A Multi-Pronged Redesign to Reduce Gaming the System (Extended Technical Report Version)

Yuanyuan Li[1], Xiaotian Zou[1], Zhenjun Ma[1], Ryan S. Baker[2]

[1] Learnta, Inc., 1460 Broadway St., New York, NY 10036, USA
[2] University of Pennsylvania, 3700 Walnut St., Philadelphia, PA 19104, USA

Abstract. Despite almost two decades of interest in reducing gaming the system in interactive learning environments, gaming continues as a key factor reducing student learning outcomes and contributing to poorer learning outcomes. Part of the problem is the creativity of students in finding new ways to game the system. In this article, we document two gaming the system strategies that students developed in the Kupei learning system: students rapidly going through all practice sets to exhaustively obtain answers, and students quitting in the middle of a problem set after completing some problems and then returning to the practice set to enter the answers they saw. In this study, we redesigned the Kupei learning system by implementing a combined set of three interventions aimed at mitigating the impact of those gaming behaviors: 1) not allowing students to complete more than 15 problems (3 practice sets of 5) on the same topic in one day to slow down students who try to game; 2) integrated meta-cognitive feedback to encourage students to engage in behaviors that lead to better learning; 3) mandatory videos or notes to provide students another chance to learn the materials missed by gaming. Our results show evidence of a possible positive effect of the combined gaming prevention intervention, significantly reducing gaming by quitting the problem set to seek external answers and improving performance within the system. However, the approach was not as successful at mitigating gaming by exhaustively completing problem sets.

Keywords: Gaming the System, Learning engineering, Iterative redesign, Interactive learning environments

1 Introduction

Interactive learning environments are intended to create opportunities for students to learn but a substantial proportion of students choose instead to game the system, attempting to succeed by taking advantage of the regularities and properties of a system rather than by learning the material [1]. Considerable research has demonstrated negative correlations between gaming the system and student outcomes, including negative correlations with content learning [2, 1, 3], standardized examinations [4, 5], and educational outcomes years later [6, 7].

Over the last 15 years, several research groups have attempted to refine learning systems to reduce the prevalence of gaming or mitigate its impacts. A range of interventions have been proposed and investigated. One approach has been to detect when gaming occurs and then adapt in real time to detection of gaming behavior, for example by giving students supplementary exercises on the material they bypass by gaming [8], or by giving students textual messages recommending better strategies [9]. Other work has intervened in more subtle fashions, providing visualizations to students of their learning efficiency in between problems [10], or offering a "reflective nudge" which allows reflection on gaming behaviors [11]. These approaches have generally been successful at reducing the amount of gaming behavior. Findings have been less consistent in terms of learning. The approach in [12] was unsuccessful at improving learning outcomes at the domain level. By contrast, the approaches in [8, 10] improved learning outcomes but were not adopted at scale. The literature does not clearly indicate why these approaches did not scale, but it may be due to intrusiveness in the overall learning experience, the difficulty of authoring them across an entire year of content, or other non-beneficial impacts of these re-designs.

An alternate approach is to try to prevent gaming behavior in the first place. For example, Murray and VanLehn [13] discussed how a delay between each level could prevent students from requesting unnecessary help or clicking through the hint too quickly. However, they found that this approach simply led students to find an alternate strategy for gaming the system. Other researchers re-designed their system to have students spend a minimum amount of time on a problem before giving a hint [14]. However, this intervention may hinder the usefulness of help-seeking for non-gaming students; Aleven et al. 's [15] prescriptive model of help-seeking recommends that a student who is completely uncertain what to do should seek help immediately. Nonetheless, approaches that can reduce gaming non-intrusively, through design, may be more straightforward to scale than detection-based approaches, making this general strategy worthy of further investigation.

In this paper, we investigate a multi-pronged redesign of an AIED system, using three interventions in tandem to reduce students' propensity to game. The three interventions are designed to reduce gaming through overall interaction re-design, i.e. general changes to the system rather than adaptivity based on detecting gaming. We conduct a within-system quasi-experiment, investigating whether the redesigned version of the system leads to reduced gaming behavior and better within-system performance.

