

Off-Task Behavior in the Cognitive Tutor Classroom: When Students “Game the System”

Ryan Shaun Baker, Albert T. Corbett, Kenneth R. Koedinger, Angela Z. Wagner

Human-Computer Interaction Institute, Carnegie Mellon University
Pittsburgh, PA, USA

rsbaker@cmu.edu, corbett@cmu.edu, koedinger@cmu.edu, awagner@cmu.edu

Abstract

We investigate the prevalence and learning impact of different types of off-task behavior in classrooms where students are using intelligent tutoring software. We find that within the classrooms studied, no other type of off-task behavior is associated nearly so strongly with reduced learning as “gaming the system”: behavior aimed at obtaining correct answers and advancing within the tutoring curriculum by systematically taking advantage of regularities in the software’s feedback and help. A student’s frequency of gaming the system correlates as strongly to post-test score as the student’s prior domain knowledge and general academic achievement. Controlling for prior domain knowledge, students who frequently game the system score substantially lower on a post-test than students who never game the system. Analysis of students who choose to game the system suggests that learned helplessness or performance orientation might be better accounts for why students choose this behavior than lack of interest in the material. This analysis will inform the future re-design of tutors to respond appropriately when students game the system.

Categories & Subject Descriptors: H5.m. [Information interfaces and presentation (e.g., HCI)]: Miscellaneous. K3.m. [Computers and Education]: Miscellaneous.

General Terms: Experimentation, Human Factors

Keywords: Intelligent Tutoring Systems, Off-Task Behavior, Field Research Methods, User Modeling

INTRODUCTION

Cognitive tutor curricula are one of the more successful and widely-used approaches to incorporating computer-aided instruction into the classroom. Cognitive tutor curricula combine conceptual instruction delivered by a teacher with problem-solving where each student works one-on-one

with a cognitive tutoring system which chooses exercises and feedback based on a running model of which skills the student possesses [17]. At this point, about 5% of US high schools and middle schools are using cognitive tutoring curricula in their algebra or geometry courses. Over the years, it has become clear that designing an optimal cognitive tutor lesson requires substantial attention to issues of student cognition: the conceptual structure of the domain [15], the prior knowledge students bring to bear on the material [14], and how students construct understanding in general [4].

However, paying attention to these factors does not complete the tutor designer’s task. How the student chooses to interact with the tutoring software may have an equally large impact on what they learn [21]. Although using cognitive tutor software has been found to increase student involvement and effort in the classroom [22], some students have also responded to the help, feedback, and support of cognitive tutoring software with a set of non-learning-oriented strategies. This set of strategies, which we will refer to as “gaming the system”, consists of behavior aimed at obtaining correct answers and advancing within the tutoring curriculum by taking advantage of regularities in the software’s responses – systematically misusing the software’s feedback or help instead of actively thinking about the material. Some examples of gaming the system include:

1. quickly and repeatedly asking for help until the tutor gives the student the correct answer [2].
2. inputting answers quickly and systematically. For instance, entering 1,2,3,4,... or clicking every checkbox within a set of multiple-choice answers, until the tutor identifies a correct answer and allows the student to advance.

In both of these cases, features designed to help a student learn curricular material via problem-solving are instead being used to solve the current problem and move forward within the curriculum.

Although this phenomenon has been repeatedly documented in students’ help-seeking behavior, both in intelligent tutor and traditional classroom research [2,3,5,22,23], its effects on learning have not been systematically studied.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2004, April 24–29, 2004, Vienna, Austria.

Copyright 2004 ACM 1-58113-702-8/04/0004...\$5.00.

Gaming the system, although it has similarities to cheating, is not identical to cheating. Gaming the system generally involves taking advantage of loopholes in a system, whereas cheating involves direct violation of that system's rules. While gaming the system is an inappropriate use of a learning opportunity, such behavior is generally considered acceptable (if not desirable) within the context of a high-stakes examination such as the SAT – for instance, test preparation companies teach students to use the structure of how SAT questions are designed in order to have a higher probability of guessing the correct answer. Cheating on the SAT, by contrast, is not generally considered acceptable.

Students who game intelligent-tutoring software are clearly not engaged in attempting to use that software to learn. However, it is an open question what gaming's effects on learning will be. It may be instructive, for instance, to compare the effects of this strategy on learning to the effects of other types of non-learning oriented behavior.

