JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2007, **88,** 103–113 NUMBER 1 (JULY)

TIME TO COMPLETION OF WEB-BASED PHYSICS PROBLEMS WITH TUTORING

RASIL WARNAKULASOORIYA[1], DAVID J. PALAZZO[1], AND DAVID E. PRITCHARD[1,2]

DEPARTMENT OF PHYSICS & RESEARCH LABORATORY OF ELECTRONICS,
MASSACHUSETTS INSTITUTE OF TECHNOLOGY,[1] AND
EFFECTIVE EDUCATIONAL TECHNOLOGIES[2]

We studied students performing a complex learning task, that of solving multipart physics problems with interactive tutoring on the web. We extracted the rate of completion and fraction completed as a function of time on task by retrospectively analyzing the log of student–tutor interactions. There was a spontaneous division of students into three groups, the central (and largest) group (about 65% of the students) being those who solved the problem in real time after multiple interactions with the tutorial program (primarily receiving feedback to submitted wrong answers and requesting hints). This group displayed a sigmoidal fraction-completed curve as a function of logarithmic time. The sigmoidal shape is qualitatively flatter for problems that do not include hints and wrong-answer responses. We argue that the group of students who respond quickly (about 10% of the students) is obtaining the answer from some outside source. The third group (about 25% of the students) represents those who interrupt their solution, presumably to work offline or to obtain outside help.

Key words: time to completion, problem-solving, physics, logarithmic time, sigmoidal functions, hints, web-based problems, humans

---

The time necessary to complete a task is an important property of individuals' interaction with that task. The importance of time is well recognized in education: years required to obtain a degree, days of school per year, hours of class time per semester, and minutes on a given test are common metrics. Time to completion is arguably of equal importance with regard to homework tasks since homework often dominates the total time spent on a course especially at higher grade levels (Cooper, 1989). Unfortunately, the unpredictable location and time at which students actually do their homework combined with its unsupervised nature have made studies of homework time impractical. Online homework administration and tutoring programs now provide opportunities to collect time data on homework tasks. This study represents an initial investigation of such data and provides novel insight into student behavior on homework.

It is well known that ''time on task'' correlates strongly with amount learned (Long, 1983; Wellman & Marcinkiewicz, 2004). However, since students can spend only a limited amount of time on homework, it is important to study the allocation of this important resource. Studies of the time taken to complete homework tasks are necessary to optimize learning per unit time, and are valuable for other reasons, as well: identifying tasks that in the instructor's judgment are taking too long; enabling prediction of the time students will take on a test or on the entire homework assignment; and identifying students who may be receiving outside assistance with the tasks or who allocate their time inefficiently.

Although detailed studies of time spent on homework tasks are sparse, the time to complete a task, and the rate at which this time is reduced by instruction (often repetitive practice)—called the ''learning curve''—are common variables in the psychology of learning involving both human and non-human animals (Coombs, Dawes, & Tversky, 1970; Luce, 1986). Performing simple tasks quickly is an indication of skill in the task domain; hence we might expect time to completion of homework problems to be an indication of skill in the associated knowledge domain.

The authors wish to thank the reviewers for their helpful comments and suggestions. This work was funded by the National Science Foundation under grant 0231268.

Mastering Physics (www.masteringphysics.com) is marketed by Addison Wesley and was made by Effective Educational Technologies (effedtech.com), of which David Pritchard was founder and in which the Pritchard family had a controlling interest when this research was conducted.

Correspondence concerning this article should be addressed to Rasil Warnakulasooriya, 71A Norwood Avenue, Newton, Massachusetts 02459. (E-mail: rasilw@gmail.com).

doi: 10.1901/jeab.2007.70-06

Here we investigate the time individual students take to complete tutorial problems in a physics online homework environment. ''Tutorial problems'' here means multipart problems that contain both spontaneous (i.e., nonrequested) responses to incorrect answers and requestable hints that help the students along an expert route to solution. The hints provided upon request for a given part of a multipart problem are basically subtasks; answering these subtasks would guide the student to the solution of that part.

METHOD

The study involved about 400 undergraduate students taking the ''Introductory Newtonian Mechanics'' course, 8.01, at the Massachusetts Institute of Technology (MIT) in fall 2003. The students were assigned homework problems (we argue later that the problems are not exercises for most students) using Mastering Physics, a web-based homework tutor. Students could work on the assigned problems at any time over a several-day period prior to the due date. There was no requirement that the problems be completed in one sitting. All students in this study were given identical homework assignments for the first six weeks of the class. After the first six weeks, students were randomly assigned to two subclasses or groups so that the mean scores on the first six assignments were not statistically significantly different between the two groups. The 12 problems on which this study is based were then given in six pairs over the rest of the course of the semester. The two groups were required to solve the two problems in each pair sequentially, but in opposite orders. For example, if group 1 had solved problem A first and then problem B in a given pair, then group 2 had to solve problem B first and then problem A. Two groups, each solving six pairs of problems, thus results in 24 instances in total. This particular experimental design enabled another study on knowledge transfer, but has no bearing on the present study.

Most of the assigned problems contained tutoring material that enabled a Socratic dialogue in which students were provided with hints and simpler subproblems (subtasks) upon request, and were given specific feedback to the majority of incorrect answers that they proposed. This pedagogy approximates mastery learning (Bloom, 1980) in that less skillful students travel a longer path, but eventually over 94% of all students obtained the correct solution. Guessing behavior was discouraged by eschewing the multiple-choice format in favor of free-response answer types for a majority of all requested responses (where students had to type in a symbolic or numerical response). Furthermore, unopened hints were given a 3% bonus per part to promote thinking before requesting help. End-of-chapter (EOC) problems based on Young and Freedman (2004) that did not contain hints or feedback also were administered through Mastering Physics in order to allow for a comparison with similar problems that did contain tutoring as described above.

