

Projecting Student Trajectories in a Computer-Assisted Instruction Curriculum

Thomas W. Malone, Patrick Suppes, Elizabeth Macken,
Mario Zanotti, and Lauri Kanerva
Computer Curriculum Corporation
Palo Alto, California

Ten models are proposed for predicting a student's final grade placement in a computer-assisted instruction curriculum from the time the student spends taking lessons. Using data from approximately 2,000 elementary students, the models are tested both for ability to predict final grade placement and for ability to describe all points observed throughout the year. Two of the simplest models, using only the most recent point and parameters estimated for the whole group, are best at prediction. The power function model with all parameters estimated individually for each student, while relatively poor at prediction, is best for describing all observed points. Several methods for predicting the amount of time a student will spend taking lessons during a year are also tested.

In this article we consider 10 models for predicting a student's progress through a computer-assisted instruction (CAI) curriculum. With each of these models, we would like to be able to predict a student's final grade placement in a CAI curriculum as a function of the amount of time the student spends taking lessons in that curriculum. If we can successfully make these predictions, then by adjusting the amount of time individual students receive during the year, we can help students reach individually determined grade-placement goals, or we can optimize, for a school or class, any objective that can be stated in terms of final grade placement. Earlier work in this area (Suppes, Fletcher, & Zanotti, 1975, 1976) concentrated on mathematically *describing* student trajectories through the curriculum by fitting curves to a set of points that have already been observed. Here we focus on the somewhat different problem of *predicting* a future trajectory from past observations.

To test our models we use observations of

a student's grade placement during the first part of a year to predict final grade placement for the student. Since our data consist of student histories for an entire year, we can observe the amount of CAI time each student actually received by the end of the year. We begin our tests of the models by using these observed final times in the prediction of final grade placement. However, in an actual application of our models, final time would not be known before the end of the year, so we next compare these first predictions of final grade placement to predictions based on time values that are themselves estimated. For comparison with the earlier work by Suppes et al., we also test how well the various models fit curves to all the points for the year.

Method

Students

We tested our models with CAI data collected during the 1975-1976 school year from approximately 2,000 third- through sixth-grade students in the Fort Worth Independent School District. The students were scheduled to take daily CAI sessions of 10 minutes each in elementary mathematics, elementary reading, or both. In all our analyses, we treated the data from the two curriculums separately.

The CAI program was funded by Title I of the federal Elementary and Secondary Education Act. Therefore,

Thomas W. Malone, Patrick Suppes, and Mario Zanotti are also at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

Requests for reprints should be sent to Elizabeth Macken Raugh, Computer Curriculum Corporation, 700 Hansen Way, Palo Alto, California 94304.

in order to be eligible to participate in the program, children had to have been designated educationally disadvantaged. The definition of educationally disadvantaged children that was used by the funding agency depended on standardized test scores as follows: third graders whose scores were at least 7 months below grade level, fourth graders whose scores were at least 8 months below grade level, fifth graders whose scores were at least 9 months below grade level, and sixth graders whose scores were at least 12 months below grade level.

Curriculums

The curriculums used in this study were developed by Computer Curriculum Corporation (CCC) to provide drill and practice in basic mathematics and reading skills. Both curriculums use a strands instructional strategy to individualize instruction, and both keep undated records of student progress. We will outline the main features of the strands approach and the progress reports in the next few paragraphs; further details of the content of the curriculums and the criteria for student movement can be found in the CCC teachers' handbooks for reading (Adkins & Hamilton, 1972) and mathematics (Suppes, Jerman, Kanz, Morningstar, Searle, & Clinton, 1975).

A strand contains a series of exercises from the same content area. For example, the mathematics curriculum includes strands on addition and division, and the reading curriculum includes strands on vocabulary and word attack. Similar exercises within a strand are grouped into equivalence classes, which are then ordered according to their relative difficulty. Grade levels are assigned to each equivalence class according to the appearance of similar exercises in elementary textbooks and standardized achievement tests. During a 10-minute lesson, a student receives a random mixture of problems from all the strands appropriate for the student's grade level. But the difficulty level of the problems in each strand is automatically adjusted to the

student's achievement level in that strand. The curriculum programs contain preset performance criteria for each strand to determine whether a student needs more practice at the current level of difficulty, should move back to an easier level for remedial work, or can move ahead to a more difficult level. These criteria are adjusted so an average student will gain roughly 1 year in grade placement during 1 year of regular lessons. The grade placement determined by these curriculum programs is therefore a combination of criterion-referenced exercises with norm-referenced motion algorithms. The programs determine a weighted average of the student's grade placement across all the strands in a course, and it is this average grade placement for each student which is analyzed in the remainder of this article. We collected the grade placement and cumulative time for each student at approximately 2-week intervals throughout the school year. Thus we have a set of 20–25 points in each curriculum for each student which describe the student's trajectory through the curriculum.

