The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate?

Mihaela COCEA [a, 1], Arnon HERSHKOVITZ [b], Ryan S.J.d. BAKER [c]
[a]*London Knowledge Lab, Birkbeck College, University of London, UK*
[b]*Knowledge Technology Lab, School of Education, Tel Aviv University, Israel*
[c] *Human Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA*

Abstract: Both gaming the system (taking advantage of the system's feedback and help to succeed in the tutor without learning the material) and being off-task (engaging in behavior that does not involve the system or the learning task) have been previously shown to be associated with poorer learning. In this paper we investigate two hypotheses about the mechanisms that lead to this reduced learning: (a) less learning within individual steps (immediate harmful impact) and (b) overall learning loss due to fewer opportunities to practice (aggregate harmful impact). We show that gaming tends to have immediate harmful impact while off-task tends to have aggregated harmful impact on learning.

Keywords: gaming the system, off-task behavior, discovery with models, educational data mining

1. Introduction

Gaming the system (taking advantage of the system's feedback and help to succeed in the tutor without learning the material) and off-task behavior (engaging in behavior that does not involve the system or the learning task) are two forms of student behavior within intelligent tutoring systems (ITSs) that are associated with reduced learning [1]. This association is weaker for off-task behavior as some studies indicate a significant negative correlation between off-task behavior and learning in intelligent tutor software [2], while in other studies the correlation is not statistically significant [1, 2]. However, the harmful impact of gaming on learning has been observed in several studies: [1, 3, 4].

Two types of gaming the system have been observed in ITSs: help abuse and systematic trial and error. However, it has been argued that a more important distinction is between "harmful" and "non-harmful" gaming [5, 6]: (a) "harmful" gaming tends to occur on steps that the learners know least well and is associated with poor learning; (b) "non-harmful" gaming tends to occur on steps that the learners already know and is not associated with poor learning. Multiple detectors of gaming behavior have been reported [4, 5, 7, 8]. Detectors of both harmful and non-harmful gaming have been validated to transfer successfully to new lessons and students they were not initially trained on [6].

1 Corresponding Author: Mihaela Cocea, London Knowledge Lab, Birkbeck College, 23-29 Emerald Street, WC1N 3QS, London, UK; E-mail: mihaela@dcs.bbk.ac.uk.

Multiple types of off-task behavior have been documented in ITSs, including talking to other students about unrelated topics, surfing the web, and disrupting other students [9, 10]. A model that can detect off-task behavior was built in [9] and has been proven to successfully distinguish between off-task behavior and on-task conversation, though some on-task conversation is still captured in the off-task model.

Several interventions have been proposed to address off-task and gaming behaviors and prevent or reduce their occurrence, building on the previously mentioned models that can detect these behaviors. These interventions include supplementary exercises on the material avoided by gaming and display of negative emotions through an animated agent; results showed a decrease in gaming behavior's occurrence and better learning when supplementary material is received [3]. Other forms of feedback, including just-in-time messages (encouraging the student to try harder or ask a teacher for help), and passive continual visual feedback have been attempted; these approaches have been shown to decrease the frequency of gaming behavior – however this decrease does not necessarily lead to better learning [11]. Nevertheless, when feedback is integrated with self-monitoring activities, positive impacts on learning have been achieved [12]. Self-monitoring has also been proposed for off-task behavior, based on successes in using self-monitoring to reduce off-task behavior in classroom settings [12, 13].

Even if there is progress towards addressing these behaviors within learning systems, however, our understanding of ***how*** these behaviors impact learning is still fairly rudimentary. In particular, why does gaming the system appear to impact learning more than off-task behavior (e.g. [1, 2])? Progress towards a complete and predictive science of learning will require understanding not just which behaviors are associated with poorer learning, but also the mechanisms which determine how those behaviors lead to poorer learning. This richer understanding may help us discern between the results of different gaming interventions, towards eventually designing systems that can respond to differences in student behavior in more precise and sophisticated fashions.