Student interactions with Mastering Physics were analyzed for time to completion, number of hints requested, and number of wrong answers submitted. Time to completion was defined as the time duration between the first opening of a problem and the submission of the completed problem. No allowance for interactions in between (even for logins and log-offs) was made. Thus, completion time may not be the continuous time interval the student spent explicitly solving the problem. However, we argue later that the completion times for a subset of students we call real-time solvers is usually the uninterrupted time they spent on solving that problem. Completion was defined as finishing all the parts of a given problem correctly. (It was not required that students solve requested subproblems, although there was a penalty for requesting the solution to them.)

RESULTS

Time to completion ranged from several seconds to several days for the problems studied. Thus, we used logarithmic time ($\mathrm{Ln}(t)$, where t is measured in s) as the independent variable, which ran from 2 to 13, corresponding to a time scale from about 7 s to 5 days. We chose to bin the data in time intervals of 0.5 in logarithmic time scale (a factor of 1.6 in real time). We studied both the ''rate of completion,'' which is the fraction of students completing the problem in a unit interval in $\mathrm{Ln}(t)$ or t and its time integral, that is, the fraction of students who have completed the problem within a given time interval

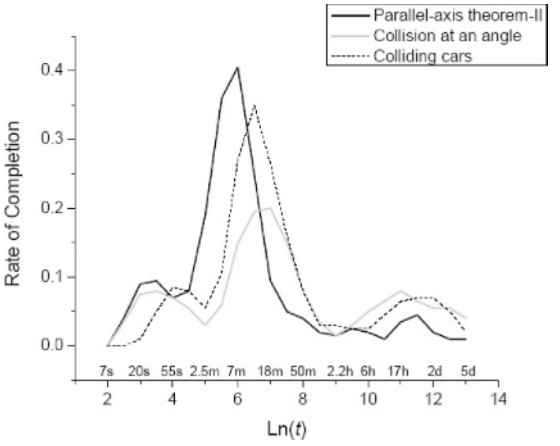

Fig. 1. The rate of completion (also the derivative of fraction complete with respect to log time) for three representative problems. The plots reveal three distinct groups. The three-group structure is preserved for each problem although each problem has a different median time as measured by the central peak.

after opening the problem. (The fraction is relative to all the students who had opened the problem.)

Figure 1 shows the rate of completion as a function of log time for three representative problems. The rate of completion was characterized by a three-peak pattern that was common to all the 12 problems that contained help (requestable hints, feedback, and advice within the problem text). The first peak consists of ''quick responders'' who predominantly answered correctly on the first attempt with a median time to completion of 1.4 ± 0.3 min. The central peak consists of the majority of the students who are ''real-time solvers.'' The median time to completion for real-time solvers is 16.9 ± 8.6 min. There is a local minimum, typically at about 2.5 min between the quick responders and the real-time solvers. Finally, the third peak corresponding to the ''delayed solvers'' is separated from the real-time solvers peak by a broader local minimum, which occurs at about 2.2 hr. These delayed solvers display long periods of inactivity, suggestive of absence from their computer terminal. Their median time for problem completion is 21.2 ± 5.8 hr. We shall argue that most members of the quick-responder and some of the delayed-solver groups receive help outside the web-based tutor. Results from 10 of the 12 problems suggest that the time duration from 2.5 min to 2.2 hr is a reasonable duration in identifying the real-time solvers marked by the prominent central peak. (For two problems involving gravitation, the central peak of the real-time solver duration shifted to later times and was between 7 min and 6 hr.) The median time to completion, hints per user, incorrect responses per user, and the interactions per user per problem for the three groups of students are given in Table 1.

Figure 2 shows the breakdown of a typical rate-of-completion curve by the type of interaction for all three groups. The students who fell within the dominant central peak (2.5 min to 2.2 hr) interacted with the tutoring features of the online environment, justifying their classification as real-time solvers. Only about 15% answered all parts of the question without error. The remainder divided into two subgroups of students – those who only submitted wrong answers (to which feedback was given), and those who also requested hints. Those who also requested hints took about 4% longer to finish the problem. For all real-time solvers, the correlation between log time to completion and the number of wrong answers submitted plus hints requested was 0.25 on average per problem, likely reflecting the fact that making wrong answers and asking for hints takes additional time. The delayed-solvers also consisted of students who made mistakes and requested hints. In contrast, the quick-responders consisted primarily of students who did not make mistakes and

Table 1

Means and standard errors of four behavioral variables for the quick responders, real-time solvers, and delayed solvers.

	Quick responders	Real-time solvers	Delayed solvers
Median time	1.4 ± 0.3 min	16.9 ± 8.6 min	21.2 ± 5.8 hr
Hints/user	0.4 ± 0.1	2.7 ± 0.6	3.8 ± 0.9
Incorrect/user	0.7 ± 0.1	4.1 ± 0.5	5.2 ± 0.8
Interactions/user/problem	6 ± 1	13 ± 2	15 ± 2

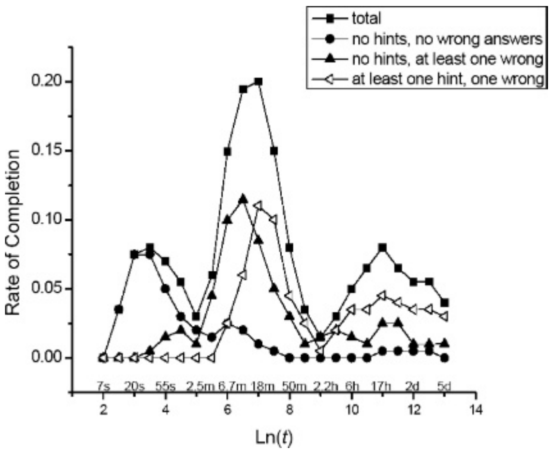

Fig. 2. Behavior underlying a typical rate-of-completion curve for one of the problems. The total curve is shown by the filled squares; the other three curves show the breakdown of the total curve depending on whether there were hint requests and wrong answer submissions.

did not requested hints. These data suggest that the quick responders behave differently from the real-time and delayed solvers.