Models

The ten models we considered are shown in Table 1. In each model a student's grade placement is expressed as a function of time. In all models we use $GP_i(t)$ to denote the grade placement of student i in the mathematics or reading curriculum after receiving t minutes of instruction in that curriculum. The learning parameters b_i , t_i , and c_i are estimated individually for student i ; the unsubscripted parameters b and k are group parameters estimated from data for the entire population and then used for each student. For student i , t_1 is the amount of time accumulated by the first data collection.

Table 1
Ten Trajectory Models for Projection and Curve Fitting

1. Power function model—individual k	$GP_i(t) = b_i t^{k_i} + c_i$
2. Power function model corrected for start time	$GP_i(t) = b_i(t + c_i)^{k_i}$
3. Power function model corrected for beginning knowledge—individual k	$GP_i(t) = b_i(t - t_1)^{k_i} + GP_i(t_1)$
4. Power function model—group k	$GP_i(t) = b_i t^{k_i} + c_i$
5. Linear regression model	$GP_i(t) = b_i t + c_i$
6. Linear model corrected for beginning knowledge	$GP_i(t) = b_i(t - t_1) + GP_i(t_1)$
7. Power function model corrected for beginning knowledge—group k	$GP_i(t) = b_i(t - t_1)^{k_i} + GP_i(t_1)$
8. Piecewise linear model—individual b	$GP_i(t) = b_i(t - t_r) + GP_i(t_r)$
9. Piecewise linear model—group b	$GP_i(t) = b(t - t_r) + GP_i(t_r)$
10. Piecewise power function model—group b	$GP_i(t) = b(t^5 - t_r^5) + GP_i(t_r)$

Note See text for explanation of variables.

The first collection was made at the end of the initial placement period; grade placement at this point is the first reliable estimate we have for the student's true beginning grade placement in the curriculum. Similarly, t_r is the time a given student had accumulated by the most recent collection date before the date from which the predictions are to be made. As noted above, data collections were made at approximately 2-week intervals. In general, t_j is the accumulated time at collection date j for student i . For simplicity of notation, the subscript i is omitted from the time variable.

We distinguish two uses of these models—namely, curve fitting and prediction. We can use the models to approximate student learning curves based on data from an entire school year, or we can use them to predict grade placement at future points based on observed data for part of the school year. When models are used for curve fitting, we can obtain better fits by adding more parameters—particularly parameters that are estimated individually. On the other hand, it is a familiar scientific experience that too many parameters, though leading to good curve fitting, can lead to bad predictions. The large number of parameters permits adjustment to idiosyncratic characteristics of the data which do not necessarily reflect the underlying tendencies of the process being studied. Accordingly, some of the simpler models may in fact be superior for our primary purpose of prediction rather than curve fitting. We will now discuss the different projection models shown in Table 1 and the rationale for each one.

Model 1: Power-function model—individual k . In Model 1 placement in the CAI program is assumed to be a power function of time, and all three parameters are estimated individually. The parameters are chosen for each student to minimize the standard error of prediction over all points for that student. Suppes et al. (1976) present a theoretical derivation of this model based on axioms about rates of introduction and sampling of information. In this formulation b and k are both information processing rate parameters, and c is meant to estimate the amount of knowledge the stu-

dent has at the beginning of the course. Suppes et al. tested this model with data from 297 students and achieved a relatively good fit with a mean standard error of .046 grade years. Based on these results and the fact that the model has three individually estimated parameters, we expected it to be one of the best of our models in terms of curve fitting.

Model 2: Power-function model corrected for start time. Model 2 is a slight variation of Model 1; the difference involves the interpretation of the parameter c . In Model 1, c translates the curve vertically and can be thought of as an estimate of the initial knowledge of the student—the knowledge at $t = 0$. In Model 2, c translates the curve horizontally and can be thought of as an estimate of the time at which learning starts. Some students may have started their learning of the subject matter in a curriculum before beginning the curriculum, whereas other students may not actually start learning until some time after the beginning of the curriculum. The parameter c serves as a correction to the starting time.

Model 3: Power-function model corrected for beginning knowledge—individual k . Model 3 attempts to exploit the interpretation of c as the amount of knowledge a student has at the beginning of the course. Instead of leaving c free as a parameter that varies to fit the curve, this model sets c equal to $GP_i(t_1)$, the first observed grade placement for the student. The time parameter must then be adjusted by subtracting t_1 so that the curve passes through $GP_i(t_1)$ at $t = t_1$. Model 3 is thus a slight variation of Model 1—the difference being in the method by which c is estimated.

