A shorter version of this paper is due to be published as follows:
Walonoski., J. & Heffernan, N.T, (2006) Prevention of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In Ashley & Ikeda (Eds) *The Proceedings of the 8th International Conference on Intelligent Tutoring System,*. Springer Verlag.
See also http://www.wpi.edu/Pubs/ETD/Available/etd-010806-205001/

(This paper has been submitted to ITS2006.)

Prevention of Off-Task Gaming Behavior in Intelligent Tutoring Systems

Jason A. Walonoski, Neil T. Heffernan

Worcester Polytechnic Institute, Computer Science Department, 100 Institute Rd,
Worcester, MA 01601 USA
{jwalon, nth}@wpi.edu

Abstract. A major issue in Intelligent Tutoring Systems is off-task student behavior, especially performance-based gaming, where students systematically exploit tutor behavior in order to advance through a curriculum quickly and easily, with as little active thought directed at the educational content as possible. The goal of this research was to develop a passive visual indicator to deter and prevent off-task gaming behavior without active intervention, via graphical feedback to the student and teachers. Traditional active intervention approaches were also constructed for comparison purposes. Our passive graphical intervention has been well received by teachers, and results suggest that this technique is effective at reducing off-task gaming behavior.

1 Introduction

Intelligent Tutoring Systems (ITS) have been shown to have a positive effect on student learning [1], however these effects may be negated by a lack of student motivation or student misuse. Research examining these issues involves studying student "gaming" of the system. A student is gaming if they are attempting to systematically use the tutors feedback and help methods as a means to obtain a correct answer with little or no work, in order to advance through the curriculum as fast or as easily as possible. Student gaming has been correlated with substantially less learning [2] therefore it is of particular importance to understand in order to maximize tutor effectiveness.

Various intervention mechanisms have been proposed to combat this off-task behavior, but they either alter the system interfaces and introduce time delays (to forcibly slow down student actions) or they are susceptible to an interesting "arms race" condition, where gaming users attempt to game the intervention techniques. These techniques tend to be active interventions by the tutoring system, which may inhibit legitimate learning efforts of on-task (non-gaming behavior) students by inappropriate invocation [2]. This research was aimed at exploring alternative methods to avoid these issues within the *Assistments* mathematics ITS [3].

We developed three gaming interventions, two being traditional active interventions, and the third being a passive deterrent or prevention mechanism. For the passive intervention, a graphical software component was designed and implemented featuring visual indicators of student actions over time, which was featured prominently on screen, for the student and any observing teacher to easily see and interpret. Effectiveness of the passive graphical component on gaming behavior was compared to the traditional active approaches and was evaluated using both subjective and objective methods.

2 Prevention of Gaming

Within ITS there have been a variety of approaches towards remediation of this undesirable behavior in students [2], which are mostly active interventions focused on combating student gaming, with few approaches focused on prevention.

Active interventions can be informally categorized as either static or dynamic. Static interventions apply to all students and are inherent in the design of the tutor, while dynamic interventions are only invoked when triggered by some mechanism. For example, one approach to static interventions on help seeking has been the removal of all bottom-out hints (the hints that directly supply the answer to a question) or only supplying them after a variable time delay. On the other hand, with active interventions that are dynamic, the system dynamically detects occurrences of gaming behavior in real-time and actively intervenes by identifying and explaining this behavior to the offending student [2]. For clarification purposes, an illustration of the differences between active and passive, and static and dynamic strategies is illustrated in Figure 1.