The large sample size (n ~ 400) allows us to study the overall shape of the fraction-complete curves for the real-time solvers. As may be seen in Figure 3, the fraction-complete curves for the real-time solvers (the integral curves of the rate of completion curves) have a sigmoidal shape, but only when plotted against logarithmic time. This relation was true for all 24 instances with tutorial help.

We explored two aspects of the sigmoidal curves in detail: first, which particular functional form of the sigmoidal curve best represented the data, and second, whether the use of $\mathrm{Ln}(t)$ as the independent variable was statistically significant. We considered three functional forms: Logistic (Equation 1), Boltzmann (Equation 2; which also has the form of a Logistic function), and Gompertz (Equation 3). The parameters a and b in the logistic model correspond to the lower and the upper asymptotes, respectively, while p serves as a shape parameter. In the Boltzmann model, the parameters a and b again correspond to the lower and the upper asymptotes with the parameter d serving as the shape parameter. At the center of the curve, the slope of the function is given by $(b\text{-}a)/4d$. In the Gompertz model the lower asymptote is zero while the upper asymptote is given by the

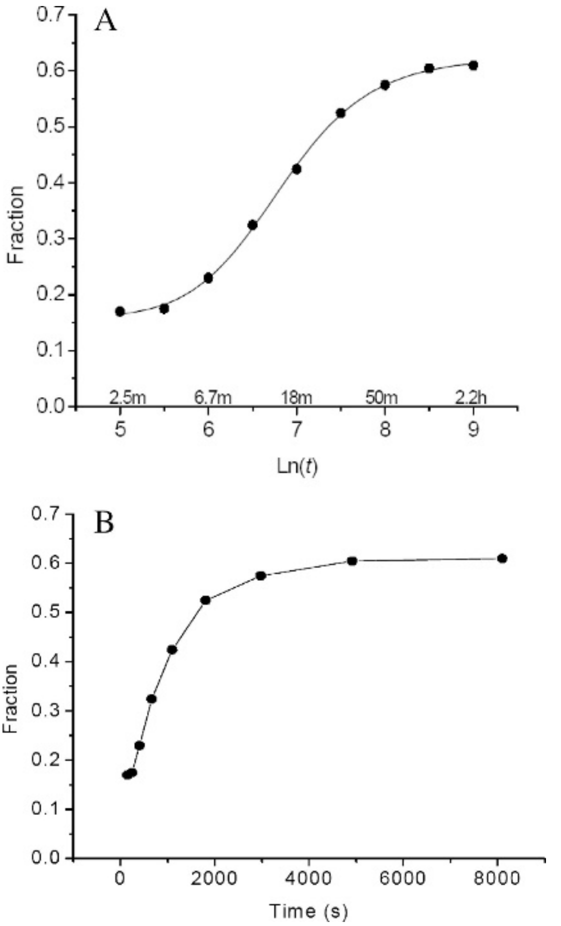

Fig. 3. A time to completion curve in log time (A) and linear time (B) for the real-time solvers for a typical problem.

parameter a with the shape being determined by k. For this model we find that the slope at the center of the curve is given by ak/e. The shape parameters in the models account for the rate of completion of the problems.

$$\text{Logistic} \quad f_L(x) = \frac{a - b}{1 + \left(\frac{x}{x_C}\right)^p} + b \qquad (1)$$

$$\text{Boltzmann} \quad f_B(x) = \frac{a - b}{1 + e^{(x - x_c)/d}} + b \qquad (2)$$

$$\text{Gompertz} \quad f_G(x) = ae^{-e^{-k(x-b)}} \qquad (3)$$

where, $x = \mathrm{Ln}(t)$ or t when the independent

Table 2

Differential model fits with independent variable Ln(t). The ΔAIC values and evidence ratios (ER) are given relative to the best-fit model (which is labeled * for ''best''). The median ΔAIC and reduced chi-squared (χ_v^2) values are shown at the bottom. Related problem pairs are shown by normal letter pairs and italicized pairs alternatively, although the fact that they are related is irrelevant to this study. Instance refers to a particular group of students for whom the problem has been assigned although it also does not have any bearing for the present study (see text).