2 Method

2.1 Platform

Our integrated intervention was developed in the context of the Kupei learning platform (Fig 1) that supports the learning of math, English and science subjects for 4th to 9th grade Chinese students. The Kupei learning platform is used in after-school learning centers and provides additional academic support for students who struggle with content learned at school, allowing students to work through it at their own pace. Rather than teacher-led instruction, the system uses algorithms that can automatically determine which content a student should work on next based on their previous performance and interactions with the system [16].

With the Kupei learning system, students usually take less than three practice sets to achieve basic mastery (probability of mastery falls between 80% and 95%) of each concept. Therefore, we define practices on the same concept after three practice sets as extra practices. We believe that a considerable proportion of extra practice will be the result of either gaming the system or struggling with the content.

Fig. 1. Example of a user's dashboard within the Kupei learning system. I) panel showing overall learning progress for the math topics; II) detailed view of a particular math topic displaying all concepts, numbers of practice completed, and mastery levels.

Fig. 2. Student's interfaces during a practice and learning the material. I) panel showing a question from a practice set that was answered correctly; II) screenshot of the video that provides the student another chance to learn the material missed by gaming.

Kupei uses Bayesian Knowledge Tracing (BKT) [17] to estimate student proficiency in real-time. BKT assesses the probability a student knows a specific concept by their history of performance involving it. If a student's probability of knowing a concept is less than 80%, the concept is considered unmastered. When the probability falls between 80% (a cut-off used by many commercial systems) and 95% (the original cut-off in) [17], then the concept is labeled as basic mastery. If a concept's level of mastery exceeds 95% then it is labeled as advanced mastery. When studying a concept using the Kupei system, BKT assesses students' probability of mastering a concept after the first 3 items are completed. If the mastery probability is more than 95% (advanced mastery), the system stops and advances them to the next concept. In the case that the probability of mastery is lower than 80% (unmastered) after the first three items are completed, the practice stops and displays the result. However, if the probability of mastery falls between 80% and 95%, the students will continue to work on two additional items. Students who did not master the concept (whether after 3 or 5 problems) are next required to complete an integrated review on the same concepts/skills (involving video and/or lecture notes). In all situations, the learning recommendation offered after each concept (i.e. what concept to work on next) will change according to students' performance during the practice. Kupei offers students a dashboard (shown in Fig 1) where students can see their overall progress, and also their detailed work on past concepts, including mastery level achieved, problems encountered, and answers.

2.2 Gaming Behaviors

Prior to the integration of the gaming prevention intervention, gaming behaviors typically observed in the Kupei system can be divided into two types:

1. Students use an exhaustive method to obtain the correct answers of the practice sets by inputting random answers for each question of each practice set until earlier questions are re-shown. They can then go to their problem history prior to starting a new practice set and view the answers, or may simply remember past answers.
2. Students open a practice set to obtain the set of questions, then quit the practice set midway to seek answers elsewhere. Then students return to the practice set within 10 minutes and typed in the correct answers.

2.3 Design

In order to prevent the gaming behaviors mentioned above and encourage students to interact in ways that improve learning, our design aimed to simultaneously accomplish two goals: first, by increasing the costs of gaming, it is hoped that students will game the system less often; second, with less gaming behaviors, we hope that students will engage in more productive behaviors and learn more effectively.

Aiming to achieve these objectives, we designed the following three gaming prevention interventions: first, we re-designed the system so that students may not complete more than two practice sets (of five problems each) on a concept more than three times a day [18], with a pause of 36 hours before they can work on a concept again. Therefore, if students game the system and work through two practice sets quickly, they will have to delay their return to the content, preventing them from making the rapid progress (without learning) that gaming the system attempts to achieve. Second, Kupei now provides meta-cognitive feedback [9]. Prior to the beginning of the second practice set, right after the student performed poorly on the first practice set, students now receive a system reminder that states “This is today’s last chance to practice this concept! Please take it seriously”. This feedback acts as a reminder to the students about the cost of gaming -- if they now game, they will have to wait 36 hours. Third, the system now provides students another way to learn the materials missed by gaming [8]. Students who responded too quickly and failed to reach basic mastery (probability of mastery falls between 80% and 95%）are required to complete an integrated review on the same concepts/skills (involving video and/or lecture notes).