The study presented in this paper aims to uncover these mechanisms by investigating how the poorer learning associated with each behavior manifests itself. In particular, is each behavior associated with poorer learning in an immediate fashion, where the student does not learn on the current opportunity to practice the skill (perhaps a student games, and does not learn the gamed step), or in an aggregate fashion (perhaps the time spent off-task results in the student having less total time to spend on learning and thus fewer opportunities to practice each skill)? Our initial hypothesis was that gaming has immediate effects and off-task behavior has aggregate effects; within this paper, we use computational modeling methods, in combination with machine-learned models of these phenomena, in order to investigate whether this hypothesis is correct.

2. Data and Data Processing

For examining the hypothesis, we used logged data from four tutor lessons (scatter plots, geometry, percents, and probability), drawn from a middle-school Cognitive Tutor mathematics curriculum [14]. Cognitive Tutor course curricula combine whole-class and small-group learning activities 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. All data came from classes which were held during 2003-2005 in two school districts in suburban Pittsburgh, PA. The tutor uses cognitive models of problem-solving that were developed based on the advanced computer tutoring theory (ACT-R) [14].

In previous studies, these data were coded for gaming [6] and off-task behavior [9], using models that make predictions as to whether each action involves gaming or off-task behavior. Models of both "harmful" and "non-harmful" forms of gaming were used [5, 6]. Observational data about off-task and gaming was also collected – details can be found in [1]. In addition, pre and post tests were given to the majority of the students for three of the lessons (scatter plots, geometry, percents) [1]. (Some students missed tests due to class absence; students using the probability lesson were inadvertently given the wrong tests).

Aggregation of actions into *steps* (aka *learning opportunities*) was done; a step is a student's set of consecutive actions involving a single knowledge component (KC) [15] (i.e. actions involving the same KC in a subsequent problem are considered a different step). Variables for each step were computed as follows:

- error – indicating whether the first action within the step was wrong (1), correct (0), or a help request (1);
- error in next step – indicating whether the next step of the same student within the same knowledge component (KC) was wrong;
- off-task (OT) behavior, harmful gaming (HG), non-harmful gaming (NHG) – were coded as '1' if at least one action within the step was detected as OT/HG/NHG, and '0' otherwise.

The data set used to examine the immediate impact hypothesis (i.e., not including pre-post grades) included 72,845 steps (296 students, 4 classes, 108 knowledge components); the dataset used to examine the aggregate impact hypothesis included 387 student-class pairs (287 students, 3 classes).

3. Results

The analytical approach used was inspired by Beck's learning decomposition method [16], where learning over time is assessed in terms of events that occur in the student's learning process.

3.1. Off-task Behavior and Immediate Learning

We assess whether off-task behavior was associated with immediate poorer learning, by setting up a logistic regression model, where performance on a given skill at a given time is predicted based on the number of steps on this skill where the student previously engaged in off-task behavior. The best fitting model is as follows, where the parameters for *Student* and *KC* (knowledge component) vary for each student (total of 296) and each KC (total of 108):

$$\mathrm{Error}_{n+1} = \frac{1}{1 + e^{.005 * \mathrm{Off\text{-}task}_n + \beta_2 * \mathrm{Student} + \beta_1 * \mathrm{KC} - 5.512}}$$

For this model, the chi-square statistic and significance, and R^2 value are displayed in Table 1, as well as the likelihood ratio significance of the individual variables, indicating their contribution to the model; a value less than .05 indicates that the contribution is statistically significant. From these results, we conclude that off-task behavior is not significantly associated with immediate learning loss.

Table 1. Results of logistic regression model for off-task behavior

Chi-square	Sig.	R^2	Variables	Variable Significance
$\chi^2(403) = 10755.34$	.000	.233	Off-task	.961
			Student	.000
			KC	.000

3.2. Gaming the System and Immediate Learning

We assess whether gaming the system was associated with immediate poorer learning, by setting up a logistic regression model. In this case, performance on a given skill at a given time is predicted based on the number of steps on this skill where the student previously engaged in gaming behavior; we distinguish between harmful gaming (HG) steps and non-harmful gaming steps (NHG). The best fitting model is as follows:

$$\mathrm{Error}_{n+1} = \frac{1}{1 + e^{-.156 * \mathrm{HG}_n - .080 * \mathrm{NHG}_n + \beta_2 * \mathrm{Student} + \beta_1 * \mathrm{KC} + 5.532}}$$

Results of the gaming-related logistic regression model are described in Table 2: chi-square statistic and significance, R^2 value and the likelihood ratio significance of the independent variables, indicating their contribution to the model. Within this model, harmful gaming is statistically significantly associated with less learning, at the step by step grain-size. Surprisingly, non-harmful gaming was also associated with less learning at the step by step grain-size, though to a much lower degree than harmful gaming, and only marginally significantly.