Problem	Instance	Boltzmann ΔAIC	Boltzmann ER	Boltzmann χ_v^2	Logistic ΔAIC	Logistic ER	Logistic χ_v^2	Gompertz ΔAIC	Gompertz ER	Gompertz χ_v^2
colliding cars	I1	2.1	3	1.072	0	*	0.851	8.6	74	2.213
	I2	~0	1	1.488	0	*	1.488	7.4	39	3.361
collision at an angle	I1	0	*	0.142	7.1	35	0.313	23.2	>1000	1.866
	I2	1.4	2	0.522	0	*	0.449	9.8	137	1.342
mass-spring system	I1	0	*	0.858	3.0	5	1.199	13.5	860	3.853
	I2	3.0	5	1.416	0	*	1.014	5.2	14	1.812
shooting a block	I1	0.3	1	0.889	0	*	0.858	8.2	61	2.142
	I2	9.8	136	3.884	6.0	20	2.549	0	*	1.304
person on a ladder	I1	0	*	1.184	3.0	4	1.652	10.1	156	3.637
	I2	0	*	0.459	3.7	6	0.689	11.1	262	1.582
calculating torques	I1	0	*	0.383	0.8	2	0.418	20.6	>1000	3.788
	I2	0	*	1.020	5.2	13	1.807	15.9	>1000	5.946
asteroid impact	I1	0.7	1	1.370	0	*	1.264	5.3	14	2.264
	I2	~0	1	2.755	0	*	2.752	3.6	6	4.109
post-collision orbit	I1	2.1	3	2.539	0	*	2.015	0.6	1	2.154
	I2	0	*	1.501	0.85	2	1.648	7.8	50	3.586
flywheel kinematics	I1	1.8	2	0.837	0	*	0.688	10.9	232	2.307
	I2	1.4	2	1.335	0	*	1.148	6.3	24	2.315
angular motion	I1	1.3	2	1.084	0	*	0.937	7.7	46	2.199
	I2	0.8	2	0.631	0	*	0.580	6.9	33	1.258
parallel-axis theorem-I	I1	0	*	1.939	1.7	2	2.335	8.5	69	4.969
	I2	0	*	4.404	2.3	3	5.686	7.7	48	10.404
parallel-axis theorem-II	I1	1.5	2	1.398	0	*	1.181	3.9	7	1.831
	I2	2.0	3	2.230	0	*	1.778	~0	1	1.780
Median		0.5		1.259	0		1.190	7.8		2.238

variable is logarithmic or linear time, respectively.

We used two statistical tools to explore the question of which function of logarithmic time would best fit the data: the chi-square goodness-of-fit coupled with the second-order Akaike Information Criterion or AIC_c (Akaike, 1973; Sugiura, 1978), which is a relation between the Kullback-Leibler distance (Kullback & Leibler, 1951) and the maximized log-likelihood functions. According to Burnham and Anderson (2002), ''AIC provides a simple, effective, and objective means for the selection of an estimated 'best approximating model' for data analysis and inference'' (p. 2). AIC penalizes a model with more parameters thus guarding against overfitting (Pitt & Myung, 2002). To explore the goodness-of-fit of the above functions to the data, we considered the differentials of the functions in logarithmic time. The differentials of the functions rather than the functions themselves were considered so that each point (i.e., time interval) affects the result independently of the others (i.e., does not depend on the sum of points up to that time). The above functions were then fitted in the range 2.5 min to 2.2 hr (i.e., for real-time solvers) assuming a Poisson error variation for each point (i.e., error proportional to the square root of the number of students in that time bin) to the rate-of-completion data. The reduced chi-square statistic, the change in AIC with respect to the best-fitting function (i.e., the function that has the lowest AIC value), and the evidence ratios (i.e., the relative likelihood of the best-fitting model relative to another) were then calculated (see Table 2).

The Logistic and Boltzmann models are significantly preferred for our data relative to the Gompertz in 23 of 24 cases. The median reduced chi-squared (χ_v^2) is of the order of 1, suggesting that the Logistic and Boltzmann functions account for essentially all information in the data. This judgment is confirmed by noting that in cases where the reduced chi-

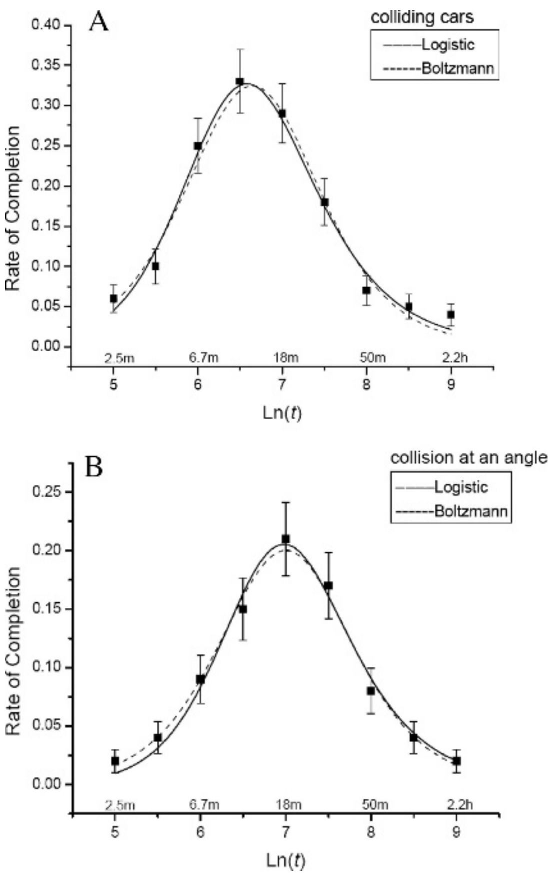

Fig. 4. Two cases where the Logistic and the Boltzmann fit better than the other in real-time solver region. In A, the Boltzmann lies below the right-most points whereas in B the Logistic lies below the left-most points. The points in between are more or less well accounted for by both models in both figures.

