Iterative Refinement of an AIS Rewards System

Karen Wang[1] and Zhenjun Ma[2] and Ryan S. Baker[3][0000-0002-3051-3232] and Yuanyuan Li[4]

[1] Worcester Polytechnic Institute, 100 Institute Rd, Worcester MA 01609, USA
[2] Learnta, 1460 Broadway, New York New York 10036, USA
[3] University of Pennsylvania, 3700 Walnut St., Philadelphia PA 19104, USA
[4] Learnta, 1460 Broadway, New York New York 10036, USA

Abstract. Gamification-based reward systems are a key part of the design of modern adaptive instructional systems and can have substantial impacts on learner choices and engagement. In this paper, we discuss our efforts to engineer the rewards system of Kupei AI, an adaptive instructional system used by elementary and middle school students in afterschool programs to study English and Mathematics. Kupei AI's rewards system was iteratively engineered across four versions to improve student engagement and increase progress, involving changes to how many points were awarded for success in different activities. This paper discusses the design changes and their impacts, reviewing the impacts (both positive and negative) of each generation of re-design. The end result of the design was improved learning and more progress for students. We conclude with a discussion of the implications of these findings for the design of gamification for adaptive instructional systems.

Keywords: Adaptive Instructional System, Gamification, Reward System.

1 Introduction

Gamification-based reward systems are an increasingly common part of the design of modern adaptive instructional systems and can have substantial impacts on learner choices and engagement [15]. For example, GamiCAD introduced arcade-style bonus missions based on successful performance, leading to faster completion of content and more completion of content, as well as self-report of better engagement [10]. The gold bars given when students master a concept in a Cognitive Tutor have also been found to be a positive incentive for many students [11]. Some research has suggested that not all students find reward systems compelling, but those who do tend to see increased motivation and learning [16]. Much of the work using rewards for gamification builds on an intellectual tradition of behavior management and modification that dates back to early behaviorist work. In that work, reward systems were studied for their impact on behavior and learning [17]. Morford and colleagues [13] note the intellectual contribution that the applied behavior analysis and behavior modification literatures have made to contemporary work on gamification.

However, as been noted since the 1970s, reward systems can impact students in negative fashions as well as positive fashions [9]. Reward systems, if incorrectly

designed, can focus learners on short-term outcomes rather than long-term outcomes and lead students to adopt behaviors that maximize the reward rather than the learning it is intended to promote. Furthermore, not all reward systems are even effective at promoting the intended behaviors [12] – the details of the design appear to matter considerably. Reward systems, if poorly designed, can also reduce long-term intrinsic motivation for the learning activity [3]. As such, considerable attention needs to be paid to the design of reward systems in adaptive instructional systems. Design principles for gamification [6] can be useful but, as discussed above, even a seemingly excellent design can have unintended consequences [5,9]. As such, the methods of learning engineering [5] are needed to ensure that designs achieve their intended goals. In particular, ongoing monitored iterative design [8] is needed, where the developer repeatedly modifies the system and tests the consequences of those modifications.

In this paper, we present the monitored iterative design of a gamified rewards system for the Kupei AI adaptive instructional system, an AI-driven learning system tutoring elementary (3rd-6th graders), middle (7th-9th), and high (9th-12th) schoolers in English and Math in China. Across four iterations, a point system was implemented in order to encourage students to master concepts and improve their learning progress. The point system was intended to give students an incentive to work on specific concepts until they reached mastery, using the system's various features appropriately and efficiently. However, students developed strategies targeted towards earning points with the maximum efficiency that were not optimal for learning. For instance, certain design choices led students to put more effort into curricular areas where it was easier to rapidly master topics. Through multiple design iterations, we were able to design a system that guided students towards more appropriate system usage.

In this paper, we present the story of the iterative design of this system, presenting each iteration in design, and empirically investigating its results on student usage behaviors and learning outcomes. In presenting this narrative, we consider broader themes around how to design effective gamification systems, distilling what we learned at each phase of redesign.

2 Methods

2.1 Platform

Kupei AI is an AI-driven learning system tutoring elementary (3rd-6th graders), middle (7th-9th), and high (9th-12th) schoolers in English and Math in China. Courses are divided up by grade level and then by units then concepts. The units are correlated with what students are learning in school. Each unit comprises a list of concepts, with the subjects the learning system recommends for the student at the top (Figure 1).