One of the most prominent accounts of the relationship between classroom behavior and learning is Carroll's Time-On-Task hypothesis [8,10]. Under this hypothesis, the longer a student spends engaging with the learning materials, the more opportunities the student has to learn. Therefore, if a student spends a greater fraction of their time off-task (engaged in behaviors where learning from the material is not the primary goal)¹, they will spend less time on-task, and learn less. This hypothesis suggests that off-task behavior will reduce learning. If it is the main reason why off-task behavior reduces learning, then other types of off-task behavior, such as talking to a neighbor or surfing the web, should have a similar effect on learning as gaming the system does.

In order to investigate how gaming the system and other off-task behaviors affect learning within interactive learning environments, we present a study in which we observed the frequency of different types of off-task behavior in students using a cognitive tutor, using an approach building upon the tradition of research in off-task behavior in the classroom [16,18,19,20].

We show how, in the classes we studied, different types of off-task behavior had different levels of correlation to learning. We compare the strength of these relationships to the relationships between learning and prior domain knowledge, general measures of classroom achievement, gender, and other factors.

We then discuss the degree to which gaming the system and other off-task behaviors were associated with better and poorer learning, determine which students game the system, and what this implies about why students game the system. We conclude with a discussion of possible ways

¹ It is possible to define on-task as "looking at the screen", in which case gaming the system is viewed as an on-task behavior. Of course, the definition of "on-task" depends on what we consider the student's task to be – we do not consider just "looking at the screen" to be that task.

cognitive tutors can respond to this behavior, and why one frequently used response may be inadvisable.

METHOD

The Cognitive Tutor Classroom

We conducted this study in a set of 5 middle-school classrooms at 2 schools in the suburbs of a medium-sized city in the Northeastern United States. Student ages ranged from approximately 12 to 14. The classrooms studied were taking part in the development of a new 3-year cognitive tutor curriculum for middle school mathematics. Seventy students were present for all phases of the study.

We studied these classrooms during the course of a short (2 class periods) cognitive tutor unit on scatterplot generation and interpretation – an earlier version of this unit is described in [7]. Scatterplots depict the relationship between two quantitative variables in a Cartesian plane, using a point to represent paired values of each variable.

The day before using the tutoring software, students viewed conceptual instruction, delivered via a PowerPoint presentation with voiceover and some simple animations (see figure 1 for an example).

The declarative instruction and cognitive tutor focused on the skills involved in choosing variables of the correct type for a scatterplot and selecting an appropriate scale for each axis, as per the recommendations in [7]. Students also had to label the values of each variable along the axis, as recommended in [6], and had to plot each of the points of the data set, as in [17]. A screenshot of the cognitive tutor is shown in Figure 2.

Data Collection

Each student in each class took a pre-test and post-test. The pre-test was given after the student had finished viewing the PowerPoint presentation, in order to study the effect of the cognitive tutor rather than studying the combined effect of the declarative instruction and cognitive tutor. The post-test was given at the completion of the tutor lesson.

Two nearly isomorphic exercises were constructed for the tests. Each exercise was used as a pre-test for half of the students, and as a post-test for the other half. In the exercises, students were given a data set with two quantitative variables to use, and two "distractor" variables (one quantitative, one nominal) which were not appropriate to use to answer the given question. The students were then asked to draw a scatterplot to show the relationship between the two quantitative variables. The tests were scored in terms of how many of the steps of the problem-solving process were correct; the items were designed so that it was often possible to get later steps in the problem correct even after making a mistake – for example, choosing the wrong variable did not always preclude selecting an appropriate scale for that variable.