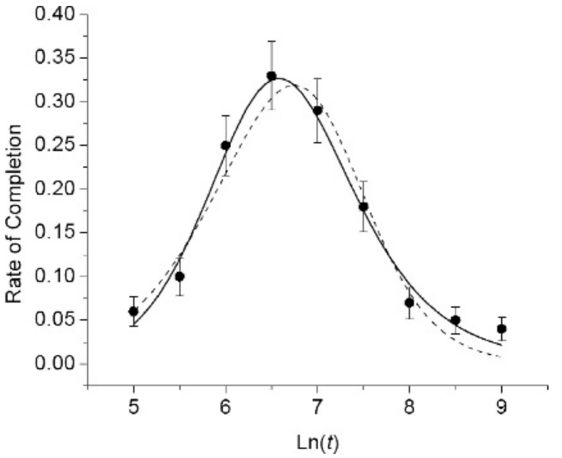

Fig. 5. Rate of completion (fraction complete) as a function of completion time for a representative problem. The dashed curve represents the fit of the differential function to the data if the fraction complete is sigmoidal in linear time ($\chi^2 = 11.2$), and the solid curve represents the fit if the fraction complete is sigmoidal in logarithmic time ($\chi^2 = 5.1$). Error bars represent ±1 standard deviation.

square is ~2 or greater, the data lie above the curve at one or the other end of the range, a likely indication that some students belonging to the quick responders or the delayed solvers spill over into the real-time solver group (Figure 4).

Sigmoidal (S-shaped or psychometric) curves have been observed in various psychological tasks (Bizo & White, 1994; Hambleton & Swaminathan, 1985; Hull, 1943; Machado & Guilhardi, 2000; Machado & Keen, 1999; Schnipke & Scrams, 1997; Tulving, Mandler, & Baumal, 1964), but as a function of linear rather than logarithmic time (i.e., taking $x = t$ in Equations 1 – 3). To explore whether the above functions (Equations 1 – 3) would fit our data better in logarithmic time or linear time, we again considered the differentials of the above functions both in logarithmic and linear time for the real-time solvers assuming a Poisson error variation at each point. We expressed the linear time differential function (df/dt) in terms of logarithmic time by using the relation $(df/dt) = (1/t)^* \, df/d(\mathrm{Ln}(t))$. This allowed us to use a common scale ($\mathrm{Ln}(t)$) and a single rate-of-completion data set for a given problem for comparisons of the fits. The linear time differential fit was then compared to the logarithmic time differential fit of the best fitting function using both the chi-square statistic and the AIC (see Figure 5).

For the 24 cases, the median superiority of the logarithmic time differential function in (unreduced) chi-squared was 3.95, and the average AIC evidence ratio of the logarithmic time differential function to that of the linear time differential function was 13.5—a considerable difference. Although there were four cases for which the linear time function provided a slightly better fit than the logarithmic time function, their median chi-square difference was 0.9 and the AIC evidence ratio was 2—not enough to rule out the logarithmic time (or accept the linear time) for these minority cases. In summary, the differential model in logarithmic time was statistically

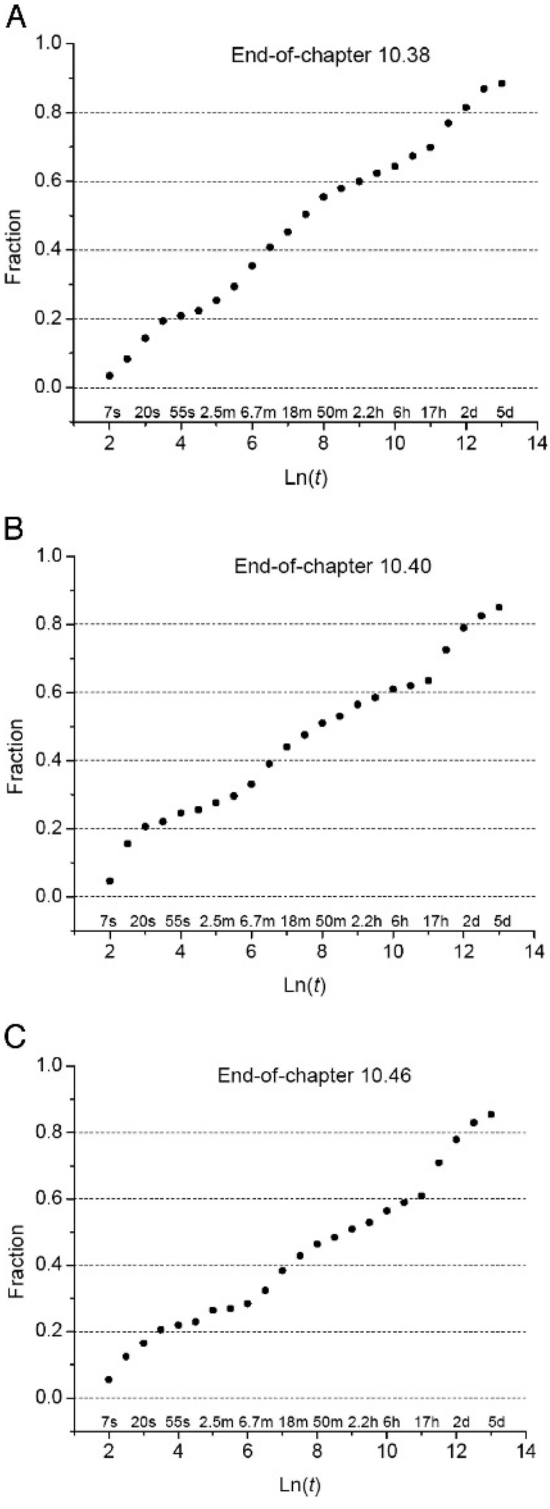

Fig. 6. Fraction complete curves for three representative End-of-chapter (EOC) problems. The numbers in the headings show the problem number in Young & Freedman (2004).

superior than in linear time, in accord with the fact that the prominent symmetrical 'S' shape is present to the eye only with respect to logarithmic time (see Figure 3).