	Static	**Dynamic**
Active	Supplying multi-level hints with a variable time delay at each successive level	Intervening after rapid and repeated hinting, with explanation and encouragement messages
Passive	Complete removal of bottom-out hints	Graphical plot of the students tutoring session illustrating progress with indicators of actions

Figure 1. Two Dimensions of Gaming Strategy with Examples

Active interventions have two drawbacks. First, the potential to fuel gaming arms races, where gaming students adapt to the interventions and even attempt to game them. And second, the unfair penalization of non-gaming students by inappropriate invocation of interventions, thereby inhibiting legitimate learning efforts. For example, by not allowing a student to be able to ask for hints when they are truly needed. Given that students who engage in off-task gaming behavior are such a small percentage of all users (estimated between 5% and 15% in the *Assistments* system), the unfairness resulting from improper application of active interventions can heavily unbalance the usefulness of the tutoring software against the majority of on-task students [2].

We hypothesize that a dynamic yet passive intervention could be developed with none of the drawbacks of active interventions, but with all of their benefits. Informally, an intervention is defined as being passive if it does not alter the operation or behavior of the tutoring software in any functional way, but effectively alters the behavior of the students. Of course, then the objective becomes the creation of a passive intervention that alters student behavior in a positive manner, such as preventing or eliminating off-task gaming.

Dynamic passive interventions should offer several hypothetical advantages over existing active approaches. The first advantage being that dynamic passive interventions should require no modification of the tutoring software's functional behavior, effectively eliminating the gaming arms race while on-task students would no longer be unfairly penalized by improper invocation of the intervention mechanisms. The second advantage would the visualization of student actions and progress, which should provide a vehicle for the summarization of student behavior through emergent visual patterns. With these emergent visual patterns featured prominently on-screen for easy viewing by the student and teachers, off-task gaming behavior might be prevented through Panopticon-like paranoia (when a fear of being watched, without knowing whether one is being watched at any given moment, causes self-corrective behavior) [4].

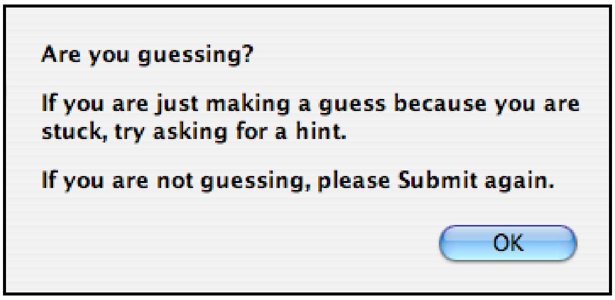
To explore the hypothetical advantages of dynamic passive interventions over traditional active interventions, a passive yet prominent graphical component was developed with the goal of preventing gaming behavior among students who gamed (and other off-task students), but ignored by non-gaming and on-task students. Such a component should (1) allow teachers to easily identify gaming behavior via emergent visual patterns, (2) thereby correcting and preventing gaming behavior in the students by the students themselves, and (3) providing a launching point for teacher intervention where gaming behavior is identified or student misunderstandings are shown.

2.1 Design and Development of Interventions

Our development of interventions focused on a dynamic passive intervention and two dynamic active interventions for comparison purposes. Since gaming behavior has two hallmark appearances (rapid-fire guessing-and-checking and hint/help abuse), two active interventions were separately developed to respond individually to each type of gaming behavior. Only one passive intervention was developed, with the intention of preventing both types of gaming simultaneously and eliminating the active intervention gaming arms race.

The two active interventions were triggered by simple gaming detection algorithms, which marked a student as guessing-and-checking or abusing-hints *prima facie* of the appropriate surface-level characteristics. Once a student was marked for a particular intervention, the student would encounter that intervention during their very next action if they attempted to answer a question (if they were suspected of rapid guessing) or if they requested a hint (if they were suspected of abusing hints), until the suspicion of gaming sufficiently decayed. On the other hand, the passive intervention had no triggering mechanism, as it was always visible and potentially passively influencing the tutoring session at all times.

