where no upper limits are set on pace or other aspects of performance, large individual differences can emerge. The EPGY calculus program, and the post-completion achievement test scores of the first group of students to have participated, has been discussed elsewhere (Ravaglia, Suppes, Stillinger & Alper, 1995; Suppes & Ager, 1995).

Overview. In this article we describe data collected on-line as students worked through first-year algebra ("Algebra 1"), second-year algebra ("Algebra 2") and precalculus courses. We first describe the structure of the courses in more detail and proceed to report the results for an early group of students, who scored in roughly the top 2% for mathematical ability on nationally recognized standardized tests. We examine their work schedules, their trajectories (course position as a function of time on task), the lengths of their self-scheduled work sessions, and their performance on the exercises presented throughout the courses. When we can, we also examine the relationships of the outcome measures to personal variables such as incoming test score, sex, and age. Simple mathematical models describing the data are invoked when they assist interpretation of the results. We conclude this article with some general remarks about the significance of the findings to educaiton of gifted students.

2 Procedures

2.1 Course structure

The core of EPGY's Algebra 1 (first-year algebra), Algebra 2 (second-year algebra) and Precalculus courses consists of a multimedia Microsoft Windows-based software package that students run on home computers. In addition to operating the software, students work through reading and problem assignments in a companion textbook, which thus plays roughly the same role as in more traditional courses. Students also consult instructors at Stanford by electronic mail and by telephone. In 1993, at the time the present data were collected, all students lived close enough to Stanford to be able to meet with instructors in person occasionally, as well. Students worked at their own pace within loose progress guidelines. At the time of data collection, final exams were offered weekly without prior appointment to anyone ready to finish a course that week.[1]

The online material is organized in the same way within all three courses. The material for each course is divided into conceptual segments, called "lessons" (of which there were 130, 72, and 106 in Algebra 1, Algebra 2 and Precalculus respectively at the time of data collection) focusing on a single topic. Each lesson, in turn, is subdivided into smaller structural segments, here called "units," presenting material in one of three formats—text, exercises, or a multimedia lecture. (There were an average of 2.2, 3.2 and 5.0 units per lesson in Algebra 1, Algebra 2 and Precalculus respectively.)

A text unit consists of a scrolling window containing written exposition for the student to read through as quickly or slowly as desired using scrolling commands from a menu. A lecture unit presents a short CD-based multimedia presentation (median length about 6, 9, and 4 minutes in Algebra 1, Algebra 2 and Precalculus, respectively) that is structured much like a short chalkboard lecture in a classroom. Students hear a narrator discuss a topic while watching his handwritten equations and graphic illustrations appear on

[1] Students now take final exams in their homes, proctored by a parent or other adult.

the screen.[2] Students are free to pause and resume these lectures or replay parts using the mouse and a set of on-screen controls resembling the control on a VCR.

An exercise unit presents a set of exercises for students to solve on-line (average = 6.1 with maximum = 14 in Algebra 1, average = 5.7 with maximum = 17 in Algebra 2, and an average = 2.3 with maximum = 20 in Precalculus). When a student enters a solution, the software announces whether it is correct or incorrect. If correct, the software either presents the next exercise in the sequence or moves on to the next unit if already at the end of the current exercise unit. If the student's answer is incorrect, the software prompts for another try, states whether the student's second attempt is correct or not, describes the solution if the student has still failed to supply the correct one, and then moves on to the next exercise in the sequence or on to the next unit, as appropriate.

Lesson 6340 is a typical lesson in Algebra 1. This lesson covers the solution of quadratic inequalities in five units: approximately five minutes of lecture on techniques for solving quadratic inequalities followed by a set of six exercises for the student to solve, a set of ten exercises, a followup lecture, and a concluding set of ten exercises. Students are able to go back and review any portion of any lesson already completed by selecting "Browse" mode from a menu. This lesson is typical in many respects, including the fact that exposition is handled entirely in lecture units. Text units are employed only occasionally in all three courses as a supplement to the lectures.

Data collection—the reporting system. The software saves a detailed history of every student's activities to a special file as he or she works through an EPGY course. Each student periodically uploads this information to the EPGY record-keeping computer at Stanford, by pointing and clicking on a graphical button in the EPGY display labelled "report," whereupon the student's computer transfers the record over the phone to an EPGY computer, where it is stored in an ASCII file under that student's name. A text-manipulation program further processs this file to convert it into input readable by statistical software.

[2]The students we describe in this article viewed the lecturer's actual handwriting; his notes have since been converted to printed text and graphics synchronized with the spoken lecture.

2.2 Participants

156 students (117 boys and 39 girls) provided the data described in this article. Of these, 86 had completed or were currently enrolled in Algebra 1 (69 boys and 17 girls), 83 had completed or were currently enrolled in Algebra 2 (65 boys and 18 girls), and 44 had completed or were currently enrolled in Precalculus (26 boys and 18 girls). Twelve additional students enrolled at the time produced no usable course data because their course reports were badly corrupted (i.e., the computer files had been rendered unreadable by software bugs or human mishandling). Table 1 gives the numbers of students involved in each possible combination of courses.[3]

Students' median ages were about 12, 12, and 13 in Algebra 1, Algebra 2 and Precalculus, respectively. Age was uncorrelated with sex ($r(68) = .06$). Their median grade levels in school were 7 in Algebra 1, 7 in Algebra 2 and 8 in Precalculus. (Age and grade level information are based on incomplete records: $n = 40$ in Algebra 1, $n = 35$ in Algebra 2, and $n = 16$ in Precalculus.) Nearly all of our students at the time resided within 25 miles of Stanford. We did not formally record ethnicity or economic circumstances.