The intended design is for students to begin a unit with a diagnostic test, a short quiz, although they have the choice to directly go into practice (Figure 1). If students decided to skip the test and go straight to practice, then a message window popped up that encouraged the student to take a diagnostic test first. The diagnostic test covers multiple concepts to determine which concepts the student should work on

within the unit. Since concepts are connected through a knowledge graph, if a student achieves advanced mastery on a concept, then the learning system infers that the student has also mastered the concept's prerequisites. Following the diagnostic test, students have the option to practice. Practice is divided into problem sets of 5 items on the same concept -- however, if a student achieves advanced mastery (defined below) after the first 3 items in a problem set, the system stops the student's work on the problem set automatically.

Once finished with either the practice or test, students are met with a concluding screen showing concepts where they achieved basic mastery and those that were not mastered (Figure 2). The learning system categorizes student proficiency on concepts students have worked on into three levels: unmastered, basic mastery, and advanced mastery. Using Bayesian Knowledge Tracing [3], the system estimates student proficiency in real-time. If a student has under an 80% probability of knowing a skill, the concept is unmastered; if the probability is between 80% (a cut-off used by many commercial systems) and 95% (the original cut-off in [3]), then the concept is labeled as basic mastered; if the concept mastery level is above 95% then it is labeled as advanced mastery. By design, advanced mastery can only be earned through practice while basic mastery can be earned through both practice and tests.

The original system (referred to below as version V0) did not incorporate gamification, but starting April 29th, 2021, the Kupei AI team released and began the process of iterating a point system. In its first version (referred to below as version V1), the point system gave students one point for each basic mastered concept. Students were able to click on a little treasure chest next to a summary of the specific learning concept they mastered to receive the point (Figure 3). Starting with V1, after completing a practice or test, students can click open treasure chests next to each concept they gained basic mastery on and earn a point. Students were also ranked against their classmates by the number of points they have, and this was shown in a leaderboard available to students to view after finishing a test or practice set.

Fig. 1. Student unit learning screen. The large orange button says “test” and the red button at the bottom right says “practice.” In this unit, there are 15 concepts to master. The progress bar next to the practice button shows mastery progress for each concept.

Fig. 2. The results screen after a diagnostic test. The left circle shows the student received a score of 15/100. The right circle shows the number of concepts mastered in the unit, which is 2/15. The bottom row of numbers from 1 to 7 are the questions the student answered in the diagnostic test: red means they answered the question incorrectly and green means they answered correctly.

Fig. 3. Students scroll to the questions they answered correctly and click on the treasure chest next to each learning concept to receive their points.

2.2 Data Collection

The entire student user base of Kupei, composed of 5726 students, was tracked from April 22nd to June 17th. From April 22 to 28, the students used the baseline version without the point system (V0). From April 29 to May 8, the students used the first version of the point system (V1). On May 9 and June 1, respectively, iterated versions V2 and V3 of the point system were introduced.

Across these iterations, the students worked on 27785 concepts, making a total of 1,393,529 actions. During this time, students achieved basic mastery on 29.7% of the concepts they attempted and achieved advanced mastery on 41.2% of the concepts they attempted. Although the system notifies the student when they achieve basic mastery, students generally did not stop at basic mastery, and went on to advanced mastery, leading to a higher proportion of advanced mastery.

2.3 Overview of Design Iterations

When the first version V1 of the point system first launched on April 29th, its primary purpose was to make learning more engaging by giving students a reward for earning basic mastery on a concept. By giving a point for each concept mastered, and showing the class's scores on a leaderboard, the intention was to incentivize students to compete [14] to earn a large number of points to surpass their classmates on the leaderboard. However, though the student's goal would be to perform well and compete, they would be learning more concepts and making more progress within the system to do so.

Before the point system, Version 0 (V0), students were unable to earn points. The first version of the point system (V1) gave the student a point the first time they earned basic mastery on a specific learning concept, and introduced the leaderboard. Version V1 gave a single point for mastering any concept, regardless of whether the concept was in English or Math.