We hypothesized that this combined gaming prevention intervention would reduce gaming and mitigate its effects. We hypothesized that the limited number of practice sets per day and wait time would be successful at reducing students’ degrees of gaming by making it harder and more costly to game. The required integrated review might also encourage students to become more engaged in the content and apply themselves more seriously to solving problems. The interventions would also prevent wheel-spinning students (students performing poorly due to struggling rather than gaming the system) [19]. from continuing content that is too difficult for them and providing them with 36 hours to catch up and understand the concept better, perhaps through incubation effects, which have been shown to reduce wheel-spinning in prior research [20].

2.4 Data Collection

A within-subjects quasi-experiment was conducted comparing two 15-day periods -- a control period before the new strategy for reducing gaming behavior was adopted, and an experimental period immediately following adoption of the new strategy within the system. We analyzed data from students who studied at least 10 concepts in both two periods. This led to a sample of 343 students who worked on math, and 345 students who worked on English. 87 students participated in both conditions. We collected interaction logs from the Kupei learning system, including detailed performance on each item (including start and end time), annotated with concepts.

2.5 Statistical Analysis

To determine if gaming behavior was reduced, we compared the frequency of gaming behavior and extra practices (more than three practice sets) between the control and experimental conditions. Also, the average time spent per item per student and the percent correctness on items per concept per student are compared to see whether students studied more carefully with the new design. To study the effectiveness of our redesigned version of the system in reducing gaming behavior, we used paired Wilcoxon signed rank tests (only among the students seen in both conditions) to compare the frequency of gaming behavior and extra practices, before and after the gaming prevention interventions were introduced. We also conducted paired t-tests (separately for both English and math) to examine if the total time spent and percent correctness were statistically significantly different before and after the gaming prevention interventions were introduced. We also compared the changes between the control and experimental groups between two non-overlapping sub-groups of students: the control-gamers, students who gamed the system in the control period and the control-non-gamers, students who never gamed in the control period (this design is also seen in [8]), using a paired Wilcoxon rank sum test (frequency of extra practices) and a repeated measures ANOVA (total time spent and percent correctness).

3 Results

3.1 Frequency of Gaming Behavior by Condition

In mathematics, 93 students were control-gamers (they gamed the system during the control period) and 250 students were control-non-gamers (they did not game the system during the control period). In English, 155 students were control-gamers and 190 students were control-non-gamers. After the gaming prevention interventions were integrated, the frequency of gaming significantly reduced in both math and English. For math, the average gaming frequency per student decreased from 0.124 to 0.064, a statistically significant difference, V=5411, $p<0.01$. There were more gaming behaviors found in English practices overall, 0.211 in the control period, and 0.157 in the experimental period, but the decrease was also significant, V=13144, $p=0.047$.

In terms of specific behaviors, there was a statistically significant reduction in gaming by quitting to seek answers. Among all students, the average frequency of gaming by quitting decreased from 0.085 to 0.031 in math, V=2511, $p<0.01$; and from 0.045 to 0.026 in English, V=3315, $p=0.029$. However, there was not a significant reduction in gaming by memorizing answers. Among all students, for math, the average frequency of students gaming by memorizing answers was 0.040 during the control period and 0.032 during the experimental period, V=2086, $p=0.51$. For English, the average frequency of gaming by memorizing answers was 0.166 during the control period and 0.132 in the experimental period, V=11339, $p=0.15$.