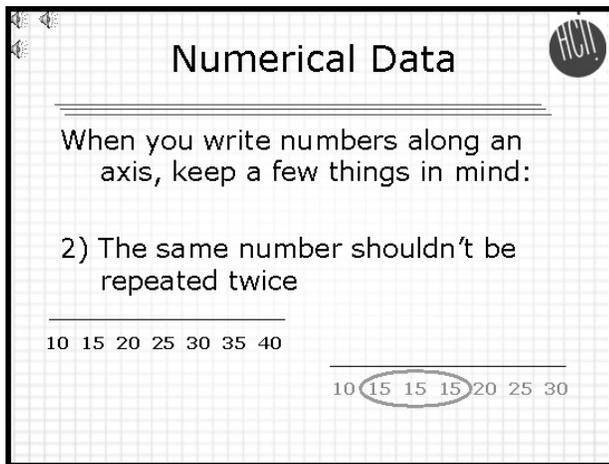


Figure 1: Declarative instruction was given via PowerPoint with voice-over

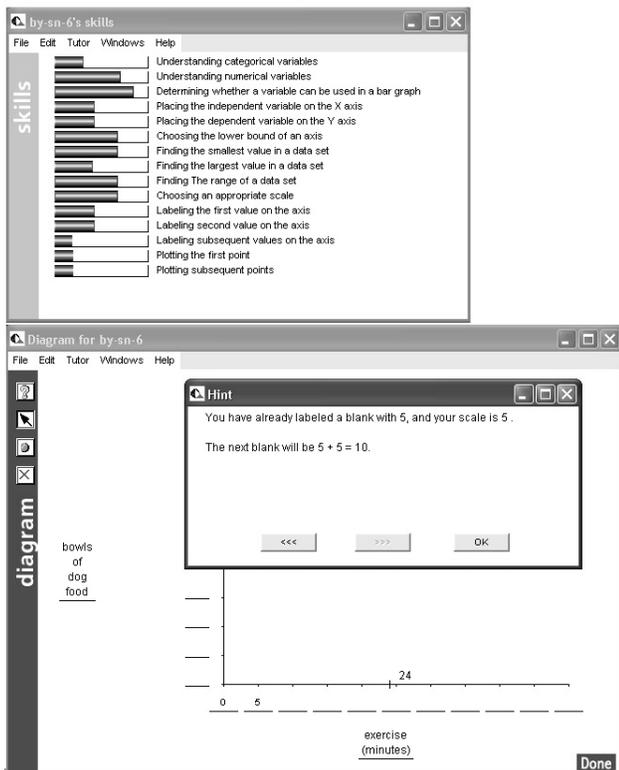


Figure 2: The last stage of a multi-stage hint: The student labels the graph's axes and plots points in the bottom window; the tutor's estimates of the student's skills are shown in the top window; the hint window (superimposed on the bottom window) allows the tutor to give the student feedback. Other windows (such as the problem scenario and interpretation questions window) are not shown.

In addition to the pre-test and post-test measures, we collected evidence about the pattern of students' on and off task behavior during tutor usage. Each student's behavior was observed a number of times during the course of each class period, by either the first or fourth author. We used outside observations of behavior rather than self-report in order to interfere minimally with the experience of using the tutor – we were concerned that repeatedly halting the student during tutor usage to answer a questionnaire (which was done to assess motivation in [12]) might affect both learning and on/off-task behavior. In order to investigate the relative impact of gaming the system as compared to other types of off-task behavior, we coded not just the frequency of off-task behavior, but its nature as well. This method differs from most past observational studies of on and off-task behavior, where the observer coded only whether a given student was on-task or off-task [16,18,19,20]. Our coding scheme consisted of six categories:

1. **on-task** -- working on the tutor
2. **on-task conversation** -- talking to the teacher or another student about the subject material
3. **off-task conversation** – talking about anything other than the subject material
4. **off-task solitary behavior**– any behavior that did not involve the tutoring software or another individual (such as reading a magazine or surfing the web)
5. **inactivity** -- for instance, the student staring into space or putting his/her head down on the desk for the entire 20-second observation period
6. **gaming the system** – systematic and rapid incorrect answers or use of help, with several such actions taking place during the 20-second observation period. Examples include entering 1,2,3,4... or clicking every checkbox in a set of multiple-choice answers until an answer is identified as correct by the system.

In order to avoid bias towards more interesting or dramatic events, the coder observed the set of students in a specific order determined before the class began, as in [20]. Any behavior by a student other than the student currently being observed was not coded. A total of 563 observations were taken, with an average of 8.0 observations per student, with some variation due to different class sizes and students arriving to class early or leaving late. Each observation lasted for 20 seconds – if a student was inactive for the entire 20 seconds, the student was coded as being inactive. If two distinct behaviors were seen during an observation, only the first behavior observed was coded. In order to avoid affecting the

**Table 1: The correlations between post-test score and our other measures.
Statistically significant relationships are in boldface**

	Prior Knowledge (Pre-Test)	General Academic Achievement	Gaming the System	Talking Off-Task	Inactivity	Off-Task Solitary Behavior	Talking On-Task	Gender	Teacher
Post-Test	0.32	0.36	-0.38	-0.19	-0.08	-0.08	-0.24	-0.08	n/a, n/s

current student's behavior if they became aware they were being observed, the observer viewed the student out of peripheral vision while appearing to look at another student.