Model 4: Power-function model—group k . Model 4 is like Model 1 except that a fixed value of k is used for all students rather than an individually estimated k . Suppes et al. (1976) considered this model as well as Model 1. They found that for values of k between 0 and 1 the mean standard error for their data was minimized when $k = .45$. We use $k = .50$ because, as we will describe in more detail below, $k = .50$ minimizes the standard error for our data.

Model 5: Linear regression model.

Model 5 is the linear version of Model 4; it sets $k = 1$ and thus becomes a simple linear regression model.

Model 6: Linear model corrected for beginning knowledge. In Model 6 we use the same intuitions about estimating c with initial grade placement as we did in Model 3, but we make the additional simplifying assumption that the relation of grade placement to time is a linear, not a power function.

Model 7: Power-function model corrected for beginning knowledge—group k. In Model 7, we use the same intuitions about estimating c from initial grade placement as we did in Model 3, but rather than estimating k individually for each student as in Model 3, we use a constant $k = .5$ for all students as in Model 4.

Model 8: Piecewise linear model—individual b. Model 8 is based on an intuition that the most recent points should influence predictions more than early points. This model forces the curve to pass through the most recently observed point $(t_r, GP_i[t_r])$. For example, if we are trying to predict final grade placement using only points observed in the first half of the year (say, Collections 1 through 10), then this model forces the grade placement curve to pass through the point observed at Collection Date 10. We predict that grade placement will then increase linearly from that point with a slope equal to b_i . The slope b_i is determined for each student as the slope of a line that passes through the most recent point and best fits the previous points already observed for that student.

Model 9: Piecewise linear model—group b. In Model 9 we use the same intuitions as in Model 8, but instead of estimating the learning rate b_i for each student, we estimate a population learning rate b based on data for all students. We predict that the curve will increase linearly from the last observed point at a rate b that is the slope of a line that best fits all the points for all students. To avoid using “knowledge of the future” in our predictions, we divided the students into even-numbered and odd-numbered halves and then used a value of b estimated from one half to make predictions for students in the other half. In practice, b could be esti-

mated using data from previous years rather than by using split halves in this way. In many ways Model 9 might be considered the simplest possible prediction method. We predict simply that the student's grade placement will continue to rise from his last observed point at a rate that is about the average rate for the whole population.

Model 10: Piecewise power-function model—group b. Model 10 is in the same spirit as Model 9, but instead of a linear increase, we postulate that a student's grade placement increases with the square root of the amount of CAI time. The exponent .50 was chosen as in Model 4 because it minimized the error for our data. The correction factor $(-t_r^{-.5})$ is necessary to move the curve so that it passes through the most recently observed point. We estimated b to minimize the error using data from half the students, as we did in Model 9, and then we used that value to predict final grade placements for the remaining students.

Summary of models. Beginning with the general three-parameter power function, we have proposed a number of alternative models embodying different parameter estimation methods. Each of the three parameters can be estimated either for each student individually or for the group as a whole. The parameter c can also be estimated by simply observing the first or the most recent grade placement for an individual student. Suppes et al. (1976) present theoretical arguments for the unique appropriateness of power functions as global models of student progress. We have not tried to list all the possible combinations of estimation methods here, but we feel we have selected all the most plausible models from the general space of three-parameter power functions.

Results

Errors in Prediction

To compare the models in terms of predictive ability, we computed the standard error of final grade-placement predictions as follows:

$$e = \left[1/n \sum_i [\hat{GP}_i(t_f) - GP_i(t_f)]^2 \right]^{1/2},$$

where $\hat{GP}_i(t)$ = predicted value of GP for student i at time t , $GP_i(t)$ = observed value of GP for student i at time t , t_f = time accumulated for student i by the end of the year, and n = number of students.

We computed a standard error for each collection date by using for the final grade-placement prediction only data observed prior to or on that date. For example, the standard error for the sixth collection is derived from predictions of final grade placement based on knowledge of grade-placement levels at the first 6 collection dates. The results of these error computations for the models that predicted best overall are shown for the mathematics and reading curriculums in Figures 1 and 2.

