Do Performance Goals Lead Students to Game the System?

Ryan Shaun BAKER, Ido ROLL, Albert T. CORBETT, Kenneth R. KOEDINGER

Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213 USA

rsbaker, idoroll, corbett, koedinger @cmu.edu

Abstract. Students approach the learning opportunity offered by intelligent tutoring systems with a variety of goals and attitudes. These goals and attitudes can substantially affect students' behavior within the tutor, and how much the student learns. One behavior that has been found to be associated with poorer learning is gaming the system, where a student attempts to complete problems and advance through an educational task by systematically taking advantage of properties and regularities in the system used to complete that task. It has been hypothesized that students game the system because of performance goals. In this paper, however, we find that the frequency of gaming the system does not correlate to a known measure of performance goals; instead, gaming is correlated to disliking computers and the tutor. Performance goals, by contrast, are shown to be associated with working slowly and avoiding errors, and are found to not be correlated to differences in learning outcomes.

1. Introduction

Understanding the student has always been a focus of intelligent tutoring research, but in recent years, there has been a distinct shift in what we are trying to understand about students. In the early years of the field, student modeling focused mostly on issues of knowledge and cognition: modeling what a student knew about the tutor's subject matter, how students acquired and constructed knowledge, and how incorrect knowledge could be modeled and responded to. This research focus led to intelligent tutoring systems that can effectively assess and adapt to students' knowledge about the educational domain, improving learning outcomes [10,17].

In recent years, there has been increasing evidence that students' behavior as they use intelligent tutoring systems is driven by a number of factors other than just their domain knowledge. There is increasing evidence that students with different motivations, beliefs, or goals use tutoring systems and other types of learning environments differently [3,7,9,11]. Furthermore, behaviors that appear to stem from factors other than student knowledge, such as abusing tutor help and feedback [1,6,8] or repeating problems over and over [19], can result in substantially poorer learning outcomes.

While these sorts of findings inform the design of more educationally effective tutors, they are by themselves incomplete. Knowing that a student possesses or fails to possess specific motivations, attitudes, or goals does not immediately tell us whether that student is in need of learning support. Similarly, observing a student using a tutor in a fashion associated with poorer learning does not tell us why that student is choosing to use the tutor in that fashion. If we observe that a specific behavior is associated with poorer learning, we can simply re-design the tutor to eliminate the behavior (cf. [8]), but if the behavior is symptomatic of a broader motivational problem, such a solution may mask the problem rather than eliminate it.

Hence, in order to design systems that can respond to student goals, attitudes, and behaviors in a fashion that positively impacts learning, it is valuable to research all of these factors together. That way, we can learn what motivations, goals, and beliefs lead students to engage in behaviors that negatively impact learning.

2. Gaming the System

In this paper, we apply this combined research approach to the question of why students choose to game the system, a strategy found to be correlated to poorer learning [6]. Gaming the system is behavior aimed at completing problems and advancing through an educational task by systematically taking advantage of properties and regularities in the system used to complete that task, rather than by thinking through the material. In [6], students were observed engaging in two types of gaming the system: systematic trial-and-error, and help abuse, where a student quickly and repeatedly asks for help until the tutor gives the correct answer, often before attempting to solve the problem on his or her own (cf. [1,23]). Within that study, gaming was strongly negatively correlated with learning; students who frequently gamed learned 38% less than students who never gamed, controlling for pre-test score. By contrast, off-task behaviors such as talking to neighbors (about subjects other than the tutor or educational domain) or surfing the web were not negatively correlated with learning. This finding was refined in later analysis, where machine learning determined that gaming students split into two behaviorally distinguishable groups, one which gamed but still learned, and another which gamed and failed to learn [4]. These two groups appeared identical to human observers, but were distinguishable to the machine learning algorithm.