The two observers observed one practice class period together before the study began. In order to avoid alerting a student that he or she was currently being observed, the observers did not observe any student at the same time. Hence, we cannot compare the two observers' assessment of the exact same time-slice of a student's behavior, and thus cannot directly compute a traditional measure of interrater reliability. However, the two observers had exceptionally high agreement in the overall proportions of each type of behavior within this class, $F(1,4)=287.77, p<0.01, r=0.99$ and very good agreement in the proportion of each type of behavior per student, $F(1,142)=238.37, p<0.01, r=0.79$. After this first practice class, the two observers observed classes separately.

Finally, we used students' end-of-course test scores (which incorporated both multiple-choice and problem-solving exercises) as a measure of general academic achievement², noted each student's gender, and collected detailed log files of the students' usage of the cognitive tutoring software.

RESULTS

The tutor was, in general, successful. Students went from 40% on the pre-test to 71% on the post-test, which was a significant improvement, $F(1,68)=7.59, p<0.01$.

Students were on-task 82% of the time, which is within the previously reported ranges for average classes utilizing traditional classroom instruction [19,20]. Within the 82% of time spent on-task, 4% was spent talking with the teacher or another student, while the other 78% was solitary. The most frequent off-task behavior was off-task conversation (11%), followed by gaming the system (3%), inactivity (3%), and off-task solitary behavior (1%). More students engaged in these behaviors than the absolute frequencies might suggest: 41% of the students were observed engaging in off-task conversation at least once, 24% were observed gaming the system at least once, 21% were observed to be inactive at least once, and 9% were

observed engaging in off-task solitary behavior at least once. 100% of the students were observed working at least once.

A student's prior knowledge of the domain (measured by the pre-test) was a reasonably good predictor of their post-test score, $F(1,68)=7.59, p<0.01, r=0.32$. A student's general level of academic achievement was also a reasonably good predictor of the student's post-test score, $F(1,61)=9.31, p<0.01, r=0.36$. Prior knowledge and the general level of academic achievement were highly correlated, $F(1,61)=36.88, p<0.001, r=0.61$; when these two terms were both used as predictors, the correlation between a student's general level of academic achievement and their post-test score was no longer significant, $F(1,60)=1.89, p=0.17$.

Gender was not predictive of post-test performance, $F(1,68)=0.42, p=0.52$. Neither was which teacher the student had, $F(3,66)=0.5, p=0.69$.

Gaming the System and Other Off-Task Behavior: Relationships to Learning

How frequently a student was off-task (with all types of off-task behavior combined) was a reasonably good predictor of their performance on the post-test, $F(1,68)=8.52, p<0.01, r=-0.33$. However, there was substantial difference between different types of off-task behavior.

The frequency of gaming the system was the only off-task behavior which was significantly correlated with the post-test, $F(1,68)=11.82, p<0.01, r=-0.38$. The impact of gaming the system remains significant even when we control for the students' pre-test and general academic achievement, $F(1,59)=7.73, p<0.01, \text{partial correlation} = -0.34$.

By comparison, the frequency of talking off-task was at best marginally significantly correlated with post-test score, $F(1,68)=2.45, p=0.12, r=-0.19$. That relationship reduced to $F(1,59)=2.03, p=0.16$ when we controlled for pre-test and general academic achievement. Furthermore, the frequencies of inactivity ($F(1,68)=0.44, p=0.51, r=-0.08$) and off-task solitary behavior ($F(1,68)=0.42, p=0.52, r=-0.08$) were not significantly correlated to post-test scores.

² We were not able to obtain end-of-course test data for one classroom, due to that class's teacher accidentally discarding the sheet linking students to code numbers.

Table 2: The correlations between gaming the system and other measures.
Statistically significant relationships are in boldface, marginally significant relationships are in italics.