The guessing-and-checking (GC) detection algorithm has a few simple rules based on student actions. If a student incorrectly answers any problem step on three consecutive attempts, they are marked as guessing-and-checking the problem. If a student GCs three or more different steps, they are marked as a GC gamer and will then be eligible for the GC gaming intervention. The GC intervention will then be fired on the next problem attempt. If a student completes an entire problem without GC a single step, then their total GC score is reduced by 1. When the GC score is greater than or equal to 3 for a given student, they are considered to be a GC gamer. When the GC score is less than 3, or is reduced to be less than 3, the student is considered to be a non-gamer. The GC gaming intervention is a JavaScript alert message, worded as carefully as possible as not to unintentionally insult the student or unjustly accuse them of doing anything improper. A screen capture of the alert message appears in Figure 2.



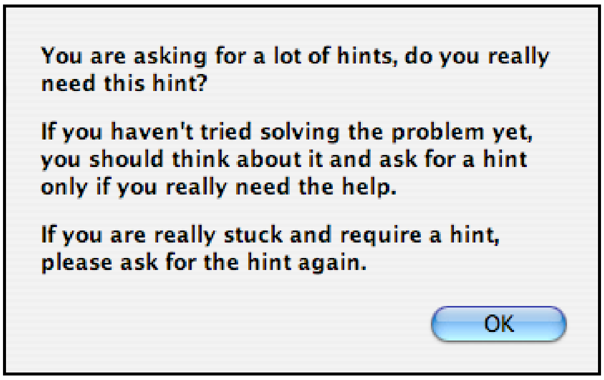
Are you guessing?

If you are just making a guess because you are stuck, try asking for a hint.

If you are not guessing, please Submit again.

Figure 2. Guess-and-Check (GC) Gaming Intervention

The hint-abuse (HA) detection algorithm is very similar to the GC detection algorithm, and also only has a few simple rules based on student actions. If a student requests a bottom-out hint, they are marked as hint-abusing the question. If a student does this on three or more different problem steps, they are marked as a HA gamer and will then be eligible for the HA gaming intervention. The HA intervention will then be fired on the next hint request (regardless if it is a bottom-out hint or not). If a student completes an entire problem without abusing hints on a single problem step, then the their total HA score is reduced by 1. When the HA score is greater than or equal to 3 for a given student, they are considered to be a HA gamer. When the HA count is less than 3, or is reduced to be less than 3, the student is considered to be a non-gamer. The HA gaming intervention is a JavaScript alert message, worded as carefully as possible as not to unintentionally insult the student or unjustly accuse them of doing anything improper. A screen capture of the alert message appears in Figure 3.



You are asking for a lot of hints, do you really need this hint?

If you haven't tried solving the problem yet, you should think about it and ask for a hint only if you really need the help.

If you are really stuck and require a hint, please ask for the hint again.

Figure 3. Hint-Abusing (HA) Gaming Intervention

Unlike the two active interventions, which are visually and functionally simple, the passive intervention is visually complex and has a more sophisticated generation mechanism. The component graphically plots (as opposed to enumerating as text) in a horizontal timeline all recorded student actions (such as problem attempts, hint requests, bottom-out hints), the amount of time each action took, and the outcome of the action. An example of the component after a few minutes of on-task use is shown in Figure 4.