As will be discussed below, Math concepts generally take longer to complete in the system than English concepts. As a result, students who wanted to be at the top of the leaderboard shifted to working mostly on concepts in English. Version V2 attempted to correct for this unanticipated incentive by making basic mastery in English worth one point and mastery in Math worth three points. This change re-balances work on the two subject areas.

Having addressed this limitation, the design team noticed that students were now much more likely to skip the diagnostic test. The diagnostic test is designed to determine which concepts in the unit the student needs to work on, avoiding allocating student time to concepts they already know. However, by skipping the test, students can master concepts they already know, gaining points without learning. Version V3, therefore, made it possible to earn three points for achieving Math basic mastery and one point for achieving English basic mastery in practice and on tests.

In the following section, we go into greater detail on the impact of each of these design changes on student behavior, learning, and learning efficiency.

3 Analysis and Results

3.1 V0 to V1 Design Change Impact

V0 represented the first version of the system, before the introduction of the point system. We added a point system and leaderboard to V0, creating V1. In V1, students could earn points when they achieved basic mastery of concepts while practicing. English and Math were both worth one point for each basic mastered concept.

First, we can compare students between V0 and V1 in terms of their mastery rates. We consider the basic mastery rate, or the number of concepts a student achieved basic mastery on divided by the total number of concepts they worked on. We also consider advanced mastery rate, the number of concepts a student achieved advanced mastery on, divided by the total number of concepts they worked on. In doing so, we only look at students who used the system during both the V0 and V1 periods and conduct a paired comparison, as only a very small number of students quit or started the system right at the switchover date.

In V0, students had a basic mastery rate of 62.5,% skewness = -0.563. In V1, students had a basic mastery rate of 64.1%, a statistically significantly faster basic mastery rate, $t(3683)=-3.3367$, $p<0.001$, skewness = -0.654. In V0, students had an advanced mastery rate of 43.1%. In V1, students had an advanced mastery rate of 45.4%, a statistically significantly faster mastery rate, $t(3368)=-4.2585$, $p<0.001$, skewness = 0.229 V0, 0.080 V1. All values of skewness for mastery rate were in the moderately skewed or approximately symmetric ranges, justifying the use of parametric tests.

From V0 to V1, the time students took to achieve basic mastery on each concept decreased by a median of 35.7 seconds in English, $V = 498656$, $p<0.001$, and went up by a median of 7.3 seconds which is statistically significant, $V = 1283079$, $p<0.001$, in Math.

However, the proportion of time spent on English versus Mathematics shifted between the two versions. In V0, 47.403% of students' completed concepts were English and 52.6% were Math. This changed considerably in V1, where students spent 59.5% of their completed concepts were English, and 40.5% of their completed concepts were Math. The increase in the proportion of time spent on English from V0 to V1 was statistically significant, $t(779)=-10.614$, $p<0.001$, skewness = 0.0692 V0, -0.441 V1. The decrease in the proportion of time spent on Math from V0 to V1 was also statistically significant, $t(779)=10.614$, $p<0.001$.

This difference in time spent led to a difference in the rate of basic mastery between the two content domains. The basic and advanced rates in each subject increased from V0 to V1. In V0, students' median English basic mastery rate was 75%. In V1, it was 74%, not a statistically significant difference (Wilcoxon used due to high skewness), $V = 285187$, p-value = 0.1878, skewness = -0.911 V0, -0.945 V1Math basic mastery rate in V1 went up as well. Students had a basic mastery rate of 59.4% in V0 which went up to 60.8% in V1, a statistically significant increase, $t(2208) = 2.28$, $p = 0.023$, skewness = -0.411 V0, -0.511V1. Additionally, the advanced mastery rate in English statistically significantly increased from 49.7% in V0 to 52.6% in V1,

t(1159) = -3.2294, p=0.001, skewness = -0.091 V0, -0.222 V1. The advanced mastery in math increased from 39.6% in V0 to 41.6% in V1, t(2208) = -2.9198, p = 0.004, skewness = -0.411 V0, -0.511 V1. Both mastery rate increases were significant.