Entrance scores. Students were usually required to pass an admissions test indicating exceptional mathematical ability when compared to their age group. The mathematics section of "retired" forms of the PSAT were used for this purpose. For those students for whom we have records, mean PSAT scores upon entrance were 482.1 ($n = 40$) for Algebra 1, 582.5 ($n = 20$) for Algebra 2 and 646.7 ($n = 6$) for Precalculus. We believe these medians are fairly representative of the group of students as a whole even though they are based on incomplete data. Scores were moderately correlated with age ($r = .51$, $p < .0001$). Although boys scored slightly higher on the whole than girls (boys' mean = 536.1, $n = 50$; girls' mean = 500.6, $n = 16$), this difference was not significant.

[3] Because only a total of 5 students enrolling in both Algebra 1 and Precalculus produced usable data, we will normally omit paired-observations comparisons of these courses in the results.

Table 1: Student Enrollment in Each Combination of Courses

Courses Taken		Algebra 1	Algebra 2	Precalculus
Algebra 1 only	Overall	51	—	—
	boys	40		
	girls	11		
Algebra 2 only	Overall	—	27	—
	boys		21	
	girls		6	
Precalculus only	Overall	—	—	18
	boys			9
	girls			9
Algebra 1 and Algebra 2	Overall	24	24	
	boys	21	21	
	girls	3	3	
Algebra 2 and Precalculus	Overall	—	19	19
	boys		13	13
	girls		6	6
Algebra 1 and Precalculus	Overall	1	—	1
	boys	0		0
	girls	1		1
All three courses	Overall	5	5	5
	boys	3	3	3
	girls	2	2	2
TOTAL	Overall	81	75	43
	boys	64	58	25
	girls	17	17	18

3 Results

We now turn to a description and analysis of the main results. We first consider the total amount of time students expended on each course, then examine the distribution of work session lengths, and finally describe exercise latencies and error rates. Our analyses and interpretations will be informed by simple mathematical models where applicable.

3.1 Total Elapsed Time and Trajectories

Total calendar time and total computer time. For each student, the software recorded the total "connect time" within each unit. Connect time represents the time the student actually spent with input focus aimed at the EPGY software (i. e., "connected" to the software). I.e., time spent with the EPGY window open but with input focus aimed at some other program was not included in connect time. Time spent watching and listening to the multimedia lectures was also excluded.[4]

The total elapsed calendar times and total connect times were calculated for students who had completed each course at the time of data collection (32, 31 and 10 students in Algebra 1, Algebra 2 and Precalculus respectively). Students required a median of about a third to a half of a calendar year to complete each Algebra course (171 days in Algebra 1, 120 days in Algebra 2) but about two-thirds of a year to finish Precalculus (median = 241 days). The interquartile ranges (134–260 days in Algebra 1, 81–174 days in Algebra 2, and 97–338 days in Precalculus) indicated wide variation in how much calendar time they expended in order to finish a course. Total connect times also varied widely, both between courses and among individual students. Algebra 1 students accumulated a median of 20 hours of connect time (excluding lectures) to finish, whereas Algebra 2 and Precalculus students both accumulated medians of 15 hours of connect time (one outlier deleted in Algebra 2). The corresponding interquartile ranges (Algebra 1: 14–30 hours, Algebra 2: 12–21 hours, Precalculus: 9–18 hours) were impressive; the third quartile was about twice that of the first quartile in all three courses. A moderate correspondence emerged between elapsed calendar time

[4]The current version of the reporting system does record lecture times, and future reports on EPGY data will include this variable. However, the exclusion of lecture intervals from connect times is not necessarily a liability for our purposes, because variation in lecture intervals should be largely independent of student.

and connect time in each course, when only those students who completed a course were considered (Algebra 1: $r(32) = .33$, Algebra 2: $r(30) = .59$, Precalculus: $r(10) = .45$).

3.2 Trajectories

We now shift our focus from the total time required to complete a course to a consideration of students' trajectories, or course position as a function of time. We defined course position simply as the percentage of course units completed. (Recall that a "lesson" corresponds roughly to a single-topic section of a textbook chapter, and a "unit" corresponds to a structural segment. There were 285, 228, and 535 units in Algebra 1, Algebra 2 and Precalculus, respectively.) Two alternative measures of elapsed time were available to us, connect times and calendar dates.

Figure 1 contrasts course position as a function of calendar date to position as a function of cumulative connect time for three students, one each in Algebra 1, Algebra 2 and Precalculus. These examples were chosen to illustrate the variety of outcomes that occurred. The first student depicted began with a very quick schedule, but then slowed down, and required a fairly large amount of computer time to finish. By contrast, the second student worked at a fairly leisurely, even pace, but required a relatively small amount of computer time to finish. The third student pursued a blazingly fast work schedule but took a moderate amount of computer time to finish.

Because of the irregularities apparent in many students' work schedules—at least half of the students who finished Algebra 1, Algebra 2 and Precalculus did not even run the software for a period of at least 27 days, 19 days and 33 days, respectively, and the fact that these irregularities were often due to events in the students' lives unrelated to their coursework, we chose to use connect times in our analyses of trajectories.

Previous research has revealed that students' trajectories through computer-based mathematics courses may be described by a power function, i.e. by an equation $y(t) = bt^k + c$, where y signifies course position, t signifies elapsed time, and k, b and c are free parameters that are individually estimated for each student (Macken, van den Heuvel, Suppes & Suppes, 1976; Malone, Suppes, Macken, Zanotti, & Kanerva, 1979; Suppes, 1992; Suppes, Fletcher & Zanotti, 1975, 1976; Suppes, Macken & Zanotti, 1978; Suppes & Zanotti, 1996). The most interesting parameter is k, for it determines the curvature of the trajectory: if $k < 1$, the rate of progress decreases with time; if $k > 1$,