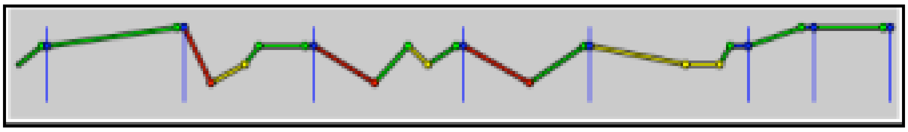
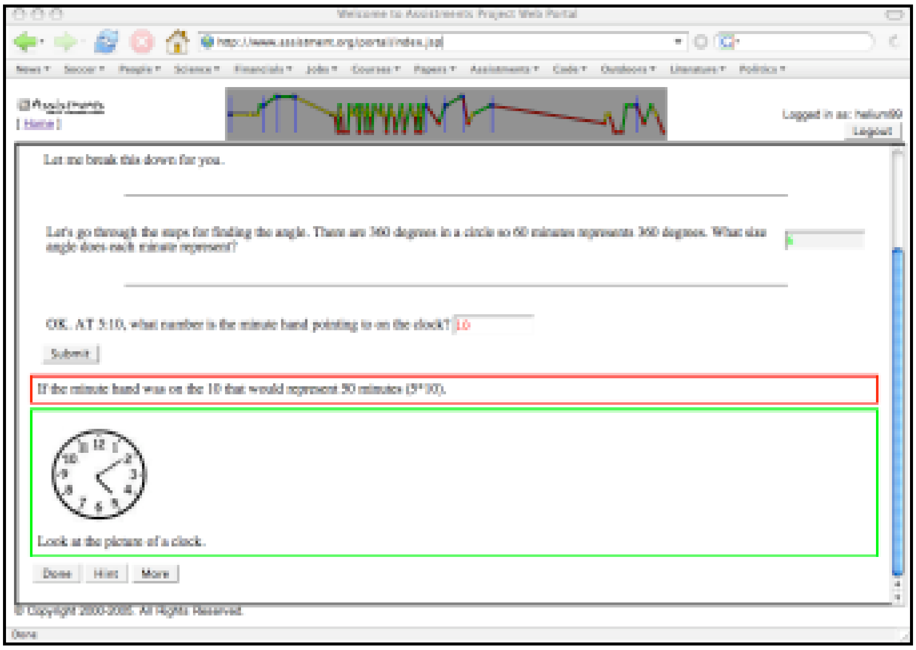


Figure 4. Passive Intervention Example, On-Task

The design is basically a timeline (time progresses from left to right) that charts actions, where each action is indicated by a small colored point, and sequential actions are connected via colored lines. Each point has associated summary text (not shown) that identifies the action and relevant details and results of the action on mouse-over. The horizontal distance between points represents the amount of time between the actions (the graphic scales as time passes, so the distance is dependent on the length of the current session). The type of action and its result determines the vertical distance between points, on a range where correct first attempts are the highest and bottom-out hints are the lowest. Furthermore, the action result is also represented by the color of the point and the color of the line connecting it to the previous point. The ubiquitous traffic-light color conventions of modern society are used here, where green is implicitly “good,” yellow is “caution,” and red is interpreted as “bad.” Green is used when questions are answered correctly, yellow is used on hints, and red is used on incorrect answers and bottom-out hints. Additionally, blue points with blue vertical lines are used to mark the transition between problems, while pink points with accompanying pink vertical lines are used to mark a problem replays. As a summary estimate of the student’s performance, the background color of the graphic (light grey in Figure 4) changes on a gradient from white to black based on the percentage correct of attempts (at one end of the spectrum, the color white is shown on greater than 90% correct, and on the other end of the spectrum, the color black is shown on less than 10% correct). Figure 5 shows another example of the component as it is situated within the *Assistments* ITS.



Welcome to Assistments Project Web Portal
http://www.assistment.org/portal/index.jsp
News Soccer People Science Financials Jobs Courses Papers Assistments Code Outdoors Literature Politics
Assistments
[Home]
Logged in as: helium99
Logout
Submit
Done
Hint
More
© Copyright 2000-2006. All Rights Reserved.
Done

Let me break this down for you.

Let's go through the steps for finding the angle. There are 360 degrees in a circle so 60 minutes represents 360 degrees. What size angle does each minute represent?

OK. AT 5:10, what number is the minute hand pointing to on the clock? 10

If the minute hand was on the 10 that would represent 50 minutes (5*10).

Look at the picture of a clock.