Another shift found was that students were much more likely to skip the diagnostic tests in V1 than in V0. In V0, 49.1% of student items were on practice and 50.9% of student items were on diagnostic tests. In V1, 64.0% of student items were on practice and 36.0% of student items were on diagnostic tests. Students were statistically significantly more likely to practice in V1 than V0, by 14.9%, t(871)=16.15, p-value < 0.001, skewness = 0.0279 V0, -0.507V1.

3.2 V1 to V2 Design Change

To try to re-balance the proportion of time spent in English and Math, we changed the points given for mastering a concept in Math from 1 point to 3 points (version V2). As Figure 6 shows, this led to a gradual shift back to Math. Students completed more Math concepts in V2 than in V1, leading to the overall proportion of concepts completed being approximately equal for English and Math in V2. The proportion of English completed by students in V1 was 55.3% and the proportion of math completed was 44.73878%. In V2, English concepts completed significantly decreased to 50.7%, t(1414)=5.92, p<0.001 and Math significantly increased to 49.3%, t(1414) = -5.92, p<0.001, skewness = -0.250 V1, -0.177 V2.

The overall proportion of practice continued to significantly increase, from 56.5% in V1 to 66.5% in V2, t(1552)=15.02, p<0.001, skewness = -0.218 V1, -0.694 V2, and the proportion of tests significantly decreased from 43.5% to 33.5% in V2, t(1552)=-15.02, p<0.001, skewness = 0.218 V1, 0.694 V2. As students could only earn points in practice, the increase in the proportion of practice is seen in both English and Math, with the change in Math being especially large. The proportion of English practice completed significantly increased from 66.7% in V1 to 72.7% in V2, t(730)=7.58, p< 0.001, skewness = -0.541V1, -0.767 V2. In addition, Math practice significantly increased from 45.0% in V1 to 59.6% in V2, t(923)=16.60, p<0.001, skewness = 0.180 V1, -0.439 V2.

However, since students completed more practice and each practice set targets one learning concept at a time, this led to a statistically significant increase in the time taken to master both Math and English concepts. In Math, the median time taken to master a concept increased from 734.67 seconds in V1 to 813.97 seconds in V2, V = 1165568, p<0.001. In English, the median time taken to master a concept increased from a median of 347.45 seconds in V1 to 377.6437 seconds in V2, also statistically significant, V = 328174, p-value < 0.001.

Despite changes in student learning behavior, the only significant change for both basic and advanced mastery rate was a decrease in the English advanced mastery rate. Math basic mastery rate was 60.0% in V1 and 60.4% in V2, which was not a statistically significant difference, t(3141) = 0.93, p = 0.350, skewness = -0.460 V1, -0.460 V2. Although, the median English basic mastery rate was 71% in V1 and 73% in V2, was a statistically significant increase, V = 892435, p-value = 0.001628, skewness = -0.907 V1, -0.824 V2. Advanced mastery for Math was 37.98% in V1 and

37.99% in V2, t(3141) = -0.021366, p = 0.983, skewness = 0.411 V1, 0.399 V2, but significantly decreased for English from 49.5% in V1 to 47.7% in V2, t(1903) = 2.89, p = 0.004, skewness = -0.151 V1, -0.011 V2.

3.3 V2 to V3 Design Change

The increase in practice in V1 and V2 implied that students were no longer using the diagnostic test as often to focus their time on the skills they needed to learn. We therefore re-designed V3 to give students a point for demonstrating mastery of a concept within the diagnostic test. In V2, students completed the majority of the concepts they encountered through practice, with 64.6% of concepts completed being practice and 35.4% being test. After V3 launched, this ratio switched, with 46.4% of concepts completed in practice and 53.6% in tests. Overall, the proportion of completed concepts in practice significantly decreased from 64.6% in V2 to 46.4% in V3, t(2489)=30.84, p<0.001, skewness = -0.521 V2, 0.157 V3, from V2 to V3, and tests significantly increased from 35.4% in V2 to 53.6% in V3, t(2489)= -30.84, p<0.001. As shown in Figures 4 and 5, this applied to both English and Math concepts. Math testing increased from 42.2% in V2 to 58.9% in V3, t(1712)=24.02, p<0.001, skewness = 0.302 V2, -0.329 V3 compared to Math practice. English testing increased from 27.4% in V2 to 46.5% in V3, t(1167)=23.44, p<0.001, skewness = -0.818 V2, -0.142 V3, compared to English practice. As shown in Figure 7, as students completed more diagnostic tests in V3, they were also able to achieve basic mastery on concepts faster. For English, the median time spent to achieve basic mastery decreased by 12.4% from 372.4 seconds in V2 to 326.2 seconds in V3, V = 2184932, p<0.001, and in Math, the median time spent decreased by 27.5% from 773.2 seconds in V2 to 560.3 seconds in V3, V = 6007553, p<0.001.