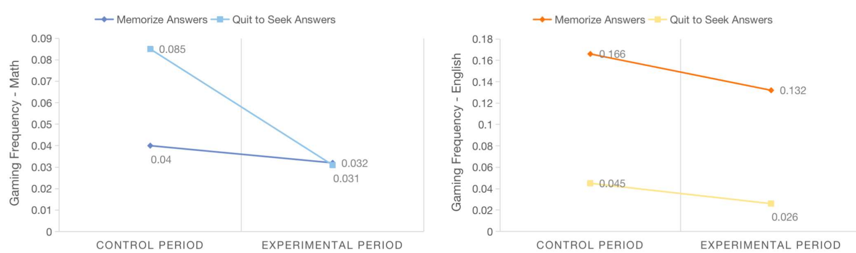

Fig. 3. The frequency of each form of gaming behavior in Math (Left) and English (Right)

3.2 Other Behavior Changes

The Proportion of Extra Practice. The average proportion of extra practice for 343 students in math decreased from 12.20% in the control period to 7.66% in the experimental period, which is statistically significant, V=25588, $p<0.01$. Similar results are found for English. The average proportion of extra practice for 345 students in English decreased from 9.00% in the control period to 4.75% in the experimental period, V=22751, $p<0.01$. Results are consistent within different student groups. The proportion of extra practice in math statistically significantly decreased for both the control-gamers, V=3246, p<0.01; and the control-non-gamers, V=10448, p<0.01. For English, the proportion of extra practice significantly decreased from control period to experimental period in both the control-gamers, V=8815, p<0.01, and the control-non-gamers, V=3128, p=0.023. As shown in Figure 3, the gaming prevention interventions produced a stronger impact on the control-gamers. For math, the proportion of extra practice in the control-gamers decreased from 23.1% in the control period to 13.4% in the experimental period, while for control-non-gamers, the proportion of extra practice decreased from 8.1% control to 5.5% experimental. According to a Wilcoxon rank sum test, the control-gamers' decrease in extra practice is significantly steeper than the control-non-gamers, W=7928, p<0.01. As for English, the proportion of extra practice in the control-gamers decreased from 15.7% control to 7.3% experimental, which is significantly steeper than the proportion of the control-non-gamers, which decreased from 3.5% control to 2.7% experimental, W=9035, p<0.01.

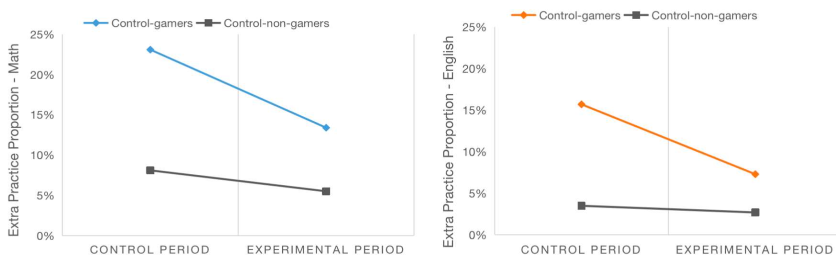

Fig. 4. The proportion of extra practice (more than 3 problem sets) during the control and experimental periods, for Math (Left) and English (Right)

Average Time Spent Per Item. Starting from the second practice set, there was a statistically significant increase in the average time spent on each item in the experimental period compared to the time spent in the control period, especially in the second practice set, as shown in Figure 5. In math learning, the average time spent per item in the first practice set decreased from 98.87s in the control period to 91.06s in the experimental period, which is significantly different, t(342)=3.32, p = 0.001 for a paired t-test. However, we observed that 340 students out of 343 completed at least two practice sets for each concept they encountered, in both the control and experimental periods. In the second practice set, the average time a student spent answering each math item increased from 79.66s in the control period to 116.90s in the experimental period, which is statistically significant, t(339)=12.75, p<0.01. We also observed 217 students out of 343 completed at least 3 practice sets for each concept they encountered, in both the control and experimental periods. In the third practice set, the average time a student spent on each math item increased from 85.05s in the control period to 97.41s in the experimental period, which is statistically significant, t(216)=2.98, p<0.01.