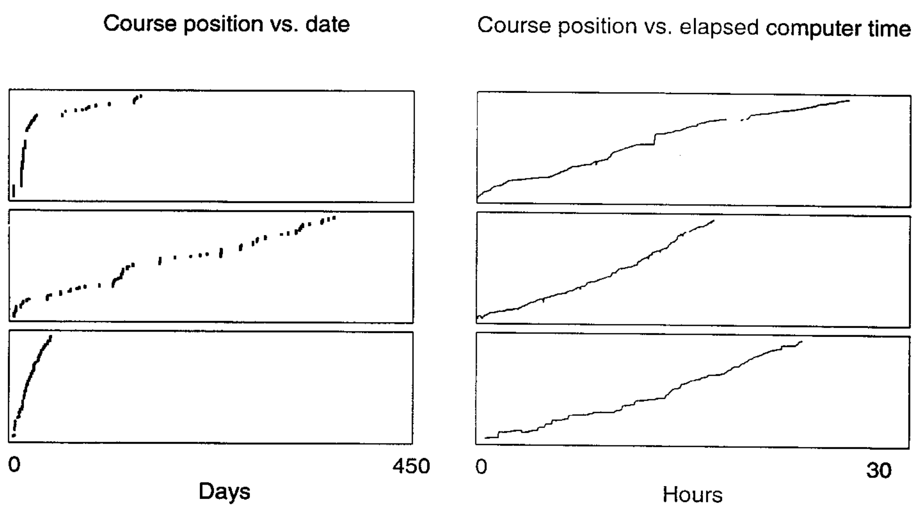

Figure 1: Trajectories of three students according to calendar time and computer time

on the other hand, the rate of progress increases with time; and if $k = 1$, the rate of progress remains constant (i.e. the trajectory is linear).

A nonlinear least-squares fitting procedure was applied to each student's trajectory in order to estimate the three free parameters in the power function model.[5] Missing position information was left missing, but missing connect time values were interpolated with the median connect time for that student in that course. Only a small percentage of all values were interpolated in this way (Algebra 1: 1.0%; Algebra 2: 0.7%; Precalculus: 7.6%). In addition, a small percentage of points were deleted that represented obvious positioning mistakes students made when resetting the software.

Table 2 presents the estimates for k, b and c as well as the root mean square error for the fits. Runs tests revealed no lack of randomness in the residuals for any student, indicating that power curves are consistent with the shape of the data. The median k estimate in all three courses was nearly 1–i.e., about hafl the students in each course sped up and about half slowed down as they proceeded; t-tests failed to detect a difference of the k estimates from 1 in any course or differences between the k estimates between courses. The finding of median k approximately equal to 1 contrasts with several earlier studies of student trajectories in computer-aided mathematics and reading courses in which the typical estimated value for k was closer to .5 (Macken *et al.,* 1976; Malone *et al.,* 1979; Suppes *et al.,* 1975, 1976; Suppes *et al.,* 1978),[6] but it is consistent with a more recent large-sample study of student trajectories through computer-based mathematics courses (Suppes & Zanotti, 1997).

Figure 2 displays the fitted trajectories, illustrating the different curvatures produced and the rapidity with which differences between students' course position grew. k estimates were regressed on students' age, sex, and entering PSAT when available. None of these personal variables taken alone predicted k; however when entered together in a linear regression they accounted for 29% of the variance in k ($F(2, 21) = 4.1$, $p < .03$). In this model,

[5]The Gauss-Newton method was employed here. This method is an iterative search procedure that may fail to converge if not supplied with reasonable starting values. Thus the procedure was run for the three starting values $k = .5,\ 1,$ and 1.5, for each student. The final sets of estimates for b, k and c were always identical whenever the process converged for more than one starting value for a single student.

[6]Malone *et al.* and Suppes, Fletcher & Zanotti (1975) reported the one value of k that best fits all individual trajectories; they did not report the actual histogram of individually estimated k's.

Table 2: Final Values of k_i, b_i and c_i for Fitted Trajectories

		Algebra 1 n=55	Algebra 2 n=55	Precalculus n=28
Parameter estimates				
	Median	1.01	0.95	1.10
k_i	Mean	1.00	1.00	1.10
	25%	.74	.77	.85
	75%	1.20	1.20	1.20
	sd	0.33	0.31	0.30
	Median	(.00)	.01	(.00)
b_i	Mean	.24	.12	.14
	25%	(.00)	(.00)	(.00)
	75%	.04	.03	.06
	sd	.97	.44	.45
	Median	12.90	9.10	13.70
c_i	Mean	25.90	21.30	46.60
	25%	1.70	3.50	5.60
	75%	36.30	28.00	72.80
	sd	48.80	37.60	70.90
Root mean square error				
	Mean	7.23	6.73	10.25
	sd	4.49	4.20	7.75

Note: Values in parentheses were small and positive, rounded to 0.

sex was a significant predictor of k in Algebra 1 when entered after PSAT ($t(21) = 2.6$, $p < .02$) such that boys tended to produce *lower* k estimates than girls after adjusting for entering PSAT score (boys' fitted mean = 1.0; girls' fitted mean = 1.4). No combination of these variables significantly predicted k estimates in Algebra 2 or Precalculus, however; Nor did any combination of sex, age, and PSAT score significantly predict any of the other outcome variables considered in the remainder of the article; hence we will not discuss them further.

Precalculus students produced the largest median estimate for k, whereas Algebra 2 students produced the smallest median estimate for k. Algebra 2 was the only course in which the median estimate was less than 1—i. e. in which the number of students whose work rate declined as they progressed exceeded the number whose work rate increased, possibly reflecting the fact that the first part of Algebra 2 consisted of review—more so than in the other two courses—allowing students to move more rapidly at first as they worked through familiar material, but causing them to slow down as new material was introduced. Nevertheless, two-sample t-tests (which included all students enrolled in either course) and paired-sample t-tests (including only those students who enrolled in both courses) revealed no difference between k estimates for any two courses.