Although basic mastery rate did not significantly change for Math from V2 to V3, it did significantly decrease from 67.8% in V2 to 64.7% in V3 for English, t(2317) = 12.674, p<0.001, skewness = -0.812 V2, -0.721 V3. In addition, advanced mastery significantly decreased for both subjects with a decrease of 2.6% in Math, V = 3733852, p<0.001, and 7.1% in English, t(2317) = 12.67, p<0.001, skewness = 0.035 V2, 0.391 V3.

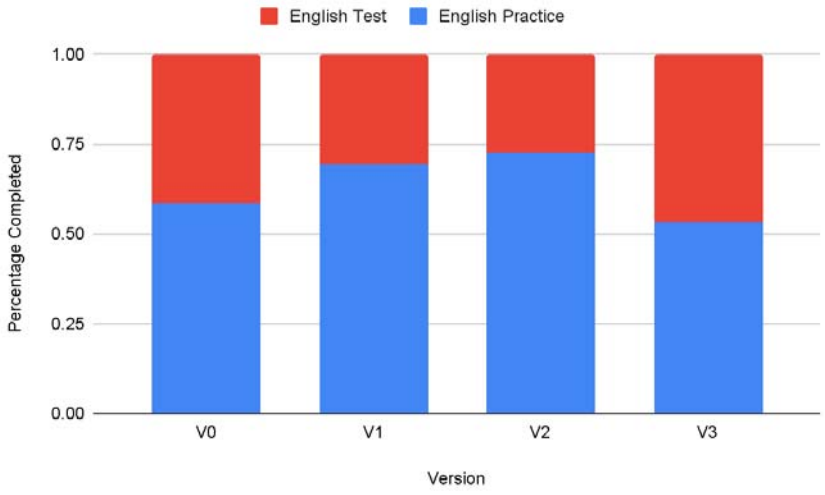

Fig. 4. Student completed English concept proportion test and practice per version.

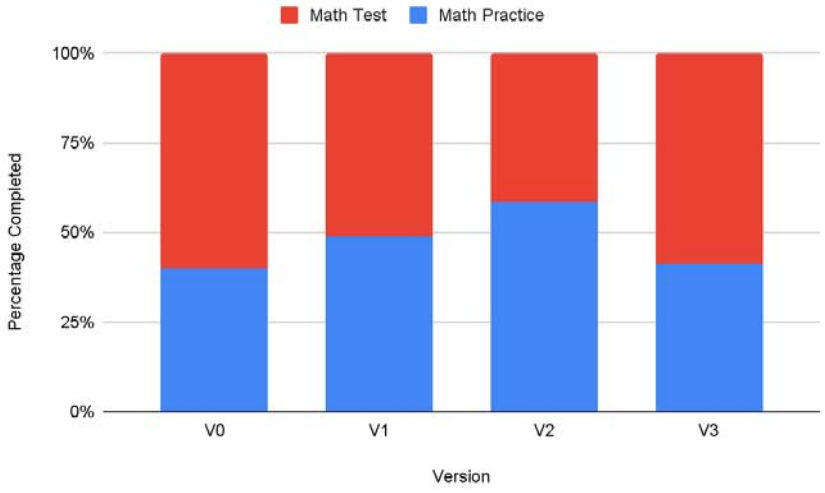

Fig. 5. Student completed Math concept proportion test and practice per version.