Figure 5. *Assistments* Screenshot featuring the Passive Intervention

The graphical component has two main functions – displaying a summary of user actions over time that would clearly classify the gaming-status of a student to an observer (such as a teacher) by emerging patterns of indictors (see Figures 6 and 7 for example of charts capturing off-task gaming behavior), and allowing a teacher to ask, "what did you do here?" and prompt an investigation into the student actions which could reveal a lack of motivation, or even a fundamental gap in the student's knowledge. The first function is addressed toward the goal of gaming prevention, and the second function is addressed toward a secondary goal of assisting a teacher in identifying student weaknesses or misunderstandings via a trace of actions through the student's session.

Both of these functions, classifying gaming-status by an emergent pattern of indicators and the identification of student weaknesses, is captured in Figure 6, which illustrates how a typical guessing-and-checking gaming student's actions would plot. As each red point and line represents an incorrect action, and the green points and lines represent a correct action, interpretation of this chart is fairly straightforward: a large series of rapid incorrect attempts is occasionally interrupted by a lone correct attempt (a lucky guess, perhaps) or the transfer to the next problem, only to be followed by a new stream of incorrect actions. This chart can only embody two possibilities: the student is engaging in off-task guessing-and-checking gaming behavior, or they have a fundamental lack of knowledge relating to the problem and are making lots of genuine errors without seeking help. Either way, teacher intervention is probably appropriate in such a case.

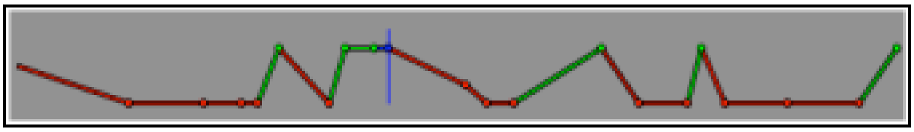


Figure 6. Passive Intervention Example, Guessing-and-Checking

Similarly, Figure 7 shows the resulting graph of help-abuse gaming. A series of rapid hint requests (the yellow points and lines in the vertical-middle of the component), followed immediately by a steep red drop (the bottom-out hint request) and then an upward green line (answering correctly after the answer was directly supplied by the bottom-out hint). This pattern occasionally appears in on-task usage, but when systematically repeated, is a clear indicator of off-task help-abuse gaming behavior.

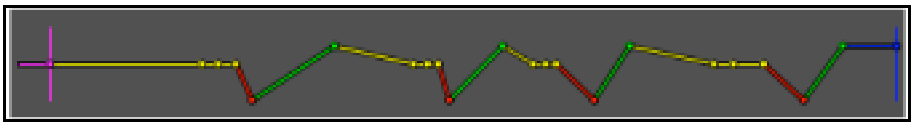


Figure 7. Passive Intervention Example, Abusing-Help

Both Figures 6 and 7 show the plots of students who were engaged in off-task gaming behavior after only a few minutes of a tutoring session. Since the graphical component scales as time progresses, the long-term identification of gaming appears slightly differently, but is even more readily apparent. Such an example appears below in Figure 8.

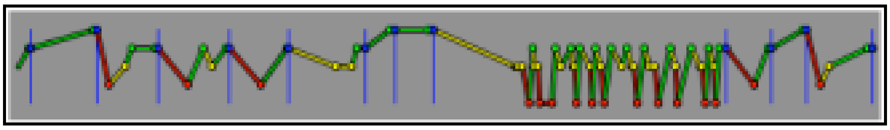


Figure 8. Passive Intervention Example, Long Session with Partial Gaming

Figure 8 shows a plot of actions from a student who has completed 12 problems (represented by the thin vertical blue lines). Usage was relatively on-task, except for the ninth problem which was completed via off-task help-abusing gaming. It is fairly obvious, even to an uninformed observer, that this portion of the graphical plot is far different than the rest of it. Not only does the graphical component plot the actions in a fairly logical manner (at the very least it is systematic), but it also does so in such a way that conspicuous patterns emerge in the case of off-task gaming behavior, especially as the length of the tutoring session increases.