Students who have performance goals, focusing on performing well rather than learning [14], have been found to engage in behaviors that appear similar to gaming, such as seeking answers before trying to solve a problem on their own [2]. For this reason, both our research group [6] and other researchers [18] have hypothesized that students game because of performance goals. A second hypothesis is that students might game out of anxiety, gaming out of the belief that they cannot succeed otherwise [6, cf. 12]. The anxiety hypothesis was supported by evidence that students who game in the harmful fashion tend to game on the hardest steps of the problem [4]. It is also worth noting that having performance goals has been found to lead to anxiety and withdrawal of effort [14] – therefore these two hypotheses may not be inconsistent.

In the remainder of this paper, we will present a study designed to investigate which student goals, beliefs and motivations are associated with gaming the system, with the goal of understanding which of these two hypotheses better explains why students game – or if students game for another reason entirely.

3. Study Methods

We studied student goals, attitudes, behavior, and learning within 6 classes at 2 schools within the Pittsburgh suburbs. All students were participating in a year-long cognitive tutor curriculum for middle school mathematics. Student ages ranged from approximately 12 to 14. 102 students completed all stages of the study; 23 other students were removed from analysis due to missing one or more parts of the study.

We studied these students during the course of a short (2 class period) cognitive tutor lesson on scatterplot generation and interpretation [5]. Within this study, we combined the following sources of data: a questionnaire on student motivations and beliefs, logs of each student's actions within the tutor (analyzed both in raw form, and through a gaming detector (cf. [4]), and pre-test/post-test data. Classroom observations were also obtained in order to improve the gaming detector's accuracy.

The questionnaire consisted of a set of self-report questions given along with the pre-test, in order to assess students' motivations and beliefs. The questionnaire items were drawn from existing motivational inventories or from items used across many prior studies with this age group, and were adapted minimally (for instance, the words "the computer tutor" was regularly substituted for "in class", and questions were changed from first-person to second-person for consistency). All items were pre-tested for comprehensibility with a student from the relevant age group before the study.

The questionnaire included items to assess:

- Whether the student was oriented towards performance or learning (2 items, 4 choices) (e.g. [20])
  "We are considering adding a new feature to the computer tutors, to give you more control over the problems the tutor gives you. If you had your choice, what kind of problems would you like best?
  A) Problems that aren't too hard, so I don't get many wrong.
  B) Problems that are pretty easy, so I'll do well.
  C) Problems that I'm pretty good at, so I can show that I'm smart
  D) Problems that I'll learn a lot from, even if I won't look so smart."
- The student's level of anxiety about using the tutor (2 items, scale 1-6) (e.g. [16])
  "When you start a new problem in the tutor, do you feel afraid that you will do poorly?"
  "When you are working problems in the tutor, do you feel that other students understand the tutor better than you?"
- The student's level of anxiety about using computers (1 item, scale 1-6) (e.g. [16])
  "When you use computers in general, do you feel afraid that you will do something wrong?"
- How much the student liked using the tutor (2 items, scale 1-6) (e.g. [20])
  "How much fun were the math problems in the last computer tutor lesson you used?"
  "How much do you like using the computer tutor to work through math problems?"
- The student's attitude towards computers (1 item, scale 1-6) (e.g. [15])
  "How much do you like using computers, in general?"
- If the student was lying or answering carelessly on the questionnaire. (1 item, 2 choices) (e.g. [21])
  "Is the following statement true about YOU? 'I never worry what other people think of me'. TRUE/FALSE"

Tutor log files were obtained as a source of data on students' actions within the tutor, for a sum total of 30,900 actions across the 106 students. For each action, we distilled 26 features (see [4] for more detail), consisting of:

- Data on how much time the current action (and recent actions) took
- The student's history of errors and help at the current skill and on recent steps
- What type of interface widget was involved in the action
- Whether the action was an error, a bug, correct, or a help request
- The tutor's assessment of the probability that the student knew the skill involved in the action [cf. 10]
- Whether the current action was the first action on the current problem step
- Whether the current problem step involved an "asymptotic" skill that most students knew before starting the tutor, or after the first opportunity to practice it

Using a combination of log files and classroom observations from this study and [6], we trained a gaming detector to assess how frequently a student engaged in harmful gaming and non-harmful gaming [4]. Within the analyses in this paper, we use this gaming detector's assessments as a measure of each student's incidence of harmful and non-harmful gaming rather than direct observations of gaming, for two reasons: First, because our direct observations did not distinguish between harmful gaming and non-harmful gaming whereas the detector could successfully make this distinction – and the two types of gaming may arise from different motivations. Second, because the gaming detector's assessments are more precise than our classroom observations – 2-3 researchers can only obtain a small number of observations of each student's behavior, but the gaming detector can make a prediction about every single student action.