	Prior Knowledge (Pre-Test)	General Academic Achievement	Other Off-Task Behavior	Talking On-Task	Gender	Teacher
Gaming the System	-0.27	<i>-0.21</i>	0.07	0.37	0.12	n/a, n/s

Unexpectedly, the frequency of talking to the teacher or another student about the subject matter was significantly negatively correlated to post-test score, $F(1,68)=4.11$, $p=0.05$, $r=-0.24$, and this remained significant even when we controlled for the students' pre-test and general academic achievement, $F(1,59)=3.88$, $p=0.05$, partial correlation = -0.25. The implications of this finding will be discussed in more detail in the following section.

We can better understand the relationship between frequency of gaming the system and post-test score by comparing the post-test scores of students who gamed with different frequencies. Using the median frequency of gaming among students who ever gamed (10%), we split the 17 students who ever gamed into a high-gaming half (8 students) and a low-gaming half (9 students). We can then compare the 8 high-gaming students to the 53 never-gaming students. The 8 high-gaming students' mean score at post-test was 44%, which was significantly lower than the never-gaming students' mean post-score of 78%, $F(1,59)=8.61$, $p<0.01$. However, the 8 high-gaming students also had lower pre-tests. The 8 high-gaming students had an average pre-test score of 8%, with none scoring over 17%, while the 53 never-gaming students averaged 49% on the pre-test. Given this, one might hypothesize that choosing to game the system is mainly a symptom of not knowing much to start with, and that it has no effect of its own.

We can address this hypothesis by comparing the 8 high-gaming students to the 24 never-gaming students with pre-test scores equal to or less than 17% (the highest pre-test score of any high-gaming student). The 24 never-gaming/low-pre-test students had an average pre-test score of 7%, but an average post-test score of 68%, which was substantially higher than the 8 high-gaming students' average post-test score (44%). Due to the small population of high-gaming students, the difference in post-test score was only marginally significant, $t(30)=1.69$, $p=0.10$, but the large magnitude of the difference is nonetheless quite suggestive.

Which Students Game The System?

In this section, we investigate what distinguished the students who chose to game the system from the other students. Gaming the system was associated with substantially less effective learning, but it was not a behavior that a majority of the students engaged in. Only 11% of the students (the "high-gaming" students from the

previous section) were observed gaming the system more than 10% of the time, and only 24% of the students in our study were observed gaming the system even once. By comparison, 51% of students were observed engaging in other types of off-task behavior at least once. Significantly fewer students were ever observed gaming the system than were observed engaging in other off-task behaviors, $\chi^2(1, N=140)=19.52$, $p<0.001$.

As mentioned in the prior section, students who gamed the system scored lower on the pre-test. While not all students with low pre-test scores gamed the system, all of the high-gaming students had low pre-test scores. More generally, there was a significant correlation between how frequently a student gamed the system and their pre-test score, $F(1,68)=5.31$, $p=0.02$, $r=-0.27$. Performance on the test of general academic achievement was marginally significantly correlated to how frequently a student gamed the system, $F(1,61)=2.77$, $p=0.10$, $r=-0.21$. When general academic achievement and pre-test score were both used as predictors of how often a student gamed the system, general academic achievement was no longer close to significance, $F(1,60)=0.22$, $p=0.64$.

As with the post-test, there was neither a statistically significant relationship between gender and the frequency of gaming the system, $F(1,68)=1.02$, $p=0.31$, nor between what teacher the student had and the frequency of gaming the system, $F(3,66)=0.99$, $p=0.41$. There was also not a significant relationship between gaming the system and other off-task behavior, $F(1,68)=0.33$, $p=0.57$. The 8 high-gaming students engaged in other off-task behaviors with about the same frequency (15%) as the never-gaming students did (14%).

By contrast, there was a fairly strong relationship between a student's frequency of gaming the system and that student's frequency of talking to the teacher or another student about the subject matter, $F(1,68)=10.52$, $p<0.01$, $r=0.37$. This relationship remained after controlling for prior knowledge and general academic achievement, $F(1,59) = 8.90$, $p<0.01$, partial correlation = 0.36. One chosen to talk to students when he/she saw them gaming, accounting for this link in a very different way.