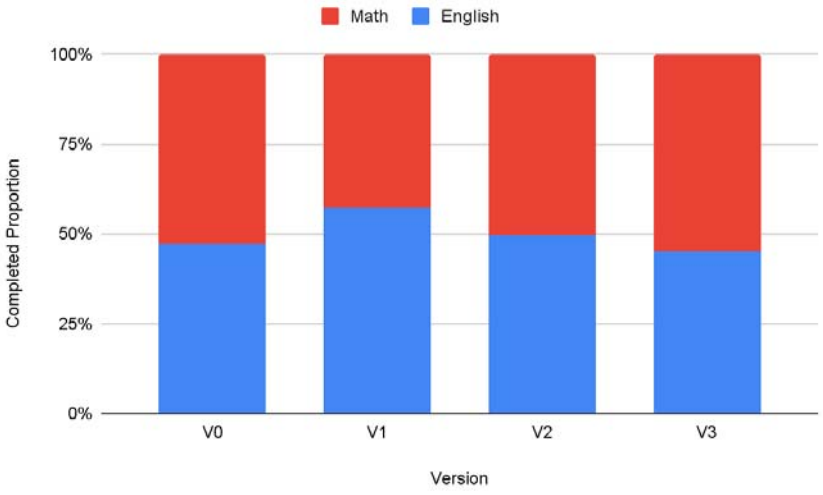

Fig. 6. Student completed concepts proportions of Math and English.

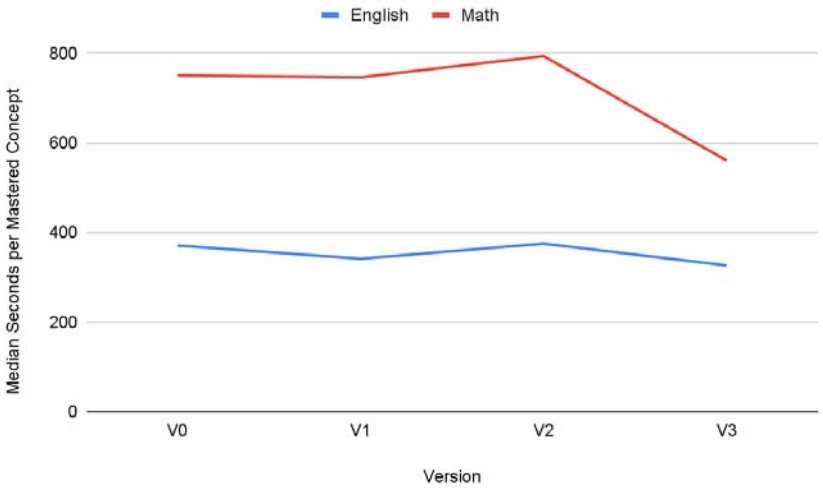

Fig. 7. Median seconds taken to achieve basic mastery on one concept per version.

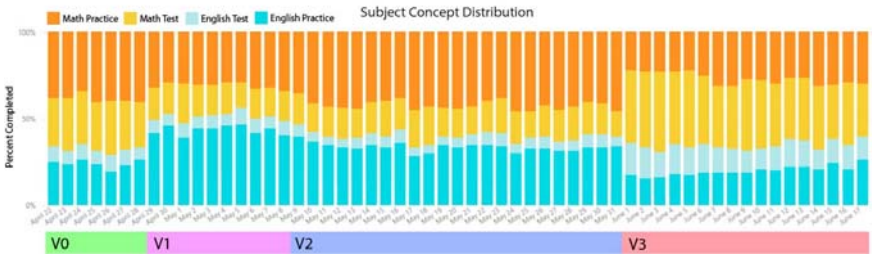

Fig. 8. Student proportion of completed concepts in Math practice, Math test, English test, and English practice by day.

4 Discussion and Conclusions

In this paper, we add gamification to an adaptive instruction system, Kupei AI, in the form of a point system and leaderboard. This paper details the iterative improvement of that point system through three versions, V1, V2, and V3. By adding a rewards system to this AIS, it was possible to shift student behavior, although there were some unexpected consequences in some of the earlier design iterations.