In English, the average time spent per item in the first practice set decreased statistically significantly from 40.87s in the control period to 37.75s in the experimental period, t(344)=2.84, p = 0.004 for a paired t-test. However, we observed that 338 students out of 345 completed at least two practice sets for each concept in both the control and experimental periods. In the second practice set, the average time a student spent on each item increased from 29.65s in the control period to 52.91s in the experimental period, which is statistically significant, t(337)=13.83, $p<0.01$. We also observed 204 students out of 343 completed at least 3 practice sets for each concept they encountered, in both the control and experimental periods. In the third practice set, the average time a student spent on each English item increased from 30.35s in the control period to 38.47s in the experimental period, which is statistically significant, t(203)=3.46, $p<0.01$.

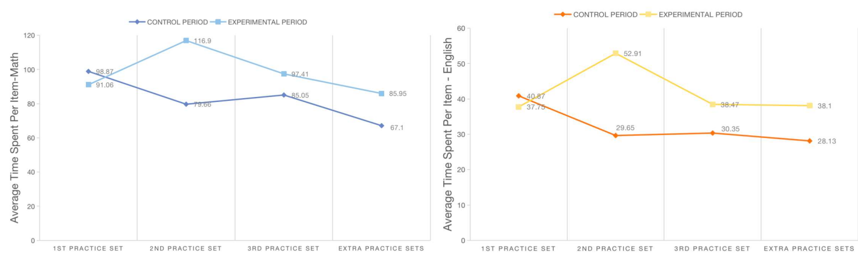

Fig. 5. The average time spent per item across practice sets, in the control and experimental periods, in Math (Left) and English (Right)

As shown in Figure 6, starting from the second practice set, students spent more time per item during the experimental period than during the control period for both subjects and groups, and students in the control-gamers on average spent less time per item than the control-non-gamers. In math, a repeated measures ANOVA indicates that the average time spent on each item is higher in the experimental time period than in the control time period, F(1,338)=120.7, $p<0.01$. Average time spent on each item is also generally higher (across both time periods) for the control-non-gamers than the control-gamers, F(1,338)=20.3, $p<0.01$. However, the interaction between time period and whether students gamed during the control period is not quite significant, F(1,338)=1.2, p=0.28. The lack of a clear interaction suggests that time increases occurred even for students who never gamed, and that there is not clear evidence that this intervention impacted the behavior of gaming students differently than non-gaming students.

For English, the average time spent per item is also significantly higher in the experimental time period than the control time period, F(1,336)=141.2, $p<0.01$. The average time spent on each item is also generally higher (across both time periods) for the control-non-gamers than the control-gamers, F(1,336)=6.3, p=0.01. However, the interaction between time period and whether students gamed during the control period is not significant, F(1,336)=0.05, p=0.82. Again, the lack of an interaction suggests that time increases occurred even for students who never gamed, and that this intervention did not impact the behavior of gaming students differently than non-gaming students.

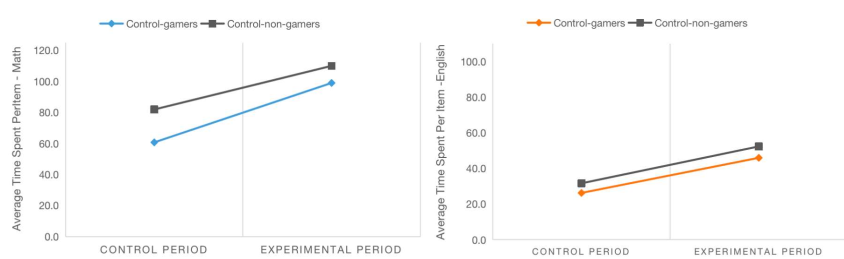

Fig. 6. The average time spent per item, in the control and experimental periods, for students who did/did not game in the control period, in Math (Left) and English (Right)