Table 2. Results of logistic regression model for harmful/non-harmful gaming-the-system behavior

Chi-square	Sig.	R^2	Variables	Variable Significance
$\chi^2(404) = 10784.78$	.000	.234	HG	.000
			NHG	.054
			Student	.000
			KC	.000

3.3. Off-task Behavior and Aggregate Learning

We assess whether off-task behavior is associated with poorer learning in an aggregate fashion, using a two-step analysis. First, the correlation between overall off-task behavior (measured as the percent of off-task steps out of all steps) and total number of steps was computed for all lessons and each lesson individually (for three out of four lessons, as data for the probabilities lesson was not available). The results displayed in Table 3 show that overall off-task behavior is negatively correlated with the number of steps, meaning that off-task behavior is associated with less practice; in other words, more time spent off-task means fewer opportunities to practice.

For the second step, linear regression was applied to study the factors that contribute to post-test performance, with pre-test and total number of steps as independent variables. The results presented in Table 4 suggest that the total number of steps is positively associated with post-test results, i.e., more practice is associated with better performance. We can explain the relationship between off-task behavior and poorer learning, at least in part as: off-task behavior reduces the number of on-task steps, and these steps represent missed opportunities for better learning.

Table 3 – Correlations between overall off-task behavior and number of steps

Overall correlation: -0.347** (N=386)					
Geometry		**Percents**		**Scatter Plots**	
Correlation	N	Correlation	N	Correlation	N
-.392**	108	-.416**	51	-.375**	227

** *p<.01*

Table 4 – Linear regression of post-test by pre-test and total number of steps

Post-test = β_2*Pre-test + β_1*Steps + β_0						
Lesson	**F**	**Model Significance**	**R^2**	**Variables**	**β_i**	**Variable Significance**
Geometry	F(2, 105)=14.64	.000	.218	Pre-test	.297	.000
				Steps	.328	.000
Percents	F(2, 32)=5.65	.008	.261	Pre-test	.165	.294
				Steps	.456	.006
Scatter Plots	F(2, 203)=23.75	.000	.190	Pre-test	.256	.000
				Steps	.297	.000

3.4. Gaming the System and Aggregate Learning

In a similar way to the previous analysis, we assess whether gaming the system is associated with poorer learning in an aggregate fashion, in a two-step manner: (1) correlation between observed gaming behavior and the number of non-gaming steps, and (2) linear regression for post-test scores, investigating the relationship to the number of gaming steps and non-gaming steps.

First, the correlation between observed gaming and the number of non-gaming steps was investigated. Ideally, we would want to investigate the effects of each type of gaming behavior (harmful and non-harmful) on the number of non-gaming steps. However, as the observational data did not differentiate between harmful and non-harmful gaming (which have not yet been distinguished by human observers, likely because the main difference appears to be the context in which the behavior occurs [e.g., 5]), there is only one category: gaming. Indicators of the overall occurrence of harmful and non-harmful gaming behavior could be derived from the gaming prediction model. However, this has a major drawback: we would correlate two outcomes of the same prediction model, which would bring into question the validity of results.

Gaming (as observed behavior) is negatively correlated with the number of non-gaming steps in the model (*r(386)*= –.133, p<.01), indicating that gaming behavior is associated with *fewer* non-gaming steps, or, in other words, with fewer opportunities to practice. Unsurprisingly, gaming is positively correlated with the number of gaming steps in the model (*r(386)*=.291, p<.01).

Second, linear regression was applied to study the relationship of gaming and non-gaming steps to post-test performance. The model, displayed in Table 5, shows that in a combined model, the number of non-gaming steps predicts better learning in each tutor lesson, while the number of gaming steps only directly predicts differences in learning in the Percents lesson.

