Using Information Visualization to Promote Students' Reflection on "Gaming the System" in Online Learning

Meng Xia
HKUST
iris.xia@connect.ust.hk

Yuya Asano
University of Toronto
yuya.asano@mail.utoronto.ca

Joseph Jay Williams
University of Toronto
williams@cs.toronto.edu

Huamin Qu
HKUST
huamin@cse.ust.hk

Xiaojuan Ma
HKUST
mxj@cse.ust.hk

ABSTRACT
"Gaming the system" is the phenomenon where students attempt to perform well by systematically exploiting properties of the learning system, rather than learning the material. Frequent gaming tends to cause bad learning outcomes. Though existing studies tackle the problem by redesigning the system workflow to change students' behaviors *automatically*, gaming students discover new ways to game. We instead propose a novel way, reflective nudge, to *reflectively* influence students' attitudes by conveying reasons not to game via information visualizations. Particularly, we identify three common gaming contexts and involve students and instructors in co-designing three context-specific persuasive visualizations. We deploy our information visualizations in a real online learning platform. Through embedded surveys and in-person interviews, we find some evidence that the designs can promote students' reflection on gaming, and suggestive data that two of them can reduce gaming compared with control groups. Furthermore, we present insights into reflective nudge designs and practical issues concerning deployment.

Author Keywords
Information Visualization; Online Learning; Gaming the system; Reflective Nudge; Reflection.

CCS Concepts
•Human-centered computing → Information visualization; •Applied computing → E-learning;

INTRODUCTION
"Gaming the system" in online education refers to the phenomenon where students systematically exploit properties and regularities of the learning system, rather than learning the material [10]. Common gaming behaviors identified by previous literature include systematic trial-and-error and abuse of help (clicking through hints to get the answer) without thinking about how to solve the problem [11, 12, 9]. Such behaviors have been observed in educational games [27, 30], intelligent tutoring systems [11], online learning environments such as Massive Open Online Courses (MOOCs), and flipped classrooms [35]. The frequency of "gaming the system" behavior has a strongly negative correlation to learning outcomes [11, 12]. Students who game the system tend to have reduced learning gains [11, 12, 13, 20] and lower long-term academic achievements [41].

Most existing studies [4, 33, 24, 18] were proposed to reduce gaming by tweaking the system to make it harder to game. Though these methods are effective to a certain extent, they force students to follow the system setups *automatically* without encouraging their reflection on why gaming is not good. As a consequence, gaming students may discover new ways to work around these micro-interventions [33]. In other words, if tweaks fail to promote people' reflection on why a behavior change is necessary, their effects may fade away quickly once removed [15]. Therefore, it is critical to design *reflective* mechanisms that can promote students' reflection on gaming behavior and reduce gaming proactively by conveying information on reasons not to game according to different contexts.

Data visualization is one of the various ways to covey this kind of information. Its persuasiveness has been revealed in a wide range of recent research [19, 36, 1, 46]. For example, visualizations contribute more to peoples' understanding and persuasion than text in estimating drug efficacy or changing the attitude toward political topics [36]. In this paper, we propose the use of visualizations to convey reasons not to game. The reasoning information and the persuasive visualization work together as the reflective nudge to encourage students' reflection, instead of directly intervening on their behaviors. As introductory programming is one of the most popular courses on E-platforms such as edX [3], we consider the context of an online programming homework platform used in a large university-level introductory programming course and investigate three research questions (RQs). RQ1: What are the typical contexts in which students may try to game the system and what are the possible negative consequences on learning when gaming occurs in these contexts? RQ2: What are

L@S '20, August 12–14, 2020, Virtual Event, USA.

ACM ISBN 978-1-4503-7951-9/20/08 ...$15.00.
http://dx.doi.org/10.1145/3386527.3405924

the ways to encode information for communicating reasons not to game in various contexts into reflective nudge to students? RQ3: What are the design considerations for creating reflective nudge to promote reflection in online learning?