The somewhat surprising result of this analysis is that for both curriculums the best predictions are given by the two simplest models, Models 9 and 10. Evidently our intuitions about the importance of the most recently observed point were correct. For mathematics Model 10 gives slightly better predictions than Model 9, but for reading the

results are almost indistinguishable. For the mathematics data, standard errors before the last point for Models 9 and 10 range from .06 to .47 grade years and from .07 to .49 grade years, respectively; for reading, the range is from .03 to .39 grade years for both models. It is particularly interesting that Model 9, with the learning rate parameter constant for all students, gives better predictions than Model 8, in which the learning rate is estimated individually. Similarly in the case of the parallel models, Models 1 and 4, Model 4, with a fixed exponent of $k = .50$, gives better predictions than the more complex Model 1, in which the exponent k is estimated individually for each student. Note that these differences are more pronounced in the mathematics data than in the reading data, where the standard errors are generally smaller. In the reading curriculum, Models 9 and 10 appear to be best over the whole year, but Models 1 and 4 do as well or better on most collection dates in the latter half of the year. The larger error in the mathematics predictions may be due to

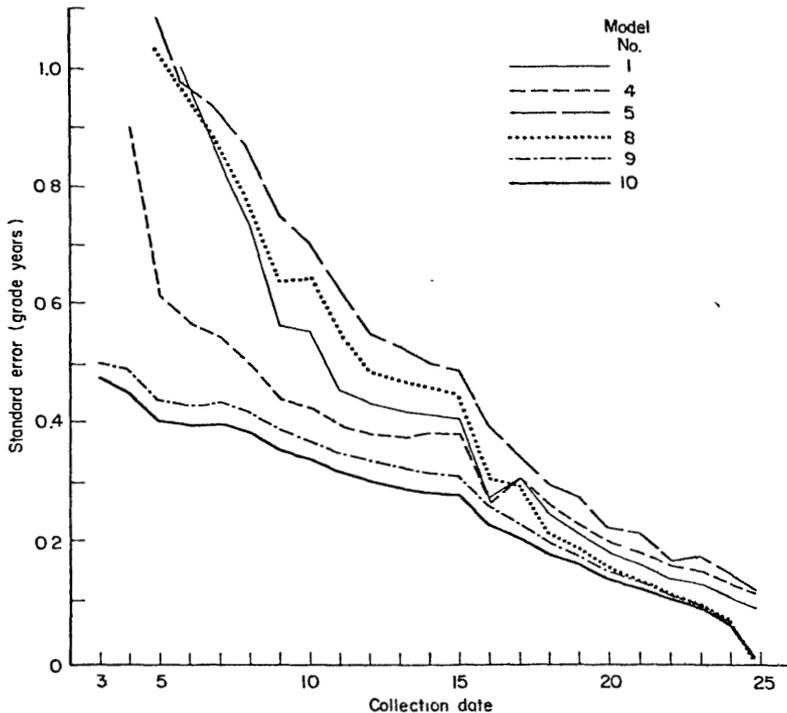


Figure 1. Standard error of prediction of final grade placement in mathematics. (Parameters for the selected trajectory models are reestimated at each successive collection date during the school year. The number of students present at different collection dates varies between 923 and 1,668.)

the fact that there is a larger mean and standard deviation for gains in the mathematics curriculum than the reading curriculum. For both sets of data the differences among models are more pronounced at the beginning of the year than at the end of the year.

Of the models that were not graphed, Model 2, the model with a correction for start time, did worst of all. Its errors were often from .1 to .2 grade years greater than the other models. Models 3, 6, and 7, the models corrected for beginning knowledge, were also omitted from the graph. Of this group, Model 3 generally did best, but seldom as well as the models on the graph.

As an indication that time does, in fact, influence the grade-placement predictions, we note that the group b coefficients in Model 9 are about .001 (.0012 in math and .0008 in reading). Since the average total lesson time in a year is about 1,000 minutes (e.g., 989 minutes in math), this indicates a predicted grade-placement gain of about

$(1,000)(.001) = 1.0$ grade years in 1 school year. This indicates that time does indeed have a significant effect on the grade placement predictions.

Errors in Curve Fitting

Even though we are primarily interested in how well our different methods predict the final grade placement, we can also look at how well they fit all the observed points throughout the year. To do this we computed the mean standard error of fit for each collection in a manner similar to our computation of the standard error of prediction. For each student we calculated a measure of curve-fitting error, e_i , as follows:

$$e_i = \left[1/m \sum_j [\hat{GP}_i(t_j) - GP_i(t_j)]^2 \right]^{1/2},$$

where m is the total number of observed points at the time of collection m for student i . We then computed the mean standard