However, as the number of non-gaming steps is associated with learning, and gaming is associated with fewer non-gaming steps, the evidence is consistent with the hypothesis that gaming the system has a negative aggregate impact on learning.

Table 5 – Linear regression of post-test by pre-test, non-gaming steps and gaming steps

Post-test = β_3*Pre-test + β_2*Non-GSteps + β_1*GSteps + β_0						
Lesson	**F**	**Model Significance**	**R^2**	**Variables**	**β_i**	**Variable Significance**
Geometry	F(2, 105)=15.33	.000	.226	Pre-test	.296	.001
				Non-GSteps	.341	.000
				GSteps	-.004	.966
Percents	F(2, 32)=11.09	.000	.512	Pre-test	.029	.828
				Non-GSteps	.569	.000
				GSteps	-.332	.016
Scatter Plots	F(2, 203)=22.84	.000	.190	Pre-test	.251	.000
				Non-GSteps	.246	.001
				GSteps	.091	.206

4. Discussion and Conclusions

In this paper, we have presented research aiming on gaining a deeper understanding of the mechanisms that lead to reduced learning among students who engage in two types of behavior: gaming the system and off-task. This work can potentially contribute to a design of better interventions to address these potentially harmful ways of interacting with learning environments.

Both gaming the system and off-task behavior have been previously shown to be associated with poorer learning. The evidence seen here suggests that they do so through different mechanisms and, consequently, may be best addressed through different types of interventions. Gaming the system is associated with both *immediate* poorer learning (strongly) and *aggregate* poorer learning (more weakly). Off-task behavior, on the other hand, appears to only be associated with poorer learning at an *aggregate* level (strongly). The apparent immediate impact of gaming, at step level, appears to be due to lack of learning at that very step where the gaming occurred; in other words, by gaming, an opportunity to learn is wasted. The apparent aggregate impact of the two behaviors, considerably stronger in the case of off-task behavior, is cumulative. Poorer performance seems to occur due to fewer learning opportunities.

Understanding these mechanisms has the potential to lead to a more informed intervention to improve learning. For off-task behavior some possibilities are: (a) remind students that going off-task will just increase the time practicing until they master the lesson; (b) graph for students each day the amount their progress was reduced by off-task behavior [12]; (c) inform the students' teacher/tutor/parents how much they are off-task.

Gaming the system is more strongly associated with learning gains (negatively) than off-task behavior is [1, 2]. The evidence in this paper suggests that this difference may be because gaming behavior reduces learning both immediately and in the aggregate, whereas off-task behavior has no immediate effects on learning. Gaming the system can also lead the learning software to incorrectly assess the student's knowledge level, if the knowledge assessment does not integrate information about gaming [e.g., 9]. Providing supplementary exercises on the material on which gaming occurred and display of negative emotions has been shown to improve students' learning [3]. This study suggests that supplementary exercises may improve learning because they disrupt the immediate negative effects of gaming. It may also be useful to inform teachers/tutors/parents as to which material a student avoided by gaming, so that additional remediation can be offered.

In our study, we used a function that labeled steps as off-task, harmful gaming or non-harmful gaming after one action within the step was detected as such. However, other functions could be used, such as average and weighted average. It would be interesting to see how the results could differ for alternative aggregation functions.

Past studies have suggested that both gaming and off-task are associated with disliking math [9, 13]. Affective states also seem to play a significant role in the occurrence of these two behaviors. Frustration has been associated with gaming in intelligent tutoring systems [4, 13]. However, time-series analyses [e.g., 17] have suggested that frustration co-occurs with gaming, but does not precede it; interestingly enough, boredom and confusion have been observed to precede and co-occur with gaming, within simulation problem-solving games (which are, it should be noted, fairly different from intelligent tutors, and which see a fairly different pattern of affect and behavior from students [cf. 18]).

Combining understanding of the affect and motivation that underlie the choices to engage in gaming the system and off-task behavior, with evidence as to the mechanisms influencing how these behaviors impact learning, creates the potential for a new generation of adaptive software that influences student behavior through interventions which are effective, individualized, and minimally disruptive. Exploring these possibilities will be an important area of future research.