2.2 Deployment of Interventions

With all three interventions designed and implemented, they were deployed within the *Assistments* system in groups with different conditions. The original plan was to have four groups: one which received only the active interventions, a second which received only the passive interventions, a third that received both interventions, and a fourth that received no interventions. Due to a variety of circumstances, these groups were not created equally. Figure 9 summarizes the different experimental groups created for evaluation purposes of the developed interventions.

Active Interventions	Passive Interventions	Number of Classes	Number of Distinct Students
False	*False*	7	442
False	*True*	4	42
True	*False*	0	0
True	*True*	63	1289

Figure 9. Summary of Experimental Groups

For the group with both the active and passive interventions, there were 63 distinct classes that ran across 31 school days with 1 to 12 classes running per day (with an average of about 7 classes per day, with a standard deviation of about 4) over an approximately 1.5 month span (31 October 2005 through 15 December 2005). Some of the classes used the tutoring system more than once during that time span. There was an average of about 78 students using the tutor per school day, with a standard deviation of about 62. The average number of students per class was about 19, with a standard deviation of about 11.

2.3 Active Intervention Results

Our primary method of analyzing the effectiveness of the active interventions was to examine which types of actions that were executed both before and after each intervention was triggered.

The stop-hinting intervention fired 1664 times in our system at the time of analysis. About 88% of the time the stop-hinting intervention was received, it was immediately followed by a hint request or bottom-out hint request, which suggests that the intervention was being ignored. The intervention was immediately followed by an attempt approximately 10% of the time, with about a third of those attempts being correct. Approximately 9% of all stop-hinting interventions were triggered at the start of a new problem before any attempt was even made. While our data suggested that the stop-hinting active intervention was being triggered at reasonable times, it was largely ineffective at stopping hint abuse.

Similar negative results were found for the effectiveness of the stop-guessing intervention. At the time of analysis, the stop-guessing intervention fired 2262 times in our system. Approximately 98% of all stop-guessing interventions were fired after incorrect attempts, and the remaining 2% were fired after problems were replayed (student tries to exit and restart a problem, which is not allowed) or a stop-hinting intervention was triggered. The stop-guessing interventions were immediately followed by hint requests about 8% of the time, and attempts the other 92% of the time. About 55% of those attempts were incorrect. This data strongly suggests that the stop-guessing active intervention was completely ineffective, as it was being almost totally ignored.

Since our simple *prima facie* algorithms triggered both interventions, there is the possibility that they were being ignored because they were being inappropriately triggered. To in attempt to evaluate these interventions without being biased by this possibility, another cluster of actions was examined using only the first intervention received by any student. This should reveal whether the interventions are effective on first encounter.

There were 236 first-time stop-hinting interventions in our system at time of analysis. Interestingly, none of the first-time encounter stop-hinting interventions were preceded by repeated hint requests; rather they were triggered immediately after a series of attempts. However, about 71% of these first-time interventions were immediately followed by hint requests (compared to the 88% overall).

At the time of analysis, there were 375 first-time stop-guessing interventions in our system. Over 96% of first-time stop-guessing interventions were triggered immediately after incorrect attempts. About 78% of these interventions were immediately followed by attempts (compared to the 92% overall). Although the active interventions are relatively more effective upon first-encounter, these numbers suggest that they are for the most part ignored by students and therefore largely ineffective. However, results of our "toggle experiment" (discussed in Section 2.5) suggest that these interventions might have some minor affect on users when combined with the passive intervention.

2.4 Passive Intervention Results

Although there were no interesting metrics available for latent response or human factors analysis, two methods were used to evaluate the passive intervention: (1) anecdotal teacher surveys and (2) analysis of the "toggle experiment" – although it is hard to separate the effects of the passive and active interventions within the toggle experiment.