The first modified version, V1, added a leaderboard and gave students a single point for achieving basic mastery in both Mathematics and English. From V0 to V1, learning efficiency improved, especially for English. Not only did they spend less time to master a learning concept for English, decreasing by 35.7 seconds, but their mastery rate, which is their total number of mastered concepts over their total completed concepts, increased for both English and Math by 2.1 and 1.4 percentage points respectively. In addition, student learning behavior changed from V0 to V1 as students focused on completing more English and more practice. Students were incentivized to complete tasks that earned them points quickly. So, since students could only earn points in practice, not test, they completed an increase in practice. Also, English took less time to master English than Math so students mastered more English concepts than Math. Despite students being more motivated to master concepts quickly to earn points, they focused on earning points the quicker way through English practice rather than through Math.

After seeing students switch to working much more on English than Math, V2 was implemented to balance the number of concepts students spent learning. Since it was easier to earn points quickly in English, V2 balanced things by giving students more points per concept mastered in Math. As such, the proportion of English worked on reduced, bringing English and Math back into greater balance. This change affected student mastery performance, with English basic mastery rate increasing by 2 percentage points. However, students spent more time to master a concept. This occurred because students completed an even greater proportion of practice compared to tests, meaning that students were practicing concepts in full that they would have bypassed through taking the diagnostic test. If students mastered a concept in the

diagnostic test, they would not earn points for it, so they skipped straight to practice to earn points. This shift from test to practice between V1 and V2 may have been a continued trend from V0 to V1, or it may have been that the greater number of points available was a stronger inducement to complete practice rather than tests.

Doing only practice is not the most effective way to learn in Kupei AI, as the diagnostic tests help students focus on concepts they need to improve rather than going through all concepts. This is especially important for Math. To address this issue and encourage students to complete the diagnostic tests, V3 made it possible for students to earn mastery points through demonstrating mastery within tests. Quickly, students began to complete more tests and less practice, as the proportion of practice completed compared to tests decreased by 18.2 percentage points. Testing increased considerably in both subjects. As a result, the time spent mastering concepts decreased significantly, with English decreasing 46 seconds and Math decreasing 213 seconds. However. The amount of Basic mastery went down in English and advanced mastery decreased in both Math and English, by 2.6 and 7.1 percentage points respectively. The decrease in English basic mastery rate was likely due to students spending more time on Math diagnostic tests. The decrease in advanced mastery was most likely because students can only achieve advanced mastery through practice, not tests. Since achieving advanced mastery did not reward students with any points, once a student mastered a concept through tests, there was no incentive for them to go back and practice that concept further to earn advanced mastery.

Although there has been considerable design progress from V0 to V3, more progress remains possible. In a sense, the design of gamification and rewards systems can be an extended journey, as each change has a mixture of positive and negative (typically unintended) consequences. In a future version of a system like this nature, it would probably make sense to also introduce rewards for achieving advanced mastery. Even if this results in some over-practice, there is likely some advantage to this in terms of preparation for learning future concepts [7] and retention of knowledge over time [1]. Further changes will become necessary, as each change shifts the overall pattern of student behavior.

In reflecting not just on the implications of our design but also on the implications of our approach to design, we can see that there are several advantages (and of course some limitations) to the approach we have developed. Switching rapidly between designs enables rapid iteration towards a better design, but it also means that the impacts of changes may bleed together. Longer periods of each version would make differences between versions clearer but requires more time to hone in on a better system. In this analysis, we switched all students to the next version at the same time and compared time periods. An alternate (more common) approach would have been to conduct A/B tests. However, there would be significant possible issues if we randomly assigned at the student level (i.e. students would have different reward structures in the same learning center, creating effects such as resentful demoralization [2]). Therefore, we would need to assign students to condition by learning center, creating challenges in the comparability of students between conditions. One of the largest reasons to use an A/B test rather than compare across time periods is the possibility that previous versions impact current behavior – however, Figure 8 shows

the relatively rapid response of students to design changes. In general, the need to iterate design based on the effects of the current intervention makes it less feasible to conduct many-condition A/B/n studies, slowing the degree to which assignment of students to condition could speed design progress in this case. Overall, rapid iteration of the type used in this study has some limitations, as students who use multiple versions may respond to later versions differently than students who start with those later versions. This risk reduces the degree to which we can infer causality and generalizability for our changes, but the result is a system that produces better results. Clear and unquestionable evidence on what caused the impact can only be produced through a randomized comparison on entirely new students, but is that always the primary goal? Ultimately, the effort needed for an unconfounded randomized study may be better spent on an efficacy study comparing to a control condition.