Finally, a pre-test and post-test (the same tests as in [5,6]) were given in order to measure student learning. Two nearly isomorphic problems were used in the tests. Each problem was used as a pre-test for half of the students, and as a post-test for the other half. The tests were scored in terms of how many of the steps of the problem-solving process were correct; in order to get the richest possible assessment of students' knowledge about the material covered in the tutor lesson, the items were designed so that it was often possible to get later steps in the problem correct even after making a mistake.

4. Results

4.1 Gaming The System

Within this study, two types of questionnaire items were found to be significantly correlated to the choice to game: a student's attitude towards computers, and a student's attitude towards the tutor. Students who gamed in the harmful fashion (as assessed by our detector) liked computers significantly less than the other students, $F(1,100)=3.94$, $p=0.05$, $r = -0.19$, and liked the tutor significantly less than the other students, $F(1,100)= 4.37$, $p=0.04$, $r= -0.20$. These two metrics were related to each other: how much a student liked computers was also significantly positively correlated to how much a student liked the tutor, $F(1,100)= 11.55$, $p<0.01$, $r= 0.32$. Gaming in the non-harmful fashion was not correlated to disliking computers, $F(1,100) = 1.71$, $p=0.19$, or disliking the tutor, $F(1,100)=0.40$, $p=0.53$.

By contrast, our original hypotheses for why students might game did not appear to be upheld by the results of this study. Neither type of gaming was correlated to having performance goals (defined as answering in a performance-oriented fashion on both questionnaire items), $F(1,100)=0.78$, $p=0.38$, $F(1,100)=0.0$, $p=0.99$. Furthermore, a student's reported level of anxiety about using the tutor was not associated with choosing to game the system, in either fashion, $F(1,100) = 0.17$, $p=0.68$, $F(1,100) = 1.64$, $p= 0.20$ and a student's reported level of anxiety about using computers was not associated with choosing to game the system, in either fashion, $F(1,100)=0.04$, $p=0.84$, $F(1,100) = 0.58$, $p=0.45$.

Table 1. Correlations between gaming the system, the post-test (controlling for pre-test), and items on our motivational/attitudinal questionnaire. Statistically significant relationships ($p<0.05$) are in italics.

	Performance Goals	Anxiety about Using Computers	Anxiety about Using the Tutor	Lying/ Answering Carelessly	Liking Computers	Liking the Tutor
Gaming the System (Harmful fashion)	0.00	-0.02	-0.04	0.06	***- 0.19***	***- 0.20***
Post-Test	0.15	-0.02	0.04	0.03	***-0.32***	0.10

The different types of gaming were associated with learning in a fashion that corresponded to earlier results. Harmful gaming was negatively correlated with post-test score, when controlling for pre-test, $F(1,97)=5.61, p=0.02$, partial $r = -0.33$, providing a replication of the finding in [6] that gaming is associated with poorer learning. Additionally, non-harmful gaming did not correlate significantly to post-test score (controlling for pre-test), $F(1, 97)= 0.76$, $p=0.38$.

Since harmful gaming is correlated to poorer learning, and harmful gaming is correlated to disliking computers, it is not surprising that a student's attitude towards computers was significantly negatively correlated to their post-test score, $F(1,97)=11.51$, $p<0.01$, partial $r = - 0.32$, controlling for pre-test. To put the size of this effect in context, students who reported disliking computers (i.e. responding 1-2 on the survey item) or being neutral to computers (i.e. responding 3-4) had an average pre-post gain of 18%, whereas students who reported liking computers (i.e. responding 5-6) had an average pre-post gain of 33%. However, the link between computer attitudes and the student's post-test remained significant when harmful gaming (along with pre-test) is partialed out, $F(1,96)= 8.48$, $p<0.01$, and the link between harmful gaming and post-test remained significant when computer attitudes (along with pre-test) are partialed out, $F(1,96)=3.54$, $p=0.06$. This indicates that, although computer attitudes and gaming are linked, and both are connected to learning, the two have effects independent of each other. By contrast, a student's attitude towards the tutor was not significantly correlated to his/her post-test score, $F(1,97) = 0.99$, $p=0.32$, controlling for pre-test.