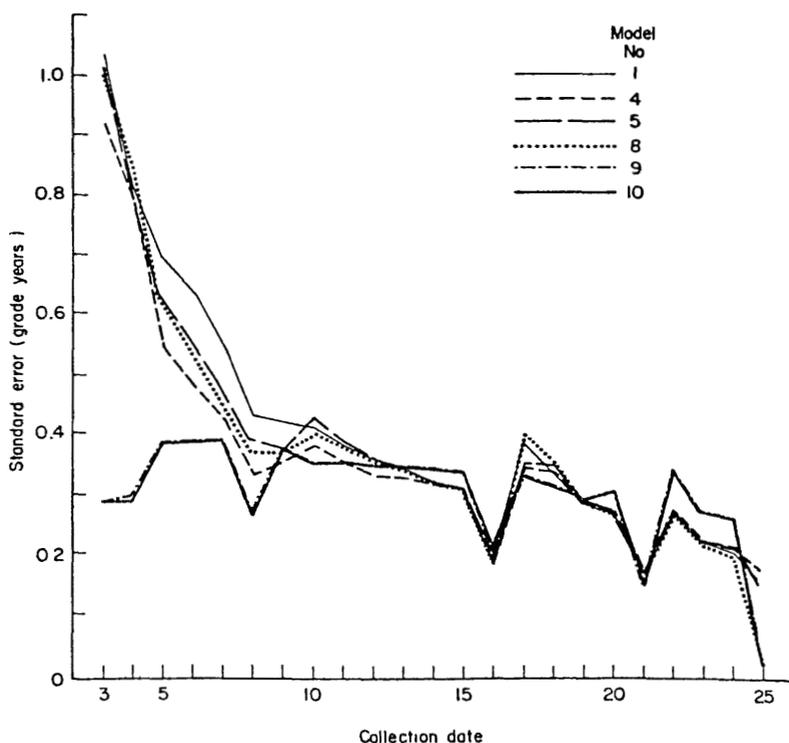


Figure 2. Standard error of prediction of final grade placement in reading. (Parameters for the selected trajectory models are reestimated at each successive collection date during the school year. The number of students present at different collection dates varies between 729 and 1,637.)

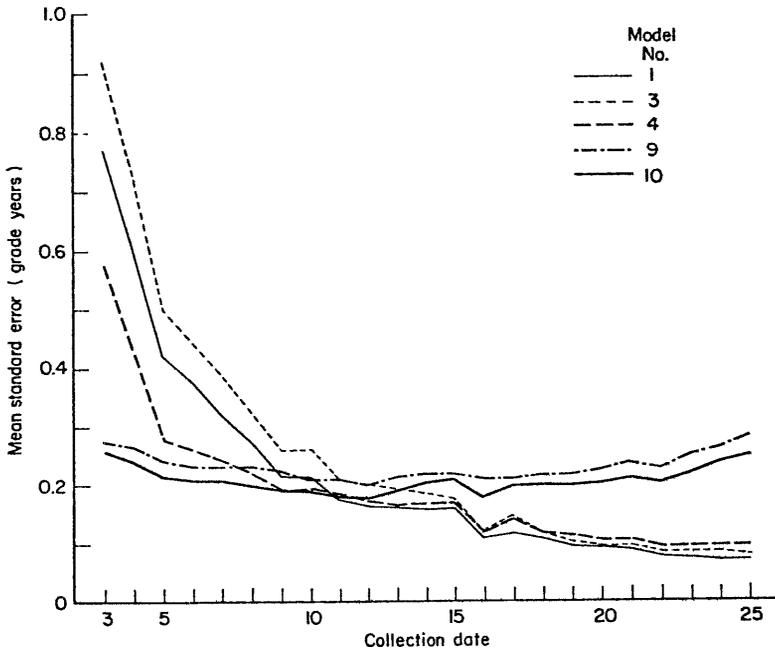


Figure 3. Mean standard error of curve fitting for mathematics. (Parameters for the selected trajectory models are reestimated at each successive collection date during the school year. The number of students present at different collection dates varies between 923 and 1,668.)

error over all students by averaging the individual standard errors:

$$e = 1/n \sum_i e_i,$$

where n is the number of students.

The results of these error computations are shown in Figure 3 for the mathematics curriculum and in Figure 4 for the reading curriculum. Although we saw above that Models 9 and 10 are the best predictive models, we see here that they do not in general give as good a fit to either the mathematics or the reading data as the other models. In fact, when 20 or more points have been observed, the fits given by Models 9 and 10 are much worse than the fits given by all other models. Evidently the extreme sensitivity of the piecewise models to the student's most recent point makes the models too inflexible to accurately fit all the points.