Average Percent Correctness. The average percent correctness was higher in the experimental period than in the control period, as shown in Figure 7. For math learning in the first practice set, the average percent correctness increased statistically significantly from 60.3% in the control period to 66.0% in the experimental period, t(342)=7.50, p<0.01 for a paired test. In the second practice set, 340 students out of 343 completed at least two practice sets for each concept. The average percent correctness increased statistically significantly from 59.3% in the control period to 66.6% in the experimental period, t(339)=7.07, p<0.01 for a paired test. In the third practice set, 217 students out of 343 completed at least two practice sets for each concept. The average percent correctness increased statistically significantly from 56.3% in the control period to 64.9% in the experimental period, t(216)=4.93, p<0.01 for a paired test.

For English learning, the average percent correctness in the first practice set increased from 68.1% in the control period to 73.5% in the experimental period, which is statistically significant, t(344)=9.03, p<0.01 for a paired test. In the second practice set, 338 students out of 345 completed at least two practice sets for each concept. The average percent correctness increased from 66.4% in the control period to 72.3% in the experimental period, which is statistically significant, t(337)=6.67, p<0.01 for a paired test. In the third practice set, 204 students out of 345 completed at least two practice sets for each concept. The average percent correctness increased from 63.8% in the control period to 69.6% in the experimental period, which is statistically significant, t(203)=3.40, p<0.01 for a paired test.

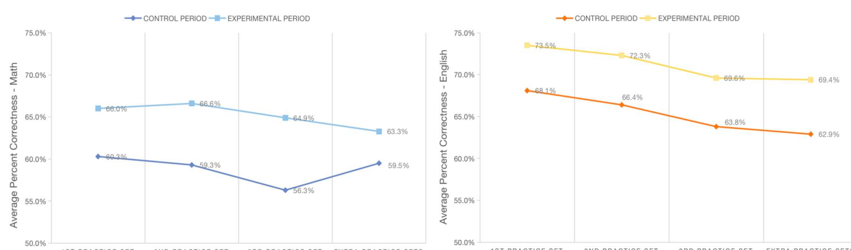

Fig. 7. The average percent correctness across practice sets, in the control and experimental periods, in Math (Left) and English (Right)

In math, using repeated measures ANOVA, we find that the average percent correctness for each math practice is significantly higher in the experimental period than the control period, F(1,341)=90.3, p<0.01. The average percent correctness was also significantly higher for the control-non-gamers than the control-gamers, F(1,341)=26.2, p<0.01, as shown in Figure 8. The interaction between time period and whether students gamed the system was statistically significantly associated with percent correctness, F(1,341)=6.6, p=0.011, suggesting that the intervention led to a larger gain in accuracy for the control-gamers than the control-non-gamers.

In English, according to a repeated measures ANOVA, the average percent correctness was statistically significantly higher in the experimental period than the control period, F(1,343)=112.6, p<0.01. Average percent correctness was also statistically significantly higher for the control-non-gamers than the control-gamers, F(1,343)=42.2, p<0.01. The interaction between time period and whether students gamed the system was statistically significantly associated with percent correctness, F(1,343)=7.6, p=0.006, suggesting that the intervention impacted the students differently, leading to a larger gain in accuracy for the control-gamers.

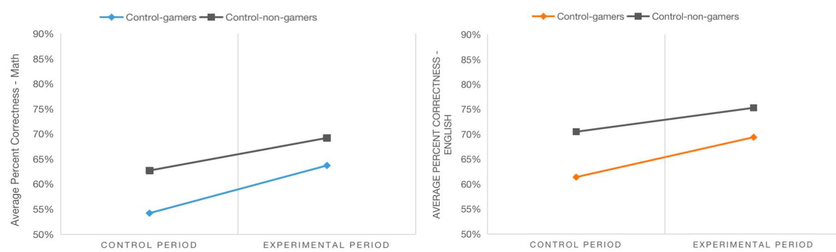

Fig. 8. The average percent correctness, in the control and experimental periods, for students who did/did not game in the control period, in Math (Left) and English (Right)