At this point, our original hypothesis (that gaming stems from performance goals) appears to be disconfirmed. On the other hand, we now know that students who game dislike computers and the tutor – but this raises new questions. Why do students who dislike computers and the tutor game? What aspects of disliking computers and the tutor are associated with gaming?

One possibility is that a student who has a negative attitude towards computers and the tutor may believe that a computer cannot really give educationally helpful hints and feedback – and thus, when the student encounters material she does not understand, she may view gaming as the only option. Alternatively, a student may believe that the computer doesn't care how much he learns, and decide that if the computer doesn't care, he doesn't either. A third possibility is that a student may game as a means of refusing to work with a computer she dislikes, without attracting the teacher's attention. All three of these possibilities are consistent with the results of this study; therefore, fully understanding the link between disliking computers and the tutor and the choice to game the system will require further investigation, probing in depth gaming students' attitudes and beliefs about computers (cf. [15]) and tutors.

4.2 Performance Goals

Entering this study, a primary hypothesis was that performance goals would be associated with a student's choice to game the system. However, as discussed in the previous section, this hypothesis was not upheld: we did not find a connection between whether a student had performance goals and whether that student gamed the system. Instead, performance goals appeared to be connected to a different pattern of behavior: working slowly, and making few errors.

Students with performance goals (defined as answering in a performance goal-oriented fashion on both questionnaire items) answered on tutor problem steps more slowly than the other students, $F(1,29276)=39.75$, $p<0.001$, controlling for the student's pre-test score and the student's knowledge of the current tutor step[1]. Overall, the median response time of students with performance goals was around half a second slower than that of the other students (4.4s .vs. 4.9s). Students with performance goals also made fewer errors per problem step than other students, $F(1,15854)= 3.51$, $p=0.06$, controlling for the student's pre-test score. Despite having a different pattern of behavior, students with performance goals completed the same number of problem-steps as other students, because slower actions were offset by making fewer errors, $t(100)=0.17$, $p=0.86$ (an average of 159 steps were completed by students with performance goals, compared to 155 steps for other students). Similarly, students with performance goals did not perform significantly better or worse on the post-test (controlling for pre-test) than other students, $F(1,97)=2.13$, $p=0.15$.

One possible explanation for why students with performance goals worked slowly and avoided errors rather than gaming is that these students may have focused on performance at a different grain-size than we had expected. We had hypothesized that students with performance goals would more specifically have the goal of performing well over the course of days and weeks, by completing more problems than other students – a goal documented in past ethnographic research within cognitive tutor classes [22]. We hypothesized that, in order to realize that goal, students would game the system. However, a student with another type of performance goal might focus on maintaining positive performance minute-by-minute. Such a student would set a goal of continually succeeding at the tutor, avoiding errors and attempting to keep their skill bars continually rising. These students could be expected to respond more slowly than other students, in order to avoid making errors – which is the pattern of behavior we observed.

An alternate account for why students with performance goals may work slowly and avoid errors comes from Elliot and Harackiewicz's 3-goal model of goal-orientation [13], which competes with the 2-goal model that our questionnaire items were drawn from [12]. In both models, students may have learning goals, but where the 2-goal model postulates a single type of performance goal, the 3-goal model states that students with performance goals may have either performance-approach goals (attempting to perform well) or performance-avoidance goals (attempting to avoid performing poorly). The 3-goal model might suggest that the students we identified as having performance goals actually had performance-avoidance goals, and that this was why these students tried to avoid making errors. That explanation would leave as an open question what sort of behavior students with performance-approach goals engaged in. However, in the 3-goal model, students with performance-avoidance goals are also predicted to have anxiety about the learning situation, and there was not a significant correlation between performance goals and tutor anxiety within our data, $F(1,100) = 1.52$, $p=0.22$ – suggesting that this questionnaire item was not solely capturing students with performance-avoidance goals.