3.3 Session duration

We next consider how much time students chose to allot to a single session working at the computer. Unfortunately, the most straightforward measure of session length, time in seconds, was not available to us. (Though the *total* amount of connect time spent within each course unit was recorded, we did not have a record of exactly when or how often *within* a unit a student shifted focus in and out of the software, ruling out a continuous time measure.) Instead, we exploited the fact that time off task *between* units was unambiguous, and let the number of course units a student worked without significant pause serve as the measure of session length. We took 10 minutes to be the maximum allowable interval between units considered to be part of the same session; longer gaps were taken to distinguish separate sessions. (This decision was based somewhat on the observed distribution of interunit intervals: the vast majority were 10 minutes long or less—6757, or 76% in algebra 1; 5737, or 80% in Algebra 2; and 7902, or 89 % in Precalculus–whereas $\leq 1\%$ were between 10 minutes and 1 hour long.) When we delimited sessions

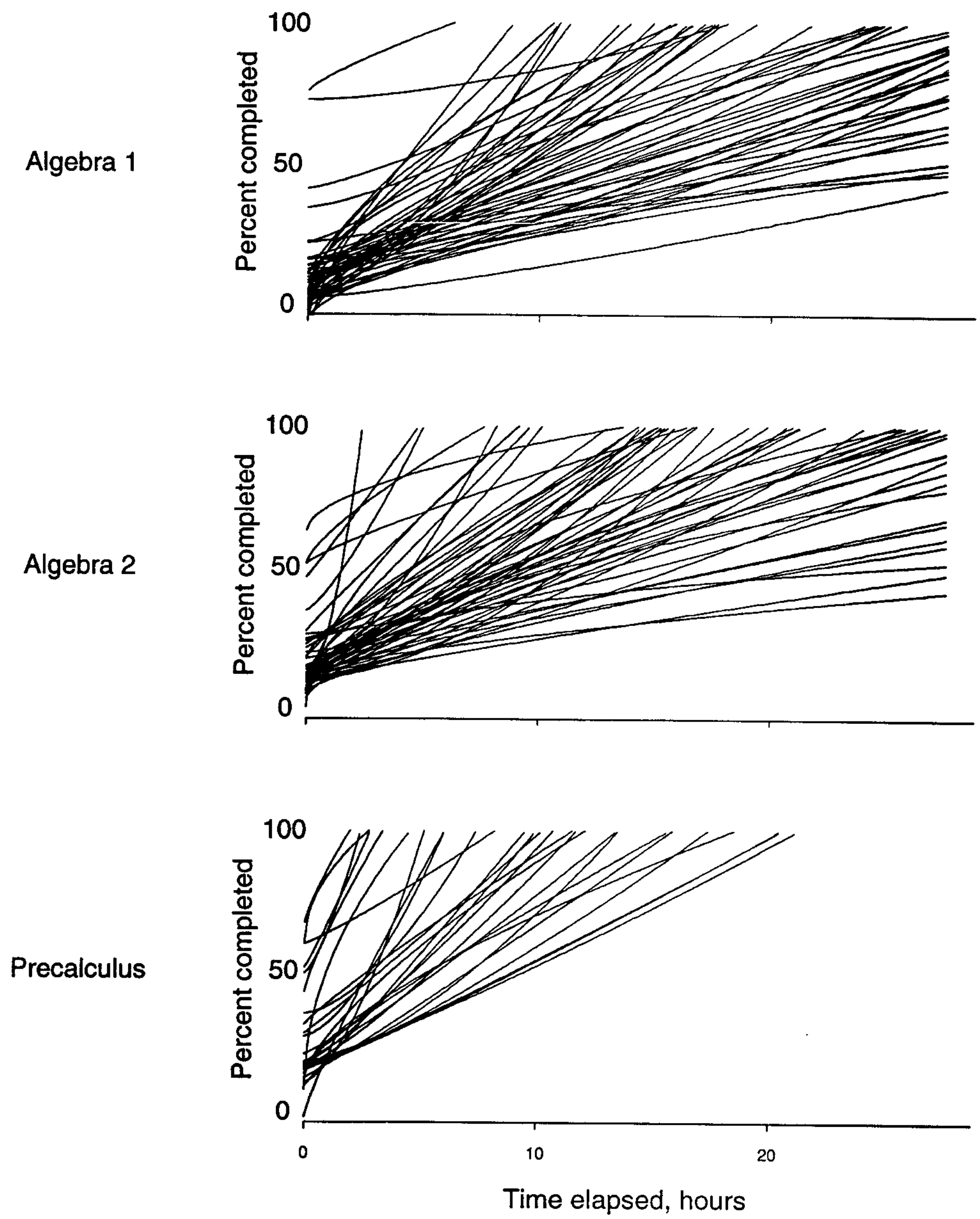

Figure 2: Fitted trajectories over elapsed computer time

in this way and counted the number of instructional units worked within a single session, we found that students in Algebra 1 worked a mean of 4.1 units per session (median = 3, 75th percentile = 5, maximum = 126, $n = 2168$ sessions); students in Algebra 2 worked a mean of 4.9 units per session (median = 3, 75th percentile = 6, maximum = 61, $n = 1481$ sessions); and students in Precalculus worked a mean of 8.4 units per session (median = 6, 75th percentile = 11, maximum = 353, $n = 1063$ sessions).

More interesting were the shapes of the observed distributions of session lengths (Figure 3 illustrates histograms for each course overall; the individual histograms had the same general shape), which did not approach normality at all; rather, these distributions closely approximated the geometric. A geometric distribution may be represented as the distribution of the number of times one must toss a weighted coin to obtain a single head; i.e., session length appears to result from a "memoryless" process–a students's probability of stopping after any instructional unit appears to be largely independent of the number of units he or she has already worked. The geometric probability density function can be written as

$$f(n) = p(1-p)^{(n-1)},$$

where n is the number of units in a session and p is the probability of stopping after any unit.

Individual student distributions. For each individual, a geometric density function was fitted to the observed distribution of session lengths using a nonlinear least-squares procedure. The obtained estimates for p are presented in Table 3 along with the root mean square errors for the fits. The p estimates vary markedly from course to course and from student to student within each course (from .07 to .67 in Algebra 1, from .05 to .50 in Algebra 2, and from .01 to .67 in Precalculus).