In summary, the students who gamed the system were of lower than average academic achievement and were among the students who had the least prior knowledge of the material in the lesson. Students who shared these characteristics but did not game the system learned

considerably more than students who gamed the system. The frequency of gaming the system remained a significant predictor of post-test performance even when these quantities were controlled for. In addition, there was no relationship, either positive or negative, between the frequency with which a student gamed the system and the frequency of other off-task behaviors. However, students who frequently gamed the system also more frequently talked on-task with the teacher and the other students.

DISCUSSION

One of the interesting features of the data we have presented is what it suggests about the relationship between off-task behavior and learning. In this case, the nature of the student's off-task behavior mattered more than its absolute quantity, with the frequency of gaming the system a much better predictor of learning than the frequency of other types of off-task behavior. This contradicts the prediction made by the Time-On-Task hypothesis. Less time on task does appear to imply less learning, but how the student chooses to use the software is more important than the exact proportion of available time the student spends using it. A moderate amount of gaming the system seems to be more harmful than a moderate amount of other off task behaviors.

One possible hypothesis (supported by the lack of relationship between the frequency of gaming the system and the frequency of other off-task behaviors) is that gaming the system is actually a sign that the student's overall approach to the software focuses on performance rather than learning. This focus could lead to constructing knowledge in a task-specific way that would not transfer outside of using the tutor [9], suggesting that students who game the system learn less even when they are not gaming the system. By contrast, the students who talk to their neighbors frequently might still be thinking carefully about the material when they actually are looking at it. Thus, gaming the system may be a signal that the student has, in general, the goal of performing rather than the goal of learning (often referred to as a "performance orientation"), and it is this performance orientation which more globally harms learning. Performance orientation has been correlated to the frequency of a classroom behavior very similar to gaming the system, called "executive" help-seeking [5], where students request help from their teacher immediately, before attempting to solve the problem at hand. This may also provide an account for the surprising relationship between more on-task conversation and lower learning.

Another hypothesis is that students may choose to game the system on exactly those problem steps where they have the most difficulty. Thus, even though they are only gaming the system a moderate amount of the time, they are selectively gaming the system exactly where it will most hurt their learning. By contrast, even if students choose to talk to a neighbor when they are faced with an especially

difficult problem step, when they are done talking they must return to that problem step. This approach would suggest that gaming the system has commonalities with the strategic behavior shown by students who have learned helplessness, attributing their early failures (which generally result from low prior knowledge, a characteristic of the high-gaming students in our study) to a global lack of aptitude, and actively avoiding difficult challenges, thus learning less than students who keep trying [13].

We do not yet have the evidence to distinguish between these two hypotheses for why students game the system; distinguishing which of these hypotheses is a better account for why students game the system should be a focus of future work in this area. However, we *can* conclude with fairly high confidence that lack of interest in the material is unlikely to be a good explanation of why students game the system. If students were gaming the system because of lack of interest, we would expect them to also engage more frequently in other types of off-task behavior. They did not do so; students who frequently gamed the system engaged in other types of off-task behavior with almost exactly the same frequency as the other students did.

Re-Designing Tutors to Appropriately Respond to Gaming the System

At this point, we have established that there is a definite link between how much a student games the system and how they perform on the post-test, even after controlling for their level of prior knowledge about the material and general academic ability. We do not yet know for certain why students choose to game the system. Thus far, many of the cognitive tutoring community's responses to students gaming the system have bypassed the question of why students game the system. Instead, they have simply re-designed tutor interfaces to make it more difficult for students to game the system. For instance, the decision to implement multi-level help within cognitive tutors instead of just giving students the answer on demand was, among other reasons, motivated by overuse of the answer on demand feature [22]. More recently, intelligent tutor lessons have been designed without "bottom-out" hints which give the answer directly to students [1] or with a time delay before the student can request each successive level of help, in order to encourage students to read each level of the help. However, the impact of these interventions has not yet been conclusively demonstrated.

Another frequent type of gaming the system, systematically and rapidly trying many answers until one turns out to be correct, might seem already counteracted by mastery learning, a feature of many cognitive tutors. When a tutor uses mastery learning, a student can only advance through the curriculum by demonstrating knowledge of the relevant problem-solving skills; the tutor infers knowledge of a skill from the student's success in responding correctly on the first try when that skill applies [11]. Eventually, however, the tutor runs out of problems to assign and re-assigns

problems the student has already seen; also, teachers have reported advancing students to the next lesson to keep them from falling too far behind the rest of the class. Thus, if the student is sufficiently patient, they can advance to the next lesson without having to learn.