When the mean standard error over all points at the last collection date is used as the measure of the descriptive power of the models, the data show that Models 1 and 3 are the most powerful. For the mathematics

data, Model 1, the power-function model with three individually estimated parameters, is better than Model 3, whereas for the reading data, the results given by the two models are almost identical. The order of the remaining models in terms of mean standard error of last collection date is nearly the same for both courses. For the mathematics data, Model 4 is the next best followed in order by Models 10 and 9; for the reading data, the order is the same except that Model 9 is better than Model 10. The five models omitted from these graphs (Models 2, 5, 6, 7, and 8) generally fell between Model 4 and Models 9 and 10. Models 9 and 10, though they seem to do worse overall, do better than the other models near the beginning of the year. Similar to the result for prediction, the differences among models for curve fitting are not as great for the reading data as for the mathematics data, and the differences among models are greater at the beginning of the year than at the end of the year.

We used the mean standard error over all points at the last collection to determine the best fixed k for Models 4, 7, and 10. Figures

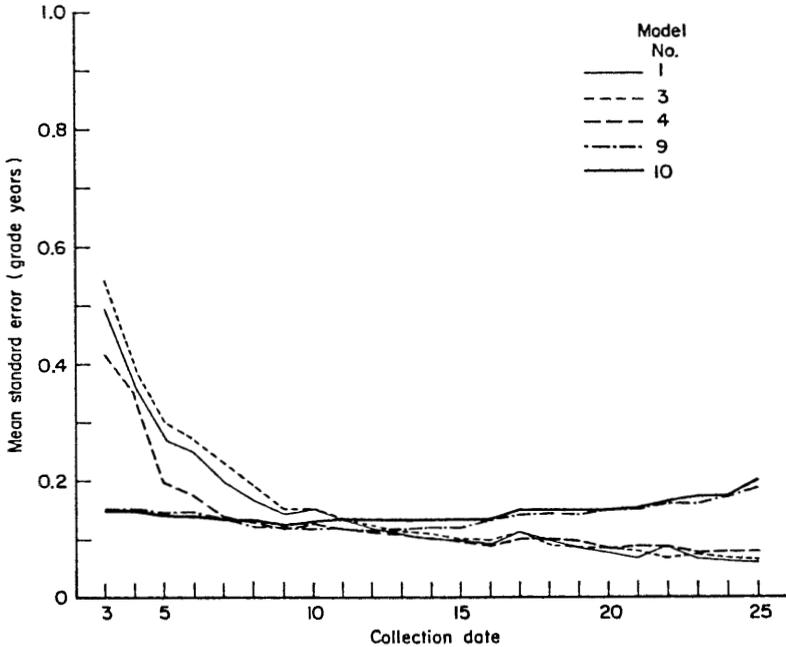


Figure 4. Mean standard error of curve fitting for reading. (Parameters for the selected trajectory models are reestimated at each successive collection date during the school year. The number of students present at different collection dates varies between 729 and 1,637.)

5 and 6 show the mean standard error over all students given by Model 4 for 100 values of k between 0 and 1. Figure 5 shows that for the mathematics course, this error is minimized at exactly $k = .50$, where it has a value of .098. In Figure 6, for the reading course, the mean standard error is minimized at .79, where it has a value of .070. But the curve in Figure 6 is relatively flat close to the minimum, and $k = .50$, which exactly minimized the error in the mathematics data, almost minimizes it for the reading data as well. Therefore, we used a

fixed $k = .50$ in Models 4, 7, and 10 for both curriculums.

Time Prediction

In our predictions of final grade placement, we were able to use observed values for the amount of CAI time each student had received by the end of the year. If we were actually making the predictions before the end of the year, we would, of course, not know the student's final time, and we would have to estimate it. Naturally, this addi-

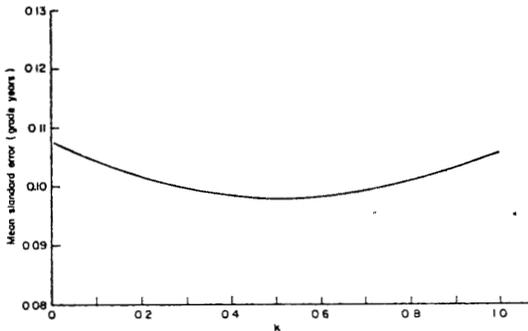


Figure 5. Mean standard error of fit at last collection for various values of k in Model 4 for mathematics.

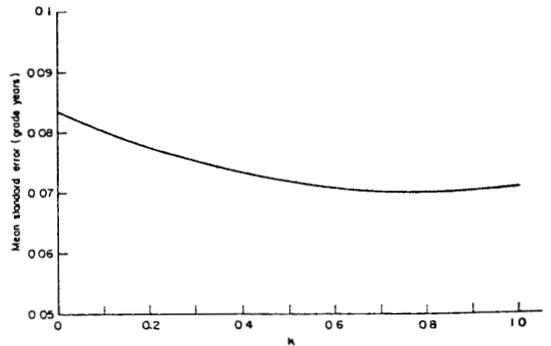


Figure 6. Mean standard error of fit at last collection for various values of k in Model 4 for reading.

tional estimation would increase the standard errors. We will now explore several methods by which we might estimate final time.