To answer these questions, we worked closely with students, instructors, and developers of the online learning system during the whole investigation. For RQ1, to identify the common beliefs that are amenable to change with the help of reflective nudge, we interviewed 16 students about the specific context(s) they would game and three course instructors (one of them is also a system developer) about the nature and ramifications of gaming in each context. For RQ2, we performed an iterative co-design with instructors and students and consolidated various design alternatives into three persuasive visualizations, targeting the gaming beliefs acquired in RQ1. For RQ3, we deployed our designs in the actual online learning system with 205 students taking the course. Students were randomly divided into four groups: one control group and three experimental groups (corresponding to the three information visualizations). For each information visualization, we collected students' ratings on the efficacy of information conveyance, reflection promotion, and visualization understanding through embedded surveys following the exercises. Additionally, students' submission logs were analyzed to check the potential gaming reduction. Furthermore, we conducted in-depth interviews with eight students to obtain their detailed feedback on the pros and cons of each visualization design as well as the room for improvement.

The key contributions of this work are summarized as follows.

1. We proposed a novel way – *reflective nudge* (reasoning information + persuasive visualization) – to address "gaming the system" behavior from the root. By interviewing students and instructors, we identified specific gaming contexts and students' beliefs that are amenable to change through persuasive visualizations.
2. We developed three persuasive visualizations to encode the reasons not to game and deployed them in an actual online homework system for students to experience. Feedback suggests that our designs evoked students to revisit their beliefs and promoted awareness of the ramifications of gaming to a certain extent. We also spotted some initial signs of reduced gaming behaviors after exposure to our designs compared with the control conditions.
3. We reported results from in-depth interviews with students who received these information visualizations to gain insights into reflective nudge designs in online learning and practical issues concerning deployment.

RELATED WORK

We first introduce the background of "gaming the system" behavior, then discuss the existing methods that address gaming behaviors and their limitations, and finally present related work on reflective nudges via information visualization.

"Gaming the System" Behavior

Gaming the system is a severe problem that exists in various forms on different kinds of learning platforms. E.g., in educational games [27, 30], students try to obtain top scores without learning any educational materials in the game. Another example in online course discussion forums [17] shows that students try to get credits by posting meaningless contents. This situation becomes even worse in intelligent systems as much gaming behavior can be exploited [4, 11, 24, 33]. Generally, there are two common types of gaming behaviors in intelligent tutoring systems [9]: quickly and repeatedly asking for help until the correct answer is provided, and quickly and systematically guessing the answers until correct (e.g., guessing numbers in order (1,2,3,4...) or ticking every checkbox or their combinations for multiple-choice questions). The previous study reported that a considerable number of students have been detected as having gaming behavior in MOOCs [35]. E.g., students even create two accounts for one course (one account is for guessing the correct answer and the answer is then used for the other account). It is reported that the ratio of students who were having any forms of gaming behavior during their study has reached a substantial ratio of 10-40% [9].

Therefore, there is an opportunity that designing technical interventions to help students regulate their attitudes and behaviors in online learning systems instead of wasting time on gaming, which is associated with less learning gains and lower long-term academic attainment [11, 12, 13, 20, 41].

Existing Mechanisms Addressing Gaming Behavior

Various researches have studied, proposed, and created interventions to address "gaming the system" behavior. In early studies, researchers added constraints to the system and applied the rules to all the students to prevent gaming behavior. For example, researchers introduced a two-second delay between each level of a multi-level hint to prevent students from clicking through hints quickly without processing them [4]. Researchers also re-designed the system to not give a hint until the student had spent a minimum amount of time on the current problem [24]. The disadvantage of these approaches is that they reduce the usefulness of the help features for non-gaming students [11]. To overcome this drawback, researchers have developed techniques on detecting gaming behavior using machine learning [9, 39] or feature engineering [23, 31]. They applied interventions only when students were detected as having gaming behavior, e.g., imposing more exercises on gaming students [18]. However, gaming students may discover new ways to work around these behavior interventions [33]. Previous research pointed out that it is hard to reform gaming students by only tackling the gaming symptoms [9].