In general, trying to redesign tutors to directly prevent students from gaming the system may lead to an arms race, with students figuring out new ways to game the system in response to our re-designed tutors. If done carelessly, such a re-design may even hinder learning for those students who do not game the system. A major reason not to adopt a one-size-fits-all approach to eliminating gaming is the fact that only a minority of students engage in it. It would be unproductive to affect (potentially negatively) all of the other students' learning experiences in order to try to prevent a small number from gaming the system. One solution might be to target such interventions only towards students with low prior knowledge, since all of the high-gaming students had low prior knowledge. However, not all students with low prior knowledge game the system, and past studies have shown that many students with low prior knowledge *do* make fruitful use of help [23]. We should not get in their way. Any change we make should be targeted as carefully as possible to only the students who game the system. The approach this would suggest is to have the tutor detect gaming in real-time and adapt to each specific student.

Once we know more about why students choose to game the system, we can re-design the tutor to adapt other elements of its presentation based on an assessment of each student's motivational needs. It may even turn out that gaming the system is only a symptom of the student's general approach to using the tutor – in which case, the most direct approaches to prevent students from gaming the system might succeed at reducing gaming but have no effect on learning.

For example, if we determine that a given student is gaming the system only on the steps where he/she is having the most difficulty, we could point that behavior out to the student and offer encouragement when the student appears to be genuinely trying and failing. On the other hand, if gaming the system arises from performance orientation, it might be worth creating an "effort meter", to go along with the already existing "skill meters", to signal when a student is gaming the system. This meter could be seen by teachers and used in their grading, making attempting to learn (or at least persuading the software you were attempting to learn) a performance goal in itself.

CONCLUSION

In this study, we found that off-task behavior was associated with less learning, but that this was not true of all types of off-task behavior. Specifically, gaming the system, where a student systematically misuses the software's help and feedback in order to get the correct answer without having to know why that answer was

correct (or having to engage in meaningful thought about the material), had a much stronger negative correlation to learning than any other type of off-task behavior.

In this study, the students who gamed the system had low knowledge of the material at pre-test, and were of low general academic achievement. However, there were many students with these characteristics who did not game the system. This suggests that while low prior knowledge and academic achievement may play a role in a student's decision to game the system, they are not sufficient to explain why students game the system.

The lack of relationship between how often a student games the system and how often that student engages in other off-task behaviors (especially when compared to the fairly strong relationship between gaming the system and talking to the teacher) suggests that it is unlikely that students game the system because of lack of interest in the material. Given the comparatively low impact of off-task behaviors stemming from lack of interest, it probably makes sense to focus our attempts to address motivation on those students who choose to misuse the software rather than those students who do not always choose to use the software. Further research, both into which students game the system, and why they do, should shed further light on this subject.

In the long-term, the re-design of cognitive tutors to detect when a student is gaming the system and respond in an appropriate fashion will help us to design tutors where students both have a more positive experience using the tutor, and where they learn more, helping us to create cognitive tutoring curriculums which address the needs of all students.

ACKNOWLEDGMENTS

We would like to thank Jane Kamneva, Heather Frantz, and Pauline Masley for assisting us in conducting this study, Jay Raspat, Meghan Naim, Katy Getman, and Pat Battaglia for welcoming us into their classrooms and giving us many helpful suggestions, and Shaaron Ainsworth, Vincent Alevan, Lisa Anthony, Samuel Baker, Darren Gergle, Elsa Golden, Andrew Ko, George Loewenstein, Jack Mostow, Santosh Mathan, Ido Roll, Deborah Small, Desney Tan, and our reviewers for helpful suggestions and comments. This work was funded by an NDSEG (National Defense Science and Engineering Graduate) Fellowship, and by NSF grant 9720359 to "CIRCLE: Center for Interdisciplinary Research in Constructive Learning Environments".

REFERENCES

- [1] Alevan, V. Helping Students To Become Better Help Seekers: Towards Supporting Metacognition in a Cognitive Tutor. Paper presented at *German-USA Early Career Research Exchange Program: Research*

on Learning Technologies and Technology-Supported Education, Tübingen, Germany (October 2001).