Even though the differences in time across students are caused only by different absence rates, in practice we would have to predict these differences. Unfortunately for the sake of simple prediction, our data show a very wide variation in time among students. For example, in the mathematics data we found a mean time of 989 minutes with a standard deviation of 342. Our methods of estimating time must predict this wide variation.

One way to predict amount of time would be to determine a student's percentile rank in the school or district according to the amount of time the student had received by the last observation period and then to estimate final time for that student by using the time observed in a previous year for a student with the same percentile rank. This prediction procedure depends on the assumption that the amount of time a student receives in comparison with the rest of the class remains relatively constant throughout the year.

In order to test this assumption, we created tables of the percentile ranks of the amount of time accumulated in the mathematics curriculum by each student at each collection date. Except for the upper 10%–20% of the distribution, these percentile ranks were relatively unstable with respect to the final percentile ranks until well past the middle of the year. The complete tables are too lengthy to publish here, but we give a sense of the magnitude of the instability in Table 2, which displays the percentile ranks for a sample of five students. This large amount of instability means that a student's percentile rank during the year is not a good predictor of the student's final percentile rank, and therefore this method would not be a good method for predicting the final amount of time.

After rejecting the percentile method of predicting time, we considered the following three alternative methods:

1. *Mean.* Each student in a group is predicted to receive the mean amount of time received by some other comparable

Table 2
Percentile Ranks for Accumulated Time for Students by Collection Date

Collection date	Student				
	1	2	3	4	5
2	42	15	28	82	54
3	42	11	24	86	47
4	33	4	30	90	50
5	20	5	41	92	51
6	10	19	55	98	60
7	7	20	62	96	60
8	7	21	65	98	63
9	4	21	67	99	62
10	3	17	65	100	64
11	2	19	64	100	65
12	2	21	66	100	64
13	1	22	63	99	68
14	2	23	66	100	72
15	2	25	64	99	76
16	2	23	65	99	76
17	2	24	66	99	79
18	1	38	73	99	82
19	2	43	72	100	81
20	1	49	73	100	81
21	1	46	75	100	79
22	1	49	75	100	79
23	1	48	77	100	80
24	1	54	75	100	80
25	1	50	75	100	83

group or by the same group in a previous year.

2. *Proportional.* Each student is predicted to receive the same proportion of the total scheduled time throughout the year as he has received so far during the year. That is, for each student i ,

$$t_f = (t_j/c_j)c_f,$$

where t_f = final amount of time for student i , t_j = amount of time at collection date j for student i , c_f = number of collection dates scheduled by the end of the year for student i , and c_j = number of collection dates for which student i was enrolled by collection date j .

To understand the rationale behind the use of c_f and c_j , note that if we assume that the number of scheduled minutes for each student is the same for each collection period, then c_f/c_j is proportional to the number of minutes scheduled for the whole year divided by the number of minutes scheduled by collection j . We also note that because

the data did not include several collection dates at the beginning of the year, we used t_1 for each student to estimate c_1 for that student. This estimate of c_1 was included in all the subsequent c_j s.

3. *Self-correcting proportional.* Each student is predicted to receive the same proportion of the total scheduled time throughout the year as he has received so far, but this prediction is adjusted by the amount it was incorrect in predicting the time at the present collection date from a previous collection date. This is formalized as follows:

$$t_f = a_0 + a_1(t_j/c_j)c_f,$$

where a_0 and a_1 are regression coefficients which best fit for all students the following curve:

$$t_j = a_0 + a_1(t_{j-k}/c_{j-k})c_j, \\ (1 \leq k < j).$$

Thus the errors in the predictions from k collections ago are used to adjust the current predictions for final time.

Table 3 shows for 6 collection dates the standard error in predicting final time in mathematics for each student according to each of the three methods. The standard error varies from a maximum of 963 minutes at the beginning of the year to a minimum of 23 minutes just before the end of the year. The mean prediction method appears to be the best of the three in the beginning of the year, the self-correcting proportional is best

Table 3
Standard Error (in minutes) for Three Methods of Predicting Final Amounts of Time in Mathematics

Collection date	Method		
	Mean	Proportional	Self-correcting Proportional
3	340	876	820
5	340	963	561
10	340	301	216
15	340	137	104
19	340	185	208
24	340	23	58

Table 4
Standard Error of Prediction of Final Grade Placement in Mathematics Using Model 10 and Various Time Prediction Methods

Collection	Observed time	Method		
		Mean	Proportional	Self-correcting Proportional
3	.468	.531	.800	.754
5	.397	.474	.838	.660
10	.334	.426	.429	.421
15	.275	.363	.314	.298
19	.162	.361	.238	.251
24	.064	.361	.066	.079

Note. The number of students present at different collections varied between 698 and 926.

in the middle of the year, and the proportional is best toward the end of the year.