A teacher survey was used to gather anecdotal evaluation of the passive intervention and was also meant as a feedback mechanism for suggestions to improve the graphical chart. A similar survey aimed at students was never administered. The teacher survey had 5 questions that were rated on a scale of 1 to 5 (1=strong disagree, 2=disagree, 3=not sure, 4=agree, 5=strongly agree), and one open response question seeking suggestions and feedback. The survey was administered to 10 teachers who had used the system with and without the interventions. Clearly, 10 teachers is not a significant sample size, but it was a useful means to gather suggestions. Teacher's had several thoughtful suggestions including altering the background gradient scheme from white-to-black to another color gradient scheme over worries of racial bias issues, keeping the background one color at all times to eliminate confusing, altering the vertical heights of plotted points to indicate increasing depth of hints or a progression of incorrect or correct attempts, removing the lines altogether and simply having various colored vertical-bars indicating actions and results, and elimination of the automatic scaling of time for consistency across different student computers. A summary of the questions and responses appears below in Figure 10.

Question	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
Do you think that the new graphical chart will aid teachers in assessing the progress and performance of their students?	0	0	1	1	8
Do you think that the new graphical chart will aid students in self-assessing their progress and performance?	0	0	1	5	4
Do you think that the new graphical chart will provide students with additional performance-based motivation?	0	0	0	5	5
Do you think that the new graphical chart will provide students with additional learning-based motivation?	0	0	2	5	3
Do you think that the new graphical chart will decrease off-task student behavior (talking, inactivity, excessive or unnecessary hinting or guessing)?	0	0	2	3	5

Figure 10. Summary of Teacher Survey Results

Other than suggestions for alterations to the stylization of the graphical plot, most teachers expressed satisfaction with the passive intervention, at least as a feedback mechanism that would assist them in helping to select which students to focus their attention on. However, the results of the toggle experiment are considered to be the definitive and overall method of evaluation for all the intervention mechanisms, including the passive intervention.

2.5 Overall Intervention Evaluation

The final evaluation for all the intervention mechanisms was completed by the "toggle experiment." The phrase "toggle experiment" simply means that the status of the active and passive interventions was toggled after an initial number of times using the *Assistments* system (between 1 and 5 sessions, depending on the school, teacher, and class a student was in), and the rate of student gaming before and after the toggling is then compared. The group that had neither passive nor active interventions before the toggle, afterwards had both enabled. And the group that originally received both interventions had neither after the toggling.

Before and after the toggling, the hint and guess scores were calculated using the same *prima facie* recognition algorithm that was used to invoke the active interventions. Before making any calculations we decided to remove any sessions that were less than 20 minutes in length (which entirely eliminated the group that had the passive intervention but not the active interventions). The results of the experiment are summarized in Figure 11.

Group	Metric	Before Toggle	After Toggle	Gaming Effect	Notes
both_none	Users (Students)	70	70	-2.82	This group originally had both the active and passive interventions turned ON, and then turned OFF after the toggling. Most kids had 20 minutes to an hour of use on days after toggling, with an average of about 30 minutes.
	Average Days	2.357	1.014		
	Avg Hint Score	2.257	0.829		
	Avg Guess Score	6.657	0.057		
	Avg Total Gaming Score	8.914	0.886		
	Avg Gaming Score Per Day	3.62	0.8		
none_both	Users (Students)	57	57	-4.45	This group originally had both the active and passive interventions turned OFF, and then turned ON after the toggling. Most kids had 20 minutes to an hour of use on days after toggling, with an average of about 30 minutes.
	Average Days	3.772	1.158		
	Avg Hint Score	8.912	1.754		
	Avg Guess Score	11.61	0.053		
	Avg Total Gaming Score	20.53	1.807		
	Avg Gaming Score Per Day	6.235	1.781		