4 Discussion and Conclusions

Gaming the system is an ongoing problem for many adaptive learning systems and intelligent tutoring systems, leading to poorer learning [1, 21, 4]. Although it was originally thought that gaming the system mainly occurred in systems that gave students multiple chances to solve the same item or with hints that give the answer (e.g. [1]), students can be highly creative in finding ways to obtain answers without learning the material (such as in [13]). In this paper, we document two ways that students game the Kupei learning platform, by exhausting the practice set to get repeated problems and by quitting practice sets and immediately entering the correct answers they just saw.

One possible approach to reducing gaming the system would simply have been to remove the history of past student work. However, many students review this history in constructive ways, including in review sessions with instructors where they discuss their work. In addition, removing the history of past student work would only have impacted one of the two fashions in which students game, likely driving students to use the other gaming strategy (as in [13]). Students could also write down the answers as they completed problems, rather than reviewing the history.

Therefore, we instead sought to address these gaming behaviors with a multi-pronged gaming prevention intervention consisting of 1) imposing a pause of 36 hours on a practice set once potential gaming behaviors (completing three practice sets on the same concept in a single day) are detected; 2) providing meta-cognitive feedback to remind students about the cost of gaming after the student has performed poorly in two consecutive practice sets on the same concept; and 3) offering students another way to learn the materials missed by gaming before doing the last practice set in a set of 3.

In this study, we hypothesized that the combined gaming prevention intervention would help mitigate the effect of gaming and encourage students to become more engaged in the content. These three interventions appear to have been successful at dissuading students from gaming the system. Supporting our hypothesis, we detected a lower frequency of gaming behaviors in both English and math after the integration of the gaming intervention. In addition, we found that fewer students used extra practice (more than three practice sets) on a concept after the implementation of the gaming intervention. Instead, students spent more time on later items in both English and math during the experimental period, possibly indicating students are practicing each item more seriously than students in the control period. In addition, we observed an increase in the percent correctness on the second practice set in both English and math.

However, there appear to be some limitations to this approach that should be considered in future work. The intervention was successful at reducing the second type of gaming — students' quitting the problem set to seek answers from problems history and improving performance within the system. However, this intervention did not appear to be as successful at addressing the first type of gaming behavior—students' exhaustively completing problem sets to obtain answers until questions are re-shown. In order to prevent this type of gaming behavior, new interventions may need to be developed in the future to -- for instance -- creating more content so that it is not practical to game until content repeats (particularly given 36-hour delays), notifying a teacher that a student is close to exhausting the content, or not displaying the history of repeated content. Ultimately, addressing this form of gaming behavior may benefit from working with teachers so that teachers identify this behavior and respond to it.

Another possible limitation is that even if some students reduce their frequency of gaming the system, they may not replace gaming with the most desirable behaviors. For example, students are required to watch the video on the content they missed once gaming behaviors are detected. The required integrated review provided students extra opportunities to review the content. However, students may have failed to complete the video once the required watch time has been reached or may refuse to watch it at all (by letting it play while engaging in off-task behavior, watching something else or doing something else). A final possible limitation is that the current interventions may result in students feeling like they are being forced to watch the video and read the notes, possibly creating broader negative feelings towards the material as well.

In this article, we have studied the potential effectiveness of a combined gaming prevention intervention designed to help mitigate student gaming behaviors in the Kupei learning system. Ultimately, we hope that our research will inform the design of systems that will reduce students' motivation to game the system, and, in turn, increase the frequency of effective self-regulated learning strategies that lead to better student learning -- hopefully reducing the longer-term impacts that gaming the system appears to have on student outcomes [7].