On the whole, within our study, students with performance goals used the tutor differently than other students, but by working slowly and avoiding errors rather than by gaming the system. It is not yet entirely clear why students with performance goals chose to use the tutor in this fashion – one possible explanation is that these students focused on performance at a different grain-size than expected. In general, it appears that performance goals are not harming student learning, since students with performance goals learned the same amount as the other students. Therefore, recognizing differences in student goals and trying to facilitate a student in his/her goal preferences (cf. [18]) may lead to better educational results than attempting to make all students adopt learning goals.

[1] It is necessary to control for the student's knowledge of the current step for this analysis, since students who make more errors would be expected to have more actions on skills they know poorly – and actions on skills known poorly might be faster or slower in general than well-known skills.

5. Conclusions

The relationships between a student's motivations and attitudes, their actions within a tutoring system, and the learning outcome can be surprising. In this study, we determined that gaming the system, a behavior associated with poor learning, appears to not be associated with having performance goals or anxiety, contrary to earlier predictions. Instead, gaming the system was linked to disliking computers and the tutor. However, we do not yet know how disliking computers and the tutor leads students to game the system; there are several possible explanations for this relationship, from students not believing that the tutor's help and feedback could be educationally helpful, to students using gaming as a means of refusing to work with a computer they dislike. In order to design systems which can respond appropriately when a student games the system, it will be important to develop a richer understanding of the connection between the choice to game, and students' attitudes and beliefs about computers and tutoring systems.

Students with performance goals did not game the system. Instead, these students worked slowly within the tutor and made fewer errors per step than other students. One potential explanation is that students with performance goals focused on performing well at a step-by-step level, rather than attempting to perform well on a longer time-scale through completing more problems than other students. Another possibility is that the students with performance goals in our study more specifically had the desire to avoid performing poorly (cf. [13]), but this explanation is inconsistent with the lack of significant correlation between performance goals and anxiety.

One other question for future work is how well the findings presented here will generalize to other educational contexts. In this paper, we studied the links between motivations/attitudes, behavior within the tutor, and learning within the context of 12-14 year old students, who use cognitive tutors as part of a full-year curriculum, in public school classrooms in the suburban northeastern United States. It is quite possible that the relationships between students' motivations/attitudes, behavior within the tutor, and learning will differ across settings and populations.

Nonetheless, the results of this study demonstrate the value of combining data about how individual students use tutors with motivational, attitudinal, and learning data. In order to design tutors that can adapt to students in a fashion that improves learning, we need to know what behaviors are associated with poorer learning, and why students engage in these behaviors. The answers to these questions can be non-intuitive: before [6], we did not expect gaming the system to be the behavior most strongly connected with poor learning; before this study, we did not expect computer and tutor attitudes to be the best predictors of gaming. However, with this information in hand, we can now focus our efforts towards designing remediations for gaming (as opposed to other behaviors), and do so in a fashion that takes into account what we know about why students choose to game (as opposed to simply trying to prevent gaming, or using an incorrect hypothesis for why students game) – improving our chances of designing intelligent tutors that can guide all students to positive educational outcomes.

Acknowledgements

We would like to thank Angela Wagner, Jay Raspat, Megan Naim, Sue Cameron, Russ Hall, Dina Crimone and Frances Battaglia for assistance in conducting the studies at the schools, and Vincent Aleven, Joseph Beck, Cecily Heiner, Steve Ritter, and Carolyn Rosé for helpful discussions and suggestions. This work was funded by an NDSEG (National Defense Science and Engineering Graduate) Fellowship, and by NSF grant REC-043779 to

"IERI: Learning-Oriented Dialogs in Cognitive Tutors: Toward a Scalable Solution to Performance Orientation".