Most of these methods are *automatic mechanisms*. According to the dual-process theories of decision-making in human behavioristics, people have two different thinking modes: automatic and reflective [26]. The automatic thinking mode dominates in contexts that demand quick decisions with little effort. It is instinctive, emotional, and unconscious. In contrast, the reflective mode allocates mental attention to make decisions. It is effortful, rational, and conscious. Little work has attempted to solve the problem using *reflective mechanisms*. One study tried to incorporate a virtual agent to express negative emotion when students were detected as having gaming behavior [18] and another study showed the performance graph to raise students' reflection [7]. However, nudging methods that lack educational effects fail to maintain efficacy once

Contexts of gaming the system	# of interviewees (out of 16)
C1: When students are busy, they may game to save time on this course.	10
C2: When faced with difficult problems, they feel frustrated and game to keep up the pace.	8
C3: They think some concepts are unimportant, thus game quickly through.	3
C4: When the video is not clear, they do not want to spend time on exercises.	2
C5: When the deadline is at noon, they can not get up early in the morning.	2

Table 1. Contexts in which students would game.

they are removed [15]. Thus, it is vital to convey information about reasons not to game in various contexts to promote students' reflection from the root.

Persuasiveness of Information Visualization

Data visualization is increasingly being adopted as a powerful and influential way to convey messages [36]. Particularly, the persuasiveness of information visualization has been mentioned in a range of research to change people's attitudes and behaviors [36, 1, 46]. One branch of work tries to utilize data visualization to change people's attitudes. For example, Anshul *et al.* compared the effect of visualization (chart) versus text (table) in terms of changing people's political attitudes. Results show that treatments with a graphical representation of the data (charts) are more likely to persuade participants who have neutral/weakly polarized opinions to change their attitudes as compared to treatments with the tabular representation of the data (tables) [36]. Other work attempts to use visualizations to change people's behavior. Agapie *et al.* [1] created an aureole around the query text box and it becomes blue when information is added to encourage longer queries in information seeking. The author observed that users typed longer than the absence of the aureole. Similarly, Turland *et al.* [46] used color (i.e. red for insecure networks and green for trusted ones) and position (i.e. placing the most secure options at the top) to label the networks' security. They found that color and positioning combined led to a significantly increasing rate of secure network selection for 60% of the participants while nudging by positioning alone was ineffective.

Studies have also been conducted in the domain of education, exploring the transformation of visualization into students' learning dashboards in order to encourage learning [2, 28, 16]. Lisa *et al.* [28] explored students' sense-making on visualizations with different frames of reference. One key finding is that the transformation of visualization in learning is complex and that students' learning dashboards should be supplemented with tailored messages that connect learning tasks with course goals or desired attributes [38]. The aforementioned work provided insights into designing persuasive visualizations conveying tailored information that promote students' reflection on their question-answering behavior and gaming behavior.

CONTEXT UNDERSTANDING OF GAMING

This section describes our process of exploring the reasoning information for not gaming the system (RQ1). We first collected the typical contexts where students may game from students' perspectives and then queried the possible negative consequences on learning from the instructors' perspectives. Finally, reasons not to game in three typical learning contexts are summarized.

Students' Perspectives

We take an online learning system that is used in a large university-level introductory programming course as our research platform. The system organizes weekly exercises according to the lectures and contains 133 tutorial videos, 309 multiple-choice questions, and 61 open-ended programming problems. We concluded that 46% of the exercises were for credits, which altogether accounted for 14% of the final grade; the others were optional. The system allows multiple submissions and provides partial scores (how many correct options in multiple-choice questions, shown in Fig. 4A for each submission. Students are required to finish the for-credit exercises before every week's deadline.