Figure 11. Results of the Toggle Experiment

Before the toggling, the "*both_none*" group that originally received both interventions had a total gaming score that was less than half their non-intervention receiving counterparts, suggesting that the interventions were perhaps having some sort of effect. However, in order to show that those differences were not the result of some sort of selection effect in the groups (for example, merely having more students in the first group who were less inclined to participate in off-task gaming behavior), the settings were toggled. After the toggling, both groups had decreased amounts of gaming. The average student decreased gaming by 3.5 points per day, after the toggling. Those who had been receiving the interventions, and then stopped receiving them, reduced gaming on average by 2.8 points per day. However, the decrease in the group that originally had no interventions and then began receiving them after the toggling, decreased their gaming by an average of 4.4 points per day. While this outcome might simply be the result of chance, it seems to suggest that the combination of the active and passive interventions has a moderately successful effect in the reduction of off-task gaming behavior.

The statistical significance of these results were analyzed using one-side t-tests on both the "*both_none*" and "*none_both*" groups, to see if the resulting change in gaming scores after the toggle was significantly different from zero, and in both cases the answer is yes ($p < 0.0001$, in both tests).

The next question we tested was whether there really was a bigger impact in the "*none_both*" group – turning the intervention mechanisms on versus off – and did so partially as a control for other effects, such as if students learn to game more or less over time. This was done with an analysis of variance (ANOVA) and the resulting p-value of .08 suggests that turning the interventions on (*none_both*) makes a bigger impact on *prima facie* gaming than turning them off (*both_none*). One possible interpretation and explanation of these results would be that with the interventions turned on the students learn not to game, and then they do not start gaming once the interventions are turned off and go away. Further analysis might reveal whether the actual invocation or receiving of the active interventions (when they were turned on) is correlated with this decrease in gaming, as opposed to simply the *possibility* of receiving them (a student might have never seen the active interventions when they were turned on if they were not gaming). Otherwise, we might be able to conclude that the decrease in gaming was due more to the passive intervention, or perhaps other factors. We leave this for future work.

3 Conclusions

Off-task gaming behavior is a major issue within the field of ITS, since it has been correlated with poor learning. The goal of this research was to explore the intervention and prevention of this phenomenon within the *Assistments* system. Three dynamic interventions were designed: two active interventions for hint-abuse and guessing-and-checking, and one passive intervention. The interventions were analyzed and evaluated with a variety of methods, but primarily by a toggling experiment with results suggesting that the combination of the active and passive interventions successfully reduces off-task gaming behavior more effectively than no intervention mechanisms. Despite a small sample size and minor suggestions for improvement, teachers were overwhelmingly positive in their reception of the passive graphical component, as much for its instant feedback and as a launching point for targeted instruction as for any purported ability to prevent off-task gaming behavior.

This research was made possible by the US Dept of Education, Institute of Education Science, "Effective Mathematics Education Research" program grant #R305K03140, the Office of Naval Research grant # N00014-03-1-0221, NSF CAREER award to Neil Heffernan, and the Spencer Foundation. All the opinions in this article are those of the authors only.

This work would not have been possible without the assistance of the 2004-2005 WPI/CMU *Assistments* team including Mingyu Feng, Andrea Knight, Ken Koedigner at CMU, Abraao Lourenco, Michael Macasek, Goss Nuzzo-Jones, Kai Rasmussen, Leena Razzaq, Steven Ritter at Carnegie Learning, Carolyn Rose at CMU, Terrence Turner, and Ruta Upalekar.

References

1. Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). *Intelligent tutoring goes to school in the big city*. International Journal of Artificial Intelligence in Education, 8, 30-43
2. Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. (2004) *Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System"*. Proceedings of ACM CHI 2004: Computer-Human Interaction, 383-390.
3. Razzaq, L, Feng, M., Nuzzo-Jones, G., Heffernan, N.T. et. al (2005). *The Assistment Project: Blending Assessment and Assisting*. To appear in the Proceedings of the 12th Annual Conference on Artificial Intelligence in Education 2005, Amsterdam
4. Bentham, Jeremy. *The Panopticon Writings*. Ed. Miran Bozovic (London: Verso, 1995).