References

[1] 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

[2] Arbreton, A. (1998) Student Goal Orientation and Help-Seeking Strategy Use, In S.A. Karabenick (Ed.), *Strategic Help Seeking: Implications for Learning and Teaching.* Mahwah, NJ: Lawrence Erlbaum Associates, 95-116.

[3] Arroyo, I., Murrary, T., Woolf, B.P. (2004) Inferring Unobservable Learning Variables From Students' Help Seeking Behavior. *Proceedings of the Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes, at the 7th International Conference on Intelligent Tutoring Systems*, 29-38

[4] Baker, R.S., Corbett, A.T., and Koedinger, K.R. (2004a) Detecting Student Misuse of Intelligent Tutoring Systems. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems,* 531-540.

[5] Baker, R.S., Corbett, A.T., and Koedinger, K.R. (2004b) Learning to Distinguish Between Representations of Data: a Cognitive Tutor That Uses Contrasting Cases. *Proceedings of the International Conference of the Learning Sciences,* 58-65.

[6] Baker, R.S., Corbett, A.T., Koedinger, K.R., and Wagner, A.Z. (2004c) Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game the System". Proceedings of ACM CHI 2004: Computer-Human Interaction, 383-390.

[7] Bartholomé, T., Stahl, E., & Bromme, R. (2004). Help-seeking in interactive learning environments: Effectiveness of help and learner-related factors in a dyadic setting. *Proceedings of the International Conference of the Learning Sciences: Embracing diversity in the learning sciences*, 81-88.

[8] Beck, J. (2004) Using Response Times to Model Student Disengagement. *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments*, 13-20.

[9] Conati, C., Maclaren, H. (2004) Evaluating a Probabalistic Model of Student Affect. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems,* 55-64.

[10] Corbett, A.T., Anderson, J.R. (1995) Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction,* 4, 253-278.

[11] deVicente, A., Pain, H. (2002) Informing the Detection of the Students' Motivational State: An Empirical Study. *Proceedings of the Sixth International Conference of Intelligent Tutoring Systems,* 933-943.

[12] Dweck, C.S. (1975) The Role of Expectations and Attributions in the Alleviation of Learned Helplessness. *Journal of Personality and Social Psychology,* 54 (1), 674-685.

[13] Elliot, A.J., Harackiewicz, J.M. (1996) Approach and Avoidance Achievement Goals and Intrinsic Motivation: A Mediational Analysis. *Journal of Personality and Social Psychology,* 70 (3), 461-475.

[14] Elliot, E.S., Dweck, C.S. (1988) Goals: An Approach to Motivation and Achievement. *Journal of Personality and Social Psychology,* 31 (4), 674-685.

[15] Frantom, C.G., Green, K.E., Hoffman, E.R. (2002) Measure Development: The Children's Attitudes Toward Technology Scale (CATS). *Journal of Educational Computing Research,* 26 (3), 249-263.

[16] Harnisch, D.L., Hill, K.T., Fyans, L.J. (1980) Development of a Shorter, More Reliable, and More Valid Measure of Test Motivation. Paper presented at the 1980 annual meeting of the National Council on Measurement in Education. ERIC Document # ED193273.

[17] Martin, J., vanLehn, K. Student Assessment Using Bayesian Nets. *International Journal of Human-Computer Studies,* 42 (1995), 575-591.

[18] Martínez Mirón, E.A., du Boulay, B., Luckin, R. (2004) Goal Achievement Orientation in the Design of an ILE. *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments*, 72-78.

[19] 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? Sixth International Conference on Intelligent Tutoring Systems (2002) 320-329

[20] Mueller, C.M., Dweck, C.S. (1998) Praise for Intelligence Can Undermine Children's Motivation and Performance. *Journal of Personality and Social Psychology,* 75 (1), 33-52.

[21] Sarason, S.B. (1978) *Anxiety in Elementary School Children: A Report of Research.* Westport, CT: Greenwood Press.

[22] Schofield, J.W. (1995) *Computers and Classroom Culture.* Cambridge, UK: Cambridge University Press.

[23] Wood, H., Wood, D. (1999) Help Seeking, Learning, and Contingent Tutoring. *Computers and Education,* 33, 153-159.