The differences between p estimates for all pairs of courses were statistically significant by two-sample t-tests (Algebra 1 vs. Algebra 2: $t(149) = 3.2$, $p < .002$; Algebra 2 vs. Precalculus: $t(111) = 4.6$, $p < .0001$; Algebra 1 vs. Precalculus: $t(120) = 6.1$, $p < .0001$). However, when only those students were considered who had enrolled in two consecutive courses, only the difference in p estimates between Algebra 2 and Precalculus was statistically significant (Algebra 1 vs. Algebra 2: $t(26) = .20$, ns; Algebra 2 vs. Precalculus: $t(21) = 4.3$, $p < .0004$). Correspondingly, the correlation between students' p estimates in Algebra 1 and Algebra 2 was a substantial .60 ($n = 27$,

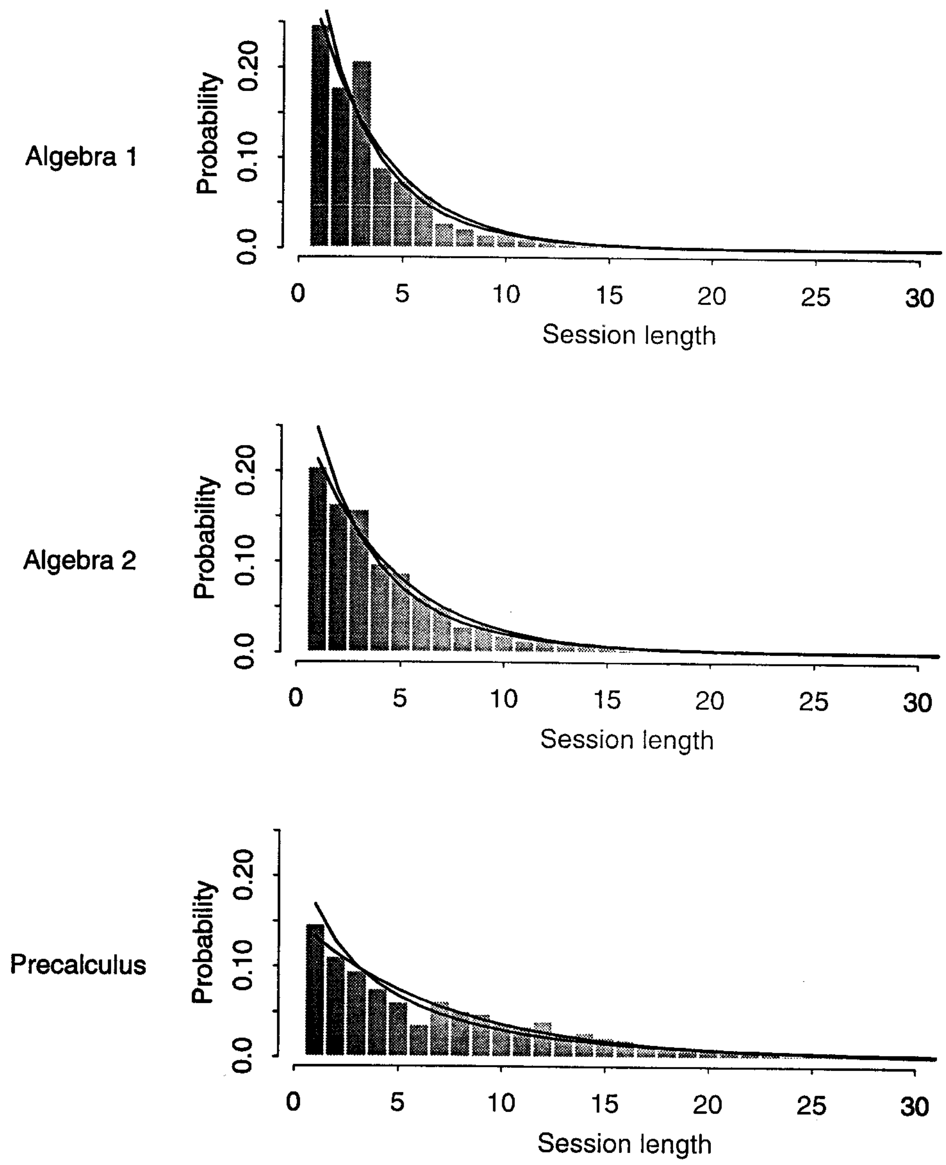

Figure 3: Histograms of session lengths in hours of students

Table 3: Estimated Parameter p for Geometric Densities Fitted to Individual and Mean Session-length Distributions

	Algebra 1 (n=80)	Algebra 2 (n=71)	Precalculus (n=42)
Individual estimates for p			
Median	.280	.240	.130
Mean	.300	.250	.160
25%	.210	.180	.090
75%	.360	.310	.190
Min	.072	.048	.008
Max	.670	.500	.670
sd	.130	.090	.120
Root mean square error			
Individual distributions			
Mean	.087	.074	.057
sd	.071	.050	.067
Mean distribution			
pure	.011	.010	.003
mixture	.008	.007	.003

$p < .01$), but the correlation betweeen students' p estimates in Algebra 2 vs. Precalculus was quite low ($r(22) = .01$, $F(1, 20) = .005$, ns. Regression calculations revealed no apparent relationship between the estimates for p and trajectory exponent k across students; nor were there any statistically significant relationships between p and any combination of sex, age and PSAT score.

Mean course distributions. As indicated by the root mean square errors in Table 3, a pure geometric density function fit the mean observed course histograms fairly well, better in fact than to any individual student's histogram. The fitted geometric density functions are drawn in on Figure 3. The overall stopping probability calculated in this way for each course was very nearly equal to the mean of the individual p's.