A further indication of the significance of the shape of the fraction-complete curves comes from a comparison of the above curves with those obtained on problems that lacked the tutoring described above. These problems are typical end-of-chapter (EOC) problems that do not contain hints or spontaneous responses to specific wrong answers (i.e., the student just receives the message "try again"). Several such curves are shown in Figure 6. It is obvious that these curves lack the steep central part observed for tutorial problems (see Figure 7). In fact, the fraction of students who completed tutorial problems within the real-time solver interval was 65 ± 4% for 10 problems (not including the two gravitational problems for which the "real-time solver" interval was between 7 min and 6 hr) compared to 33 ± 5% for five EOC problems within the same time interval. This reduction in the fraction for EOC problems may indicate that some students are "getting stuck" and subsequently waste time or go away from their computers to seek help. The key point is that the shape of the curve reveals a strong qualitative difference that most probably arises due to the hints and wrong-answer responses in the tutorial problems (Simon & Reed, 1976).

DISCUSSION

This study shows that time to completion provides valuable insight into how students approach problems in their online homework assignments. Such an approach allows one to identify students who are solving a particular problem in real time and those who, usually after no real attempt at a solution, delay for perhaps a day their attempt of the problem (delayed solvers). Finally, we can identify students who apparently are obtaining the solution from outside sources without making an honest attempt to solve it (quick responders).

The quick responders differed from the real-time and delayed solvers in that the vast majority of the quick responders made no mistakes. We suggest that the "quick-responder" group is made up largely of students who

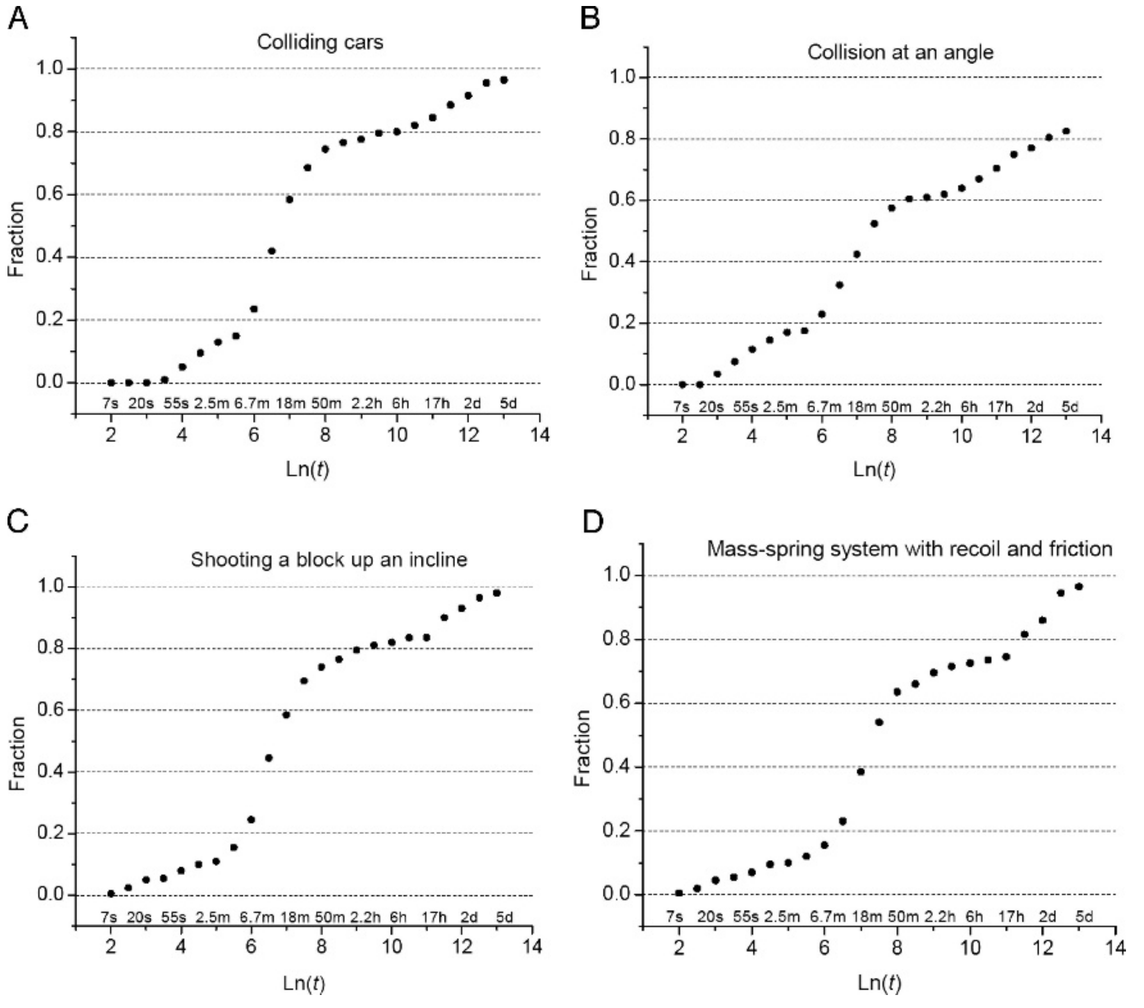

Fig. 7. Fraction complete curves for four representative tutorial problems. The fraction of real-time solvers (within 2.5 min – 2.2 hr) is about 65%.