To understand the frequency of and reasons behind learners' gaming behaviors in this system, we recruited students on-site after the lecture and office hours. We told students that none of the researchers are the instructors and their responses would be kept anonymous and for research only (not affecting their course performance) in the consent form before each interview. Altogether, we conducted semi-structured interviews with 16 students from the first to the third year of study (12 males, age: 23$\pm$ 3.38), eight of whom are from the CS department. Each interview lasted about 15 mins. We introduced the purpose of our interview and asked mainly two questions: 1) How often do you indulge in gaming behavior, if at all? 2) Under what circumstances are you likely to game the system and why? We did not ask about students' experience with the system since our pilot interviews showed that when such questions were presented, students tended to focus more on problems specific to certain system features instead of gaming behavior that could happen in any e-learning systems.

From the interview results, we learned that all the interviewees had indulged in gaming behavior on 10-35% of all the multiple-choice questions during their learning processes on the platform by trying different options or option combinations quickly (A, B, C, D, AB...) and using the partial score (Fig. 4A) as the hint. We conducted a thematic analysis of the interview results and identified five typical contexts in which students may game the system (Table 4.1). The most common context is (C1), which was reported by 10 out of the 16 interviewees: when they are busy, they want to game to reduce the time spent on the course. *"(I gamed for) two weeks, when I had the midterm, or I had stuff in my other courses. And I had other stuff to work. So I'm really busy. I rushed to save time"*. The second most frequent gaming context is (C2) as mentioned by eight students: when they encounter difficult problems, they may give up and game to keep up the pace. *"I assume other students may spend less time on these problems* *and it was the only way [for me] to keep up with others and get credits"*. The other three situations were brought up by two or three interviewees. E.g., when students determine that some exercises are not important (C3), find tutorial videos hard to understand (C4), or have trouble getting up to meet the deadline at noon (C5), they would feel it is acceptable to just game the questions rather than spending time watching videos and solving the problems carefully.

Some of our results are consistent with previous research findings made by Baker *et al.* [9], who explored why students game online learning systems. They surveyed 210 high school students aged 12-14 who used intelligent tutoring systems (Cognitive Tutors [6] and ASSISTments [44]) as a complementary way to learn math. The top three reasons to game reported in their work were the dislike for math, lack of motivation, and frustration with difficult problems. Both their work and ours find that students tend to game when they do not have sufficient ability (e.g., C2 difficult question) or motivation (e.g., C3 trivial concepts in our case). There are two possible reasons why our interviewees seem to put more concerns over time (C1 and C5) than on the course's subject matter (as in Baker *et al.*). First, college students have more flexibility in choosing majors and courses based on their own interests compared to high-school students. Second, college students need to manage their own time [25].

Instructors' Perspectives

We conducted another round of semi-structured interviews on the context-specific consequences of "gaming the system" behavior with three course instructors (I1, I2, and I3) from our university (I1 is also a system developer). All three instructors have used this online learning system to assist their teaching (*Software Tools and Systems Programming* and *Introduction to Programming*) for at least three years. Each interview lasted 30-60 mins, with main questions including but not limited: what are the intentions behind the initial design of the system? what's your observed students' practice on the system? what are your attitudes toward certain practices? what are the suggestions and potential solutions? After transcribing all three interviews, two of the authors reviewed and coded the transcriptions independently. Then the same two authors compared codes to merge the similar codes and extracted themes.