Under the assumption that the individual distributions are indeed geometric and that individual students' true stopping probabilities differ, the mean distribution would have the form of a mixed geometric distribution. We thus produced a predicted mean density function from the individual p estimates as follows:

$$f(n) = \alpha_i \sum p_i(1 - p_i)^{(n-1)}; \;\; \sum \alpha_i = 1, \;\; n = 1, 2, \ldots,$$

where n is the number of consecutive units worked in a session, the p_i are the individual estimates, and the weights α_i are equal to the number of sessions run by student i divided by the total number of sessions run by all students in the same course. The mean density functions predicted from this mixture of individual geometric distributions, superimposed on the obtained histograms depicted in Figure 3, appear to follow the observed histograms reasonably well. Root mean square errors of the predicted mixed density function from the observed densities over all observed n indicate that the predicted mixture–and hence the model assuming that students' distributions are geometric with different values of p– indeed fits the data considerably better than did a pure geometric density function.

3.4 Exercises

Student as unit of observation. All three courses included online exercises for students to work. Error rates and latencies within each course are shown in Table 4, where it is apparent that individual students' average latencies varied dramatically, ranging from about 14 seconds to nearly

5 minutes in each course. (Because these means include students who completed only part of a course, we must caution that the extremes may be exaggerated.) The reporting system recorded correctness and students' latencies on these exercises, with one complication: when students submitted a wrong answer, the software would usually allow them to try the exercise again. Thus the "latency" for a wrong answer actually represented the cumulative time for at least two attempts. Students on the whole answered somewhat fewer exercises correctly in Precalculus than in the Algebra courses, but these differences were not statistically significant, by either paired or unpaired comparisons. Error rates were correlated at a moderate level for students enrolling in both Algebra 1 and Algebra 2 ($r(27) = .58$, $p < .002$) and at a high level for students enrolling in both Algebra 2 and Precalculus ($r(23) = .86$, $p < .0001$).

Exercise latencies were significantly shorter for correct answers than for incorrect answers in the algebra courses but not in Precalculus (Algebra 1: $t(78) = 6.12$, $p < .0001$; Algebra 2: $t(71) = 6.00$, $p < .0001$; Precalculus: $t(39) = .60$, ns). Nevertheless, Precalculus students' latencies were shorter than those of students in Algebra 1 or 2, despite or perhaps because of their greater tendency to make errors, a tendency which on closer inspection turned out to result mostly from shorter latencies for *incorrect* answers in Precalculus than in the algebra courses (Algebra 1 vs. Precalculus: unpaired $t(12) = 2.27$, $p < .03$; Algebra 2 vs. Precalculus: unpaired $t(112) = 2.63$, $p < .01$). Students' latencies in Algebra 1 and Algebra 2 were correlated strongly for correct answers and to a lesser extent for incorrect answers (correct: $r(27) = .68$, $p < .001$; incorrect: $r(26) = .35$, $p < .08$; overall: $r(27) = .77$, $p < .0001$). Correlations for students taking both Algebra 2 and Precalculus were weaker (correct: $r(20) = .17$, ns; incorrect: $r(23) = .38$, $p < .08$; overall: $r(23) = .20$, ns).

The correlation across students between error percentage and estimated trajectory exponent within each course was small (Algebra 1: $r = .02$, $n = 54$, ns; Algebra 2: $r = .11$, $n = 54$, ns; Precalculus: $r = .22$, $n = 27$, ns). Not surprisingly, a negative correlation emerged across students between average exercise latency and k within each course, a correlation that reached statistical significance in the case of Algebra 2 (Algebra 1: $r(54) = -.06$, ns; Algebra 2: $r(54) = -.32$, $p < .05$; Precalculus: $r(27) = -.09$, ns).

Table 4: Error Rates and Latencies, Student as Unit of Analysis

		Algebra 1 n = 80	Algebra 2 n = 73	Precalculus n = 42
Error rate (percent)	Median	23.0	25.0	31.0
	Mean	27.0	26.0	31.0
	sd	17.0	14.0	19.0
	25th %ile	13.0	17.0	17.0
	75th %ile	38.0	34.0	43.0
	n	80	73	42
Latency, right answers (sec)[a]	Median	73.7	70.8	46.5
	Mean	82.5	79.7	59.1
	sd	47.6	48.2	43.9
	25th %ile	46.1	46.0	34.7
	75th %ile	107.2	92.3	71.3
	n	80	73	41
Latency, wrong answers (sec)[b]	Median	109.6	127.3	75.9
	Mean	129.5	136.2	95.7
	sd	85.0	87.9	62.0
	25th %ile	71.6	72.4	56.4
	75th %ile	173.5	171.3	118.5
	n	80	72	42
Latency, all exercises[a]	Median	82.0	86.3	58.1
	Mean	95.8	92.8	69.6
	sd	56.2	52.2	44.7
	25th %ile	55.5	56.9	44.5
	75th %ile	120.4	119.2	77.1
	n	80	73	41

[a]Precalculus n = 41. [b]Algebra 2 n = 72.

Table 5: Error Rates and Latencies, Exercise as Unit of Analysis

	Algebra 1	Algebra 2	Precalculus
Error rate			
Mean	.31	.31	.35
sd	.28	.27	.30
Median	.23	.25	.27
25th %ile	.08	.09	.10
75th %ile	.48	.48	.53
Latency			
Mean	101.50	95.30	75.50
sd	220.10	129.30	299.90
Median	65.20	61.70	35.60
25th %ile	30.20	29.70	16.60
75th %ile	124.80	120.50	76.80

Exercise as unit of observation. Table 5 tabulates the observed distributions of average error rates and average latencies for the exercises obtained by first calculating an average latency or percent incorrect over students for each exercise. Correlations between latencies and error rates were low and not statistically significant.