obtain the answer from external sources without either inspecting or working the problem themselves. This suggestion is buttressed by three observations. First, skillful expert problem solvers in this domain attempting to work problems rapidly take longer than 2.5 min to solve the problems and display a higher error rate than the quick-responder students (the minimum at about 2.5 min likely reflects the minimum reading and comprehension time required for answering a given problem for students who are not quick-responders). For five problems on which the real-time solvers averaged 20.4 ± 5.3 min, one of the authors (DEP), working rapidly at the cost of making frequent mistakes, took 7.6 ± 2.5 min whereas the quick-responder students took only 1.6 ± 0.2 min on average and made few mistakes. Second, for most of the problems the time needed to read the problem statement and question and enter the answer exceeds 2.5 min. Finally, there was a negative correlation between solving a higher percentage of tutorial problems as a quick responder and doing well on the final exam. Figure 8 shows that students who responded quickly to more than 50% of the problems average 1.3 standard deviations below the class average in the paper-based final examination, hardly the behavior expected for students who are truly able to solve the majority of problems more quickly and with far fewer mistakes than either

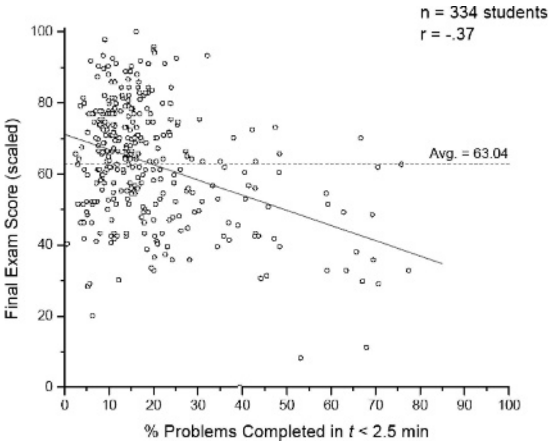

Fig. 8. Final exam score as a function of percentage of problems completed as a quick solver (i.e., in less than 2.5 min). The percentage is out of ~200 problems that were completed in less than 2.5 min during the semester.

the skilled instructor or the best real-time solvers. The opportunity directly to study homework copying (and how it develops in time, in response to poor performance, etc.) reveals the power of analyzing web-based homework and allows us to confirm and extend existing research on academic dishonesty that is based solely on self-reporting (Jensen, Arnett, Feldman, & Cauffman, 2002; Murdock, Hale, & Weber, 2001).

We suggest that by 2.2 hr, nearly everyone working continuously has given up while by then a few students who have received advice from friends are returning for a second attempt, accounting for the minimum observed at about 2.2 hr between the real-time solvers and the delayed solvers. With regard to the delayed solvers, only 39.6 ± 5.6% of delayed solvers gave any response (either correct or incorrect) to any main part of a problem before 2.2 hr. Hence, about 60% of the delayed solvers give their first response 2.2 hr or more after opening the problem. Also, over 65% of the interactions with the tutor for all delayed solvers took place after 2.2 hr. These interactions included hint requests, submitting solutions (correct/incorrect) to main or subparts, and requesting solutions. This pattern suggests that many delayed solvers decide to postpone their attack on the problem (possibly to seek some outside help) upon initial inspection of the problem, or possibly open it only to print it for later offline solution. Indeed, within the delayed-solvers group, there was a subset of students who made no interactions with the problem before closing and opening another problem. For ten of the problems, 35% of the delayed solvers apparently did this. When they returned to solve the bypassed problem, about 20% made no interactions with the tutor at all other than to submit the correct answer. The other delayed-solvers had multiple interactions. The mean duration of this problem completion period for this group of students was 2.3 ± 0.3 hr, just about the top of the transition from a real-time solver to a delayed solver, suggesting that they may have paused again to seek additional help.

The results shown in Figure 2 support our use of the term ''problem'' to describe the tasks, at least for the real-time and delayed solvers. In the problem-solving literature, a ''problem'' is distinguished from an ''exercise'' for which the route to solution is apparent to the student upon initial inspection (Kahney, 1993; Mayer, 2003; Smith, 1991). The observation that the vast majority of students submit at least one wrong answer, that typically half feel sufficiently confused to sacrifice some credit in order to receive a hint, and that they take substantially longer than expert solvers to solve the problem all argue that most are not able to treat the tasks as an exercise.

Within the real-time solvers group, the overall fraction of students completing a given problem followed a sigmoidal curve reminiscent of earlier studies with both humans and nonhuman animals using simpler tasks (Bizo & White, 1995; Hull, 1943; Hunt, Lunneborg, & Lewis, 1975; Machado & Guilhardi, 2000; Machado & Keen, 1999; Schnipke & Scrams, 1997; Tulving et al., 1964). Unlike those results in which the sigmoidal shape occurred in linear time, however, we find that the curves fit standard sigmoidal model functions better when logarithmic time rather than time itself is the independent variable. This is a novel finding. This difference may well indicate a specific way in which solving complex tasks like physics problems is different from simple tasks such as memory recall or pigeon learning tasks, which occur on a scale of several seconds or less (Bizo & White, 1994; Tulving et al., 1964). Furthermore, our results show additional discrimination in that the Boltzmann and Logistic models are significantly better fits than is the Gompertz.

In order to see the extent to which the sigmoidal functions considered in this article would fit the earlier data, we reanalyzed the data reported by Tulving et al. (1964) on human word recognition and sentence completion using a log time scale. The original data plotted the probability of a correct response as a function of linear exposure duration time measured in milliseconds. When plotted against logarithmic time (discounting the 0 ms exposure duration), the results were best fit by a linear model with typical evidence ratio > 100 with respect to the three models reported in this study except in the case of the zero-word context where Gompertz was favored ($\Delta AIC = 4.9$). This comparison suggests that the shape of the time-to-completion curves depends significantly on the type of task which in turn suggests the possibility that explicit mathematical models of individuals' learning and problem solving (Atkinson, Bower, & Crothers, 1965; Bizo & White, 1995; Coombs et al., 1970; Friedman, Massaro, Kitzis, & Cohen, 1995; Hull, 1943; Luce, 1986; Machado & Guilhardi, 2000) may provide connections between the types of processes and the shape of the time-to-completion curves. We are not aware of any learning model that gives $\mathrm{Ln}(t)$ as the natural independent variable.