First, all three instructors verified that there are indeed many students who game the system. *"The other thing is that students are not taking it seriously at first. Many students including good students will just game it and try all combinations of multiple choice because somehow they think that's faster and the best use of their time at that moment. Get the credit and then go on. I have people who come back to my office hours later and say I didn't really do it seriously and now I have a question about this thing."* as stated by I2. I3 added, *"When they come to me in office hours, they're very embarrassed because they did this."*

Second, instructors were strongly against "gaming the system" behavior. I2 noted that *"it's such a waste of time and seems probably not the best use of time. If people can take it a bit more seriously in the first place, that would be a better use of their time in many ways. It would make things better, better use of lecture time. There are so many reasons why doing it the first time making sense."* In addition, I1 pointed out that *" the problem with this is that once they see they got it (by gaming) and get the green check (an icon appears when students get the correct answer), they still get that mental cake."* We enquired of the system developer why repeated submissions are penalty-free. He explained that this design is to encourage students to hone their skills through spaced repetition [8], especially during the reviewing process at the end of the course.

Finally, we summarized instructors' reflections and suggestions on each of the aforementioned gaming contexts. As for C1, all instructors stressed that gaming can not save time. *"If you do it halfway, and then you have to come back. You're spending one and a half times longer than students who just try right for the first place. I try to avoid that with myself. I don't read your email through and then come back to read in detail because if I read twice then it's costly. But little tips and tricks that make people think that they're using their time well. Doing it in this way may not going to be best for them in the long run."* For C2, instructors suggested that showing the other students' number of attempts and the overall pass rate could reduce students' anxieties and potentially break the habit of giving up too soon. *"So the median of the attempts and the first-time pass rate gives you a sense of how difficult the problem is."* For C3, instructors implied learning is an accumulative process and that every concept matters. *"The other thing is that it's (the course) very cumulative in nature. So you couldn't realistically jump from week 2 to week 6 and be able to do week 6."* Instructors believed C4 and C5 could be addressed by improving the course design in the next semester.

In brief, we conclude three reasons not to game in three typical learning contexts (C1, C2, and C3). 1) Students game the homework by randomly guessing answers with the intent to save time, which would cost them much more time in the review period. 2) Students game in the face of difficult problems assuming it is the only way to keep up with their peers, but difficult problems also take other students considerable effort to solve. 3) Students game problems related to seemingly unimportant concepts, but the negligence of those concepts may hinder the mastery of later concepts depending on them.

VISUALIZATION DESIGN ITERATIONS

Based on the three contexts of gaming and reasons not to game, this section describes how to encode that reasoning information into persuasive visualizations to raise students' reflection on gaming (RQ2). First, we derived a set of design goals based on the previous interviews with the instructors. We then followed an iterative design process to create the persuasive visualizations, which involved both instructors and students. Three high-fidelity persuasive visualizations are generated and presented at last.

Design Goals

Before mocking up the visualization, we first derived the design goals according to the interviews with instructors and the previous literature using visualization in learning [15, 47, 14].

G1. Be informative: encode the reasons not to game. As mentioned by all three instructors (I1, I2, and I3), students know it is not good to guess the correct answer, but they do not know the severity of this behavior sometimes. E.g., they may have no ideas about how the learning concepts are connected and neglect the fact that missing one concept may affect the learning of future concepts. Thus, we should design different visualizations to convey various reasons not to game.

G2. Be persuasive: encourage alternative behavior instead of gaming. All instructors mentioned that the visualizations should encourage taking alternative next steps such as spending more time on reflection and reviewing prerequisites instead of gaming. It was also reported in previous work that students learning dashboards should offer enough actionable intelligence to optimize students' self-regulated learning [28].

G3. Be intuitive: make visualizations easy to understand. Since the target users are college students, most of whom have no professional training on the visualization literacy [5] (the capability "to read, comprehend, and interpret" graphs), it is necessary to encode the information intuitively and straightforwardly so that they are easy to understand [40, 28].