The latency histograms, Figure 4, are quite smooth with a shape that is reminiscent of a gamma distribution. Indeed, a gamma distribution of exercise latencies is exactly what would result if all exercises were composed of k steps, where all steps had identical exponentially-distributed latencies.[7]

[7] If a single exercise comprises a set of k steps, and the average latency for each step is an exponentially distributed random variable having parameter λ_j, then the latency for that exercise would be the sum of exponentially distribted random variables. Then if we assume that the exponential parameter for each step is the same, the average latency for that exercise is the sum of n independent exponential random variables having the same parameter λ, which is a gamma random variable; i.e., the average latency has a density function of the form:

$$f(x) = \frac{\lambda e^{\lambda x}(\lambda x)^{n-1}}{\Gamma(n)},$$

where $\Gamma(n)$ is the gamma function, defined by

$$\Gamma(n) = \int_0^\infty e^y yn - 1 dy.$$

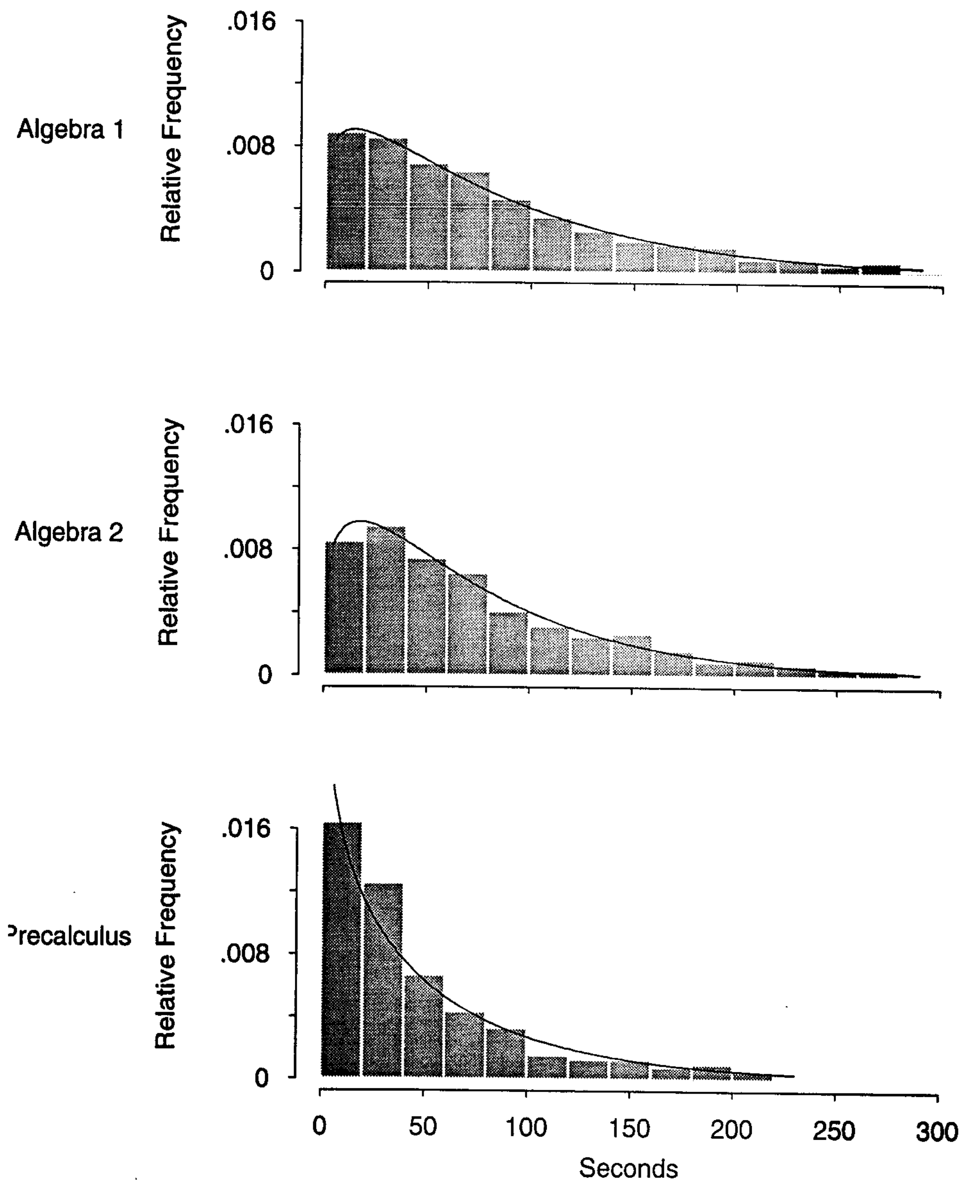

Figure 4: Latency histograms of students in each course.

To see how well this greatly simplified model describes the data, the natural logarithm of a gamma density function was fitted to the natural logarithm of the data via nonlinear least-squares (fitting to the logarithms rather than to the raw data was a programming measure necessary to avoid arithmetic overflow). The fitted gamma density functions, which have been superimposed on the histograms of observed data in Figure 4, in fact seem to follow the data remarkable well. In Algebra 1 the parameter estimates were $\lambda = .014$, $t = 1.18$; in Algebra 2 the estimates were $\lambda = .017$, $t = 1.33$; and in Precalculus they were $\lambda = .014$, $t = .779$. The corresponding root mean square errors were .23 (df = 27), .35 (df = 27), and .32 (df = 21).

4 Discussion

The most prominent feature of the foregoing results is the great range of individual differences the students displayed on every aspect of the courses we observed. Their trajectories, or progress with respect to time, furnished a particularly interesting example–about half accelerated through each course while half decelerated, contributing to the rapid increase in differences among students' course location illustrated in Figure 2. Table 6 reveals that indeed on most measures, the 75th percentile value was about twice that of the 25th percentile. Hence, although mathematically gifted students may seem alike when they are compared to the general population on a scale of ability, they will differ widely on aspects of their mathematics performance when offered a learning environment that imposes few constraints on their progress and work habits. A similar conclusion follows from the 1973 experiment of Macken et al. (1976) with highly gifted students, age 10–14 years, taking at home Stanford's university-level logic course, where the mean connect time was 37.7 hours, with the minimum 26.0 hours and the maximum 55.3 hours. The point that seemingly similar individuals can exhibit differences when restrictions are removed is in fact a very general one that will be familiar to personality psychologists and test-theorists as well as to educational psychologists and

Since n is an integer, $\Gamma(n) = (n-1)!$ and the density function reduces to

$$f(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!},$$

If we further assume that λ and n are approximately the same for all exercises, then the mean distribution of latencies over exercises should approximate a gamma distribution.