Further research may also want to look for the impact of these design iterations on other aspects of student behavior. For instance, since Kupei AI's learning system has many different grade levels available for students to work on, students may be working on problems beneath their grade level just to earn points quickly. Relatedly, it may be worth looking into the degree to which the changes observed across versions differs for younger and older students using the platform. It is also possible that these design changes may be influencing the frequency and form of gaming the system (Baker et al., 2004) that is occurring in the system.

Overall, over the course of these modifications, we were able to rapidly iterate Kupei AI's points system, removing the unintended consequences emerging in earlier versions and increasing the speed with which students mastered concepts through the system. This higher efficiency creates several opportunities, including the ability to cover a greater proportion of the year's content in a limited amount of after-school time and focus student time better on the content they need to learn. A next step will be to look at whether this improved distribution of time leads to greater learning gains on an external examination of knowledge, and the impacts of the re-design on student retention in the afterschool program. Overall, by allocating student time more effectively, we can lead to more efficient learning and in the long term, hopefully, better outcomes for learners.

Acknowledgements

We would like to thank the developers of the Kupei AI system and the students, teachers, and parents who used the system, for their participation in this research.

References

1. Cen, H., Koedinger, K. R., Junker, B.: Is Over Practice Necessary?-Improving Learning Efficiency with the Cognitive Tutor through Educational Data Mining. In Proceedings of the International Conference on Artificial Intelligence and Education, 551-558. IOS Press (2007).
2. Cook, T. D., Campbell, D. T.: Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston, Houghton Mifflin, (1979).

3. Corbett, A. T., & Anderson, J. R. Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction, 4(4), 253-278 (1995).
4. Deci, E. L., Koestner, R., Ryan, R. M. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin, 125, 627—668 (1999).
5. Dede, C., Richards, J., Saxberg, B. (Eds.). Learning engineering for online education: Theoretical contexts and design-based examples. Routledge (2018).
6. Dicheva, D., Dichev, C., Agre, G., Angelova, G. : Gamification in education: A systematic mapping study. Journal of educational technology & society 18(3), 75-88 (2015).
7. Koedinger, K. R., Corbett, A. T., Perfetti, C. The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive science 36(5), 757-798 (2012).
8. Koedinger, K. R., Sueker, E. L. F.: Monitored design of an effective learning environment for algebraic problem solving. Technical report CMUHCII-14-102. Carnegie Mellon University (2014).
9. Lepper, M. R., Greene, D.: The hidden costs of reward: New perspectives on the psychology of human motivation. Psychology Press (2015).
10. Li, W., Grossman, T., Fitzmaurice, G.: GamiCAD: a gamified tutorial system for first time autocad users. In Proceedings of the 25th annual ACM symposium on User interface software and technology, pp. 103-112. Association for Computing Machinery (2012).
11. Long, Y., Aleven, V.: Students' understanding of their student model. In International Conference on Artificial Intelligence in Education (pp. 179-186). Springer (2011).
12. Long, Y., Aleven, V.: Gamification of joint student/system control over problem selection in a linear equation tutor. In International conference on intelligent tutoring systems, pp. 378-387. Springer (2014).
13. Morford, Z. H., Witts, B. N., Killingsworth, K. J., Alavosius, M. P.: Gamification: The intersection between behavior analysis and game design technologies. The Behavior Analyst 37(1), 25-40 (2014).
14. Muñoz-Merino, P. J., Molina, M. F., Muñoz-Organero, M., & Kloos, C. D. An adaptive and innovative question-driven competition-based intelligent tutoring system for learning. Expert Systems with Applications 39(8), 6932-6948 (2012).
15. Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., Akinkuolie, B.: Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020. Computer Assisted Language Learning, 1-38 (2021).
16. Tahir, F., Mitrovic, A., Sotardi, V. : Investigating the effects of gamifying SQL-Tutor. Proceedings of the International Conference on Computers in Education (2020).
17. Terrell Jr, G., Kennedy, W. A.: Discrimination learning and transposition in children as a function of the nature of the reward. Journal of Experimental Psychology 53(4), 257 (1957).