- [2] Aleven, V., Koedinger, K.R. Limitations of Student Control: Do Students Know When They Need Help? *Proceedings of the 5th International Conference on Intelligent Tutoring Systems* (2000), 292-303.
- [3] Aleven, V., Koedinger, K.R. Investigations into Help Seeking and Learning with a Cognitive Tutor. In R. Luckin (Ed.), *Papers of the AIED-2001 Workshop on Help Provision and Help Seeking in Interactive Learning Environments* (2001), 47-58.
- [4] Anderson, J.R. *Rules of the Mind*. Lawrence Erlbaum Associates, Hillsdale, NJ (1993).
- [5] Arbretton, A. Student Goal Orientation and Help-Seeking Strategy Use. In S.A. Karabenick (Ed.), *Strategic Help Seeking: Implications For Learning And Teaching*, Lawrence Erlbaum Associates, Mahwah, NJ (1998), 95-116.
- [6] Baker, R.S., Corbett, A.T., Koedinger, K.R. The Resilience of Overgeneralization of Knowledge about Data Representations. *Presented at American Educational Research Association Conference* (2002).
- [7] Baker, R.S., Corbett, A.T., Koedinger, K.R., Schneider, M.P. A Formative Evaluation of a Tutor for Scatterplot Generation: Evidence on Difficulty Factors, *Proceedings of the Conference on Artificial Intelligence in Education* (2003), 107-115.
- [8] Bloom, B.S. *Human Characteristics and School Learning*. McGraw-Hill, New York (1976). Ch. 7.
- [9] Bransford, J.D., Schwartz, D.L. Rethinking Transfer: A Simple Proposal With Interesting Implications. In Iron-Nejad, A., Pearson, P.D. (Eds.) *Review of Research in Education*, 24 (1998), 61-101.
- [10] Carroll, J. A Model For School Learning. *Teachers College Record*, 64 (1963), 723-733.
- [11] Corbett, A.T., Anderson, J.R. Student Modeling and Mastery Learning in a Computer-Based Programming Tutor. *Second International Conference on Intelligent Tutoring Systems* (1992), 413-420.
- [12] de Vicente, A., Pain, H. Informing the detection of the students' motivational state: an empirical study. *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems* (2002), 933-943
- [13] Dweck, C.S., Leggett, E.L. A Social-Cognitive Approach to Motivation and Personality. *Psychological Review*, 95 (2) (1988), 256-271.
- [14] Hall, V.C. & Edmondson, B. Relative Importance of Aptitude and Prior Domain Knowledge on Immediate and Delayed Posttests. *Journal of Educational Psychology*, 84 (2) (1992), 219-223.
- [15] Heffernan, N.T., Koedinger, K.R. A developmental model for algebra symbolization: The results of a difficulty factors assessment. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (1998), 484-489.
- [16] Karweit, N., Slavin, R.E. Time-On-Task: Issues of Timing, Sampling, and Definition. *Journal of Experimental Psychology*, 74 (6) (1982), 844-851.
- [17] Koedinger, K.R., Anderson, J.R., Hadley, W.H., & Mark, M.A. Intelligent Tutoring Goes to School in the Big City. *International Journal of Artificial Intelligence in Education*, 8 (1997), 30-43.
- [18] Lahaderne, H.M. Attitudinal and Intellectual Correlates of Attention: A Study of Four Sixth-Grade Classrooms. *Journal of Educational Psychology*, 59 (5) (1968), 320-324.
- [19] Lee, S.W., Kelly, K.E., Nyre, J.E. Preliminary Report on the Relation of Students' On-Task Behavior With Completion of School Work. *Psychological Reports*, 84 (1999), 267-272.
- [20] Lloyd, J.W., Loper, A.B. Measurement and Evaluation of Task-Related Learning Behavior: Attention to Task and Metacognition. *School Psychology Review*, 15 (3) (1986), 336-345.
- [21] Mostow, J., Aist, G., Beck, J., Chalasani, R., Cuneo, A., Jia, P., Kadaru, K. A La Recherche du Temps Perdu, or As Time Goes By: Where Does the Time Go in a Reading Tutor that Listens? *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems* (2002), 320-329.
- [22] Schofield, J.W. *Computers and Classroom Culture*. Cambridge University Press, Cambridge, UK (1995).
- [23] Wood, H., Wood, D. Help Seeking, Learning, and Contingent Tutoring. *Computers and Education*, 33 (1999), 153-169.