Table 6: Summary of Individual Differences: Ratio of 75th Percentile to 25th Percentile in Student Distributions

	Algebra 1	Algebra 2	Precalculus
Calendar time to completion	1.94	2.01	2.85
Computer time to completion	2.15	1.78	1.96
Latencies, all exercises	2.17	2.09	1.73
Latencies, exercises answered correctly	2.33	2.01	2.05
Latencies, exercises answered incorrectly	2.42	2.37	2.10
Error rates	2.84	2.02	2.44
Stopping probability p	1.71	1.72	2.11

mathematics educators.

Students' work habits, also unconstrained for the most part by EPGY's structure, presented several interesting features. Somewhat unexpected was the observation that students' decision whether to quit or continue a work session at a given point depended very little on how long that session had lasted so far. Though students differed in how likely they were to stop at any particular time, that likelihood appeared to be constant for a single student across work sessions. This finding becomes less problematic with an appreciation of the host of random influences that can affect students' work schedules in a completely self-paced setting at home. Such factors, if numerous and influential enough, might simply combine to overwhelm any systematic plans a student might have for any given work session.

While this group of gifted students exhibited large individual differences, they also produced low correlations between course outcome measures and "person variables"–PSAT score, sex and age–that could reasonably have been expected to predict outcomes. The PSAT was developed, along with its relatives, to predict academic success, and so might have been expected to covary with one or more of the variables we observed. We might also have expected sex to covary with one or more measures, given that a great many investigators over the years have observed differences between boys' and girls' distributions of mathematical facility; and these differences, though small toward the center of the distributions, would be enhanced at the extremes (e.g. Ross & Nisbett, 1991). Investigators have in fact found differences between gifted girls' and boys' mathematics performance (cf. Benbow, 1988 and Dweck, 1986), here raising some intriguing questions abut the contrast between the social and cognitive learning environments created by computer-mediated and human-mediated mathematics teaching systems. For example, we might ask whether girls get a performance "boost" from machine-based mathematics learning environments that actually stems from the *absence* of inhibitions normally hindering performance in the presence of human teachers. Analyses of larger groups of students with the inclusion of posttest scores are necessary to answer this and other questions about social and cognitive factors distinguishing computer-based learning from more traditional modes. A long-term, comprehensive computer-based program such as that of EPGY, which now offers a complete online mathematics curriculum from kindergarten through college level, provides the opportunity to track the long-term impact of computer-based learning and contrast it to more traditional educational experiences throughout the course of students' entire primary and secondary school career.

References

[1] R. C. Atkinson and D. N. Hansen. Computer-assisted instruction in initial reading: The stanford project. *Reading Research Quarterly*, 2:5–25, 1966.

[2] C. P. Benbow. Sex differences in mathematical reasoning ability in intellectually talented preadolescents. *Brain and Behavioral Sciences*, 1988.

[3] C. Dweck. Motivational processes affecting learning. *American Psychologist*, 1986.

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

[5] T. W. Malone, P. Suppes, E. Macken, M. Zanotti, and L. Kanerva. Projecting student trajectories in a computer-assisted instruction curriculum. *Journal of Educational Psychology*, 71(1):74–84, 1979.

[6] R. Ravaglia, P. Suppes, C. Stillinger, and T. M. Alper. Computer-based mathematics and physics for gifted students. *Gifted Child Quarterly Special Issue: Developing talent in science and mathematics.*, 39(1):7–13, Winter 1995.

[7] L. Ross and R. E. Nisbett. *The Person and the Situation: Perspectives of Social Psychology.* McGraw-Hill, 1991.

[8] P. Suppes, editor. *University-level computer-assisted instruction at Stanford: 1968–1980.* Stanford University, Institute for Mathematical Studies in the Social Sciences, Stanford, Calif., 1981.

[9] P. Suppes. Estes' statistical learning theory: Past, present and future. In A. F. Healy, S. M. Kosslyn, and R. M. Shiffrin, editors, *From Learning Theory to Connectionist Theory: Essays in Honor of William K. Estes*, volume 1, chapter 1, pages 1–20. Lawrence Erlbaum, Hillsdale, N. J., 1992.

[10] P. Suppes and T. Ager. Computer-based advanced placement calculus for gifted students. *Instructional Science*, 22:339–362, 1995.

[11] P. Suppes, J. D. Fletcher, and M. Zanotti. Performance models of american indian students on computer-assisted instruction in elementary mathematics. *Instructional Science*, 4:303–313, 1975.

[12] P. Suppes, J. D. Fletcher, and M. Zanotti. Models of individual trajectories in computer-assisted instruction for deaf students. *Journal of Educational Psychology*, 2:117–127, 1976.

[13] P. Suppes, M. Jerman, and D. Brian. *Computer-assisted instruction: Stanford's 1965–66 arithmetic program.* Academic Press, New York, 1968.

[14] P. Suppes, E. Macken, and M. Zanotti. The role of global psychological models in instructional technology. In R. Glaser, editor, *Advances in Instructional Psychology*, chapter 4, pages 229–259. Lawrence Erlbaum, Hillsdale, N. J., 1978.

[15] P. Suppes and M. Morningstar. Four programs in computer-assisted instruction. In W. H. Holtzman, editor, *Computer-assisted instruction, testing, and guidance*, chapter 12, pages 233–265. Harper and Row, New York, 1970.

[16] P. Suppes and M. Morningstar. *Computer-assisted instruction at Stanford, 1966–68: Data, models, and evaluation of the arithmetic programs.* Academic Press, New York, 1972.

[17] P. Suppes and M. Zanotti. Mastery learning of elementary mathematics: Theory and data. In P. Suppes and M. Zanotti, editors, *Foundations of Probability with Applications*, pages 149–188. Cambridge University Press, 1996.