In the first part of this article, we predicted final grade placement assuming that we knew the total amount of time a student would receive. Now, however, we are also trying to predict a student's total time in order to predict final grade placement. Table 4 shows how this additional prediction decreases our predictive accuracy for mathematics; comparable results were obtained for reading. In this table, the standard errors of the piecewise power function predictions (Model 10) using the three time prediction methods are compared with the predictions we obtained using the observed final times.

For the mathematics data we can predict final grade placement after the 10th collection almost as well using either of the proportional methods for predicting time as when we used the observed final times. The standard errors are increased by no more than .1 when time is predicted.

Discussion

The fact that the simpler models give better predictions of end-of-year achievement may be due in part to our use of the models to project far outside the range of observations. The more parameters that are

available to fit the observations, the more sensitive the curve is to small random fluctuations in the data, and therefore the more radically it can be wrong outside the range of the data. For example, if we have three points generated by small random variations above and below a straight line, we can fit the points much better with a quadratic curve than with a line. (In fact, we can fit three points exactly with a quadratic curve.) But outside the range for which data are available, the quadratic curve will be very different from the underlying linear distribution.

Another important feature of the piecewise models (9 and 10) is that they are constantly self-correcting. Since progress in the CAI curriculum contains a large stochastic element, both these models give great weight to the most recently observed point.

It is important to emphasize that the success of Models 9 and 10 in predicting final grade placement does not indicate that students all learn at the same rate. Rather, what we have shown is that the points we observe during the year seem to be too much affected by random variations for us to use them in making a reliable estimate of individual learning rates. We can, however, make reasonably good midyear predictions for the final grade placement of an individual using only an average learning rate for the group and the individual's most recently observed grade placement.

Conclusion

Throughout this article we have considered grade placement as a function of time spent taking CAI lessons. This is only one of several possible levels of analysis. We could also consider grade placement as a function of calendar time; that is, grade placement would be predicted according to time values that were estimated either explicitly or implicitly from calendar time. We have in fact suggested in the section on time prediction how this relationship between CAI time and calendar time might be explicitly formulated.

At a lower level of analysis we could consider grade placement as a function of the number of exercises worked, or at an even

lower level, as a function of the number of exercises worked correctly. But the number of exercises worked correctly is almost the same measurement as gain in grade placement. The exact nature of the relationship between grade-placement gain and number of correct exercises depends not on predictions about human learning but only on the exact criteria imposed by the curriculum designers for movement in the course.

In principle, one could choose to predict grade-placement gain from any of these levels: calendar time, CAI lesson time, total number of exercises worked, or number of exercises worked correctly. By predicting explicitly from any level, one is implicitly making predictions about all the lower levels. We have chosen to concentrate on predicting grade placement from lesson time here, since lesson time is the variable that can be most easily controlled in order to reach a grade-placement goal. After exploring the most plausible models from the family of general power function models, we found that the two models that were simplest and most convenient to use were also, happily, the best predictors for our curriculums.

The quantitative approach to instructional prediction that we have begun to develop in this article points the way to a broader and more powerful theory of instruction and thus more effective management of instructional resources. Large-scale practical applications of these ideas, however, will almost certainly require greater precision of measurement and prediction than we have so far achieved.

References

- Adkins, K., & Hamilton, M. *Teacher's handbook for reading, grades 3-6*. Palo Alto, Calif.: Computer Curriculum Corporation, 1972.
- Suppes, P., Fletcher, J. D., & Zanotti, M. Performance models of American Indian students in computer-assisted instruction in elementary mathematics. *Instructional Science*, 1975, 4, 303-313.
- Suppes, P., Fletcher, J. D., & Zanotti, M. Models of individual trajectories in computer-assisted instruction for deaf students. *Journal of Educational Psychology*, 1976, 68, 117-127.
- Suppes, P., Jerman, M., Kanz, G., Morningstar, M., Searle, B., & Clinton, J. P. M. *Handbook for mathematics strands, grades 1 through 6*. Palo Alto, Calif.: Computer Curriculum Corporation, 1975.

Received April 10, 1978 ■