Acknowledgements. The authors would like to thank to Ken Koedinger for his valuable comments and suggestions. This work was funded by the Pittsburgh Science of Learning Center, National Science Foundation award SBE-0354420, and NSF grant REC-043779 to "IERI: Learning-Oriented Dialogs in Cognitive Tutors: Toward a Scalable Solution to Performance Orientation". This research was also supported by the MiGen Project, funded by the ESRC/EPSRC Teaching and Learning Research Programme (Technology Enhanced Learning), award RES-139-25-0381, and the Science and Technology Education Center (SATEC), Tel Aviv University.

References

[1] Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., and Wagner, A.Z. (2004). Off-task behavior in the cognitive tutor classroom: when students "game the system". In: *Proceedings of SIGCHI conference on Human Factors in Computing Systems*, 383-390. Vienna, Austria.

[2] Baker, R.S.J.d., Corbett, A.T., Roll, I., Wagner, A.Z., and Koedinger, K.R., *The Relationship Between Gaming the System and Learning in Cognitive Tutor Classrooms*. Manuscript under review.

[3] Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Evenson, E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., and Beck, J.E. (2006). Adapting to When Students Game an Intelligent Tutoring System. In: *Proceedings of 8th International Conference on Intelligent Tutoring Systems*, 392–401.

[4] Walonoski, J.A. and Heffernan, N.T. (2006). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In: *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 382-391. Jhongli, Taiwan.

[5] Baker, R.S.J.d., Corbett, A., and Koedinger, K., *Detecting student misuse of intelligent tutoring systems*, in *the 7th International Conference on Intelligent Tutoring Systems*. 2004: Maceió, Brazil, 531-540.

[6] Baker, R.S.J.d., Corbett, A.T., Roll, I., and Koedinger, K.R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction, 18*(3), 287-314.

[7] Beal, C.R., Qu, L., and Lee, H. (2006). Classifying Learner engagement through integration of multiple data sources. In: *Proceedings of The 21st National Conference on Artificial Intelligence*, 2-8. Boston, MA.

[8] Johns, J. and Woolf, B. (2006). A dynamic mixture model to detect student motivation and proficiency. In: *Proceedings of the 21st National Conference on Artificial Intelligence*, 163-168. Boston, MA.

[9] Baker, R.S.J.d. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In: *Proceedings of Proceedings of the SIGCHI conference on Human factors in computing systems*, 1059-1068. San Jose, California, USA.

[10] Stage, S.A. and Quiroz, D.R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review, 26*(3), 333-368.

[11] Roll, I., Aleven, V., McLaren, B.M., and Koedinger, K.R. (2007). Can help seeking be tutored? Searching for the secret sauce of metacongitive tutoring. In: *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, 203-210. Marina del Ray, CA.

[12] Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., and Wool, B.P. (2007). Repairing disengagement with non-invasive interventions. In: *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, 195-202. Marina del Ray, CA.

[13] Baker, R.S.J.d., Walonoski, J.A., Heffernan, N.T., Roll, I., Corbett, A.T., and Koedinger, K.R. (2008). Why Students Engage in “Gaming the System” Behavior in Interactive Learning Environments. *Journal of Interactive Learning Research, 19*(2), 185-224.

[14] Koedinger, K.R. and Corbett, A.T., Cognitive tutors: Technology bringing learning sciences to the classroom, in *The Cambridge handbook of the learning sciences*, R.K. Sawyer, Editor. 2006, Cambridge University Press: New York, NY, 61-77.

[15] Vanlehn, K. (2006). The Behavior of Tutoring Systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227-265.

[16] Beck, J.E. (2006). Using learning decomposition to analyze student fluency development. In: *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems*, 21-28. Jhongli, Taiwan.

[17] Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sevilla, L.R.S., Sugay, J.O., Tep, S., and Viehland, N.J.B. (2007). Affect and usage choices in simulation problem solving environments. In: *Proceedings of the 13rd International Conference on Artificial Intelligence in Education*, 145-152. Marina del Ray, CA.

[18] Rodrigo, M.M.T., Baker, R.S.J.d., D’Mello, S., Gonzalez, M.C.T., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sugay, J.O., Tep, S., and Viehland, N.J.B. (2008). Comparing learners' affect while using an intelligent tutoring systems and a simulation problem solving game. In: *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 40-49. Montreal, Canada.