Iterative Design Process

We followed an iterative design process similar to [45] for the persuasive visualizations. During our initial ideation and prototyping phase, we developed 10+ low-fidelity (sketch) design alternatives based on the above design goals. For each of the three contexts, we mocked up three to four alternative designs. Specifically, for the first context (gaming to save time), we considered using a scatter plot or a line chart to show the relationship between time spent on solving problems and time on reviewing before the exam, or showing two examples (good students and struggling students) of ways to spend their time. Previous work [21] pointed out that students should be evaluated based on both learning outcomes and habits. Instructors confirmed that "good students" finish every exercise on time and have high scores in exams; "struggling students" miss deadlines sometimes and fail the exams. For the second context (facing difficult problems), we listed average time spent on this problem, first-time pass rate, or mean attempts to represent the difficulty level; for the third context (neglecting some problems), we proposed using a five-star rating or prerequisite-relationship graph among problems/learning concepts to show the importance of the current learning concept.

We then conducted two participatory interviews with two instructors (I1 and I2) one by one to get feedback on each of the visualizations. For the first context, they recommended using two examples (good student and struggling student), which may be more effective in showing the causality compared with the line graph or the scatter plot [48]. The line graph or scatter plot are also not easy to interpret since there are some points not linearly distributed. Then instructors selected an example of "good students" and one of "struggling students" from last semester's data. For the second context, they suggested combining the mean attempt and first-time pass rate. They also informed us that there is no data on accurate problem-solving time. For the third context, they chose the prerequisite graph of learning concepts since the five-star rating is not informative for future learning. After this process, we got three candidates with medium-fidelity.

Moreover, we performed informal testing with seven students (two females, five males, age: 24 ± 2.85) to improve the visual designs. To avoid affecting students' perceptions in the real deployment later in the on-going course, all seven students were recruited from previous semesters by snowball sampling. These tests revealed that a certain amount of text annotations were needed. They gave us many suggestions on the title and legends such as adding "the word 'exam"', "some titles like the sequence of your answers", and so on. They also provided suggestions on the layout; "horizontal is better for saving space". We modified the three visualizations according to students' suggestions and obtained the final three high-fidelity prototypes as shown in Fig. 1- 3.

Visualization Designs

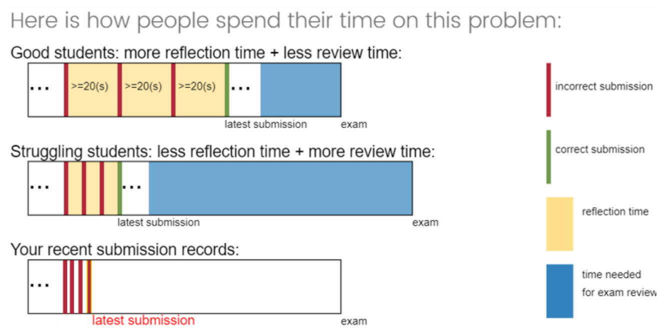

Figure 1. V1 shows how good students and struggling students spend their time on solving the problem and reviewing before the exam in order to convey gaming the system by submitting frequently may need to spend much more time reviewing the question before an exam.

V1 (Fig. 1) is designed to convey that gaming the system through frequent submissions may result in spending a much longer time reviewing questions before an exam (G1). There are three rows in this visualization. The first row describes how good students spend time on this question, the second row shows how struggling students spend time on this question, and the third row is the student's own submission sequence. Each bar on the row is a submission, whose colors, red and green, represent incorrect and correct submission, respectively. Yellow blocks mean the reflection time while blue ones represent the reviewing time before the exam. To achieve G2, first, by comparing the second row with the first row, we see that the yellow block is smaller between each bar while the blue trunk is much longer, by which we want to enhance the persuasiveness on spending more time between two submissions. Second, students can compare his/her records with the good students and struggling students to remind themselves of reflection instead of gaming. As shown in Fig. 1, this student has four recent submissions, and the time interval between each submission is short. Therefore, he/she tends to be similar to a struggling student, which may make it possible to persuade him/her to spend a longer time on reflection. We try to make it intuitive by using rectangles and basic colors (G3).

V2 (Fig. 2) is designed to convey that difficult problems take considerable effort for other students to solve and that it is unproductive to give up quickly by resorting to gaming the system (G1). According to instructors' suggestions, we use mean