References

1. Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom. Proceedings of the 2004 Conference on Human Factors in Computing Systems - CHI '04 (pp. 383-390).
2. Aleven, V., McLaren, B.M., Roll, I., & Koedinger, K. (2006). Toward Meta-cognitive Tutoring: A Model of Help Seeking with a Cognitive Tutor. International Journal of Artificial Intelligence in Education. 16. 101-128.
3. Rowe. J., McQuiggan, S., Robison, & J., Lester, J. (2009). Off-Task Behavior in Narrative-Centered Learning Environments. Proceedings of the Int'l Conf on AI in Education. 99-106.

4. Pardos, Z.A., Baker, R.S., San Pedro, M.O.C.Z., Gowda, S.M., & Gowda, S.M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end of year learning outcomes. Journal of Learning Analytics, 1(1), 107-128.
5. Fancsali, S. (2014). Causal Discovery with Models: Behavior, Affect, and learning in Cognitive Tutor Algebra. In Proc. of the Int'l Conf on Educational Data Mining (pp. 198-211)
6. San Pedro, M.O.Z., Baker, R.S.J.d., Bowers, A.J., Heffernan, N.T. (2013) Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. In Proceedings of the 6th Int'l Conference on Educational Data Mining, 177-184.
7. Almeda, M. V. Q., & Baker, R. S. (2020). Predicting Student Participation in STEM Careers: The Role of Affect and Engagement during Middle School. Journal of Educational Data Mining, 12(2), 33–47.
8. Baker, R.S., Corbett, A.T., Koedinger, K.R., Evenson, S., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J.E. (2006). Adapting to when students game an intelligent tutoring system. In Proc. of the Int'l Conf on Intelligent Tutoring Systems (pp. 392-401).
9. Roll, I., Aleven, V., Mclaren, B., Koedinger, K.. (2007). Can Help Seeking Be Tutored? Searching for the Secret Sauce of Metacognitive Tutoring. In Proc of the Int'l Conf on Artificial Intelligence in Education, 203-210.
10. Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Mehranian, H., Fisher, D., Barto, A., Mahadevan, S., Woolf, B. (2007). Repairing Disengagement With Non-Invasive Interventions. In Proc of the Int'l Conf on Artificial Intelligence in Education, 195-202.
11. Xia, M., Asano, Y., Williams, J. J., Qu, H., & Ma, X. (2020). Using information visualization to promote students' reflection on "gaming the system" in online learning. In Proceedings of the Seventh ACM Conference on Learning @ Scale.
12. Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. Learning and Instruction, 21(2), 267–280.
13. Murray, R. C. & Vanlehn, Kurt. (2005). Effects of Dissuading Unnecessary Help Requests While Providing Proactive Help. In Proc of the Int'l Conf on Artificial Intelligence in Education, 887-889.
14. Beck, J.E. (2005). Engagement tracing: using response times to model student disengagement. In Proc of the Int'l Conf on Artificial Intelligence in Education (pp. 88-95).
15. Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2004). Toward tutoring help seeking. Proc. of the Int'l Conf. on Intelligent Tutoring Systems, 227–239.
16. Zou, X., Ma, W., Ma, Z., Baker, R. (2019) Towards Helping Teachers Select Optimal Content for Students. *Proceedings of the 20th International Conference on Artificial Intelligence and Education.*
17. Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction, 4(4), 253-278.
18. Heffernan, N. T., & Heffernan, C. L. (2014). The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. International Journal of Artificial Intelligence in Education, *24*(4), 470–497. https://doi.org/10.1007/s40593-014-0024-x
19. Beck, J & Gong, Y. (2013). Wheel-Spinning: Students Who Fail to Master a Skill. In Proc. of the Int'l Conf on Artificial Intelligence in Education (pp. 431-440).
20. Kai, S., Heffernan, C., Almeda, M,V,. Heffernan, N,. & Baker, R.S. (2018). Decision tree modeling of wheel-spinning and productive persistence in Skill Builders. Journal of Educational Data Mining, 10 (1), 36-71.
21. Beck, J., & Rodrigo, M.M.T. (2014). Understanding wheel spinning in the context of affective factors. In Proc. of the 14th Int'l Conf on Intelligent Tutoring Systems (pp. 162-167).