We have no explanation for the occurrence of the sigmoidal shape in logarithmic time nor the preference seen for the Logistic and the Boltzmann models over the Gompertz. The functional forms considered in this article resemble growth curves that start, accelerate, and then gradually slow (or saturate). The logistic model (in linear time) occurs in the pacemaker model in the behavioral theory of timing (BeT) investigated in the context of pigeon responses (Bizo & White, 1995), whereas the Gompertz model has been successfully applied to growth of fish (Burnham & Anderson, 2002). Clearly, our results are in need of a theory on the connection between the mathematical shape of the fraction complete and the nature of the task, for example a mathematical model of the cognitive process of physics problem solving (Newell, 1990; Pirolli & Wilson, 1998).

The fact that the real-time solver group was twice as large on the tutorial problems as on the EOC problems (which do not have tutorial help and where there are more delayed solvers) is a strong suggestion that the tutoring dramatically reduces the time it takes students to work problems. The EOC and tutorial problems were not variants of the same problem, but they otherwise seem to us to be of comparable difficulty. This suggests that the hints and wrong-answer responses generate significant reduction in time to solution, and a more significant reduction in the fraction of delayed solvers (i.e., those who have to regroup and go elsewhere for help after getting stuck).

Finally, although the presence of delayed solvers for a tutorial problem may indicate that some students are unable to solve it with the aid of only the available tutorial hints and wrong-answer responses, it may be that some students elect to seek outside help rather than accept the hint penalty. Together with the fact that the shape of the time curves of EOC problems differs systematically from that of the tutorial problems, the presence of delayed solvers may allow authors of problems to identify (and rectify) problems for which the current hints and wrong-answer responses are inadequate.

A significant aspect of this study is that its results were obtained during the regular course of instruction in the area (homework) that consumes a large fraction of students' time and hopefully imparts a correspondingly large amount of learning. The ability to isolate the real-time solvers and track their actions offers an opportunity to make educational studies inexpensively with very large samples and a correspondingly high signal-to-noise ratio. Importantly, improvements (e.g., in learning per unit time) found in such studies will be directly applicable to future students' homework without requiring further development, class testing, and attention to issues of scalability and transferability.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Atkinson, R. C., Bower, G. H., & Crothers, E. J. (1965). *An introduction to mathematical learning theory.* New York: Wiley.

Bizo, L. A., & White, K. G. (1994). The behavioral theory of timing: Reinforcer rate determines pacemaker rate. *Journal of the Experimental Analysis of Behavior*, *61*, 19–33.

Bizo, L. A., & White, K. G. (1995). Biasing the pacemaker in the behavioral theory of timing. *Journal of the Experimental Analysis of Behavior*, *64*, 225–235.

Bloom, B. S. (1980). *All our children learning*. New York: McGraw-Hill.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer-Verlag.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs, NJ: Prentice-Hall.

Cooper, H. M. (1989). *Homework*. New York: Longman.

Friedman, D., Massaro, D. W., Kitzis, S. N., & Cohen, M. M. (1995). A comparison of learning models. *Journal of Mathematical Psychology*, *39*, 164–178.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York: Appleton-Century-Crofts.

Hunt, E., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology*, *7*, 194–227.

Jensen, L. A., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology*, *27*, 209–228.

Kahney, H. (1993). *Problem solving: Current issues*. Philadelphia: Open University.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79–86.

Long, D. R. (1983, November). *Time-on-task in beginning students of Spanish at the University level: A case study*. Paper presented at the annual meeting of the American Council on the Teaching of Foreign Languages, San Francisco, CA. (ERIC Document Reproduction Service No. ED 239514).

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University.

Machado, A., & Guilhardi, P. (2000). Shifts in the psychometric function and their implications for models of timing. *Journal of the Experimental Analysis of Behavior*, *74*, 25–54.

Machado, A., & Keen, R. (1999). Learning to time (LET) or scalar expectancy theory (SET)? A critical test of two models of timing. *Psychological Science*, *10*, 285–290.

Mayer, R. E. (2003). *Learning and instruction*. Upper Saddle River, NJ: Merrill.

Murdock, T. B., Hale, N. M., & Weber, M. J. (2001). Predictors of cheating among early adolescents: Academic and social motivations. *Contemporary Educational Psychology*, *26*, 96–115.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Pirolli, P., & Wilson, M. (1998). A theory of the measurement of knowledge content, access, and learning. *Psychological Review*, *105*, 58–82.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*, 421–425.

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, *34*, 213–232.

Simon, H. A., & Reed, S. K. (1976). Modeling strategy shifts in a problem-solving task. *Cognitive Psychology*, *8*, 86–97.

Smith, M. U. (1991). A view from biology. In M. U. Smith (Ed.), *Toward a unified theory of problem solving: Views from the content domains* (pp. 1–21). Hillsdale, NJ: Erlbaum.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics, Theory and Methods*, *A7*, 13–26.

Tulving, E., Mandler, G., & Baumal, R. (1964). Interaction of two sources of information in tachistoscopic word recognition. *Canadian Journal of Psychology*, *18*, 62–71.

Wellman, G. S., & Marcinkiewicz, H. (2004). On-line learning and time-on-task: Impact of proctored vs. unproctored testing. *Journal of Asynchronous Learning Networks*, *8*, 93–104.

Young, H. D., & Freedman, R. A. (2004). *University physics*. San Francisco: Addison Wesley.

Received: October 16, 2006
Final acceptance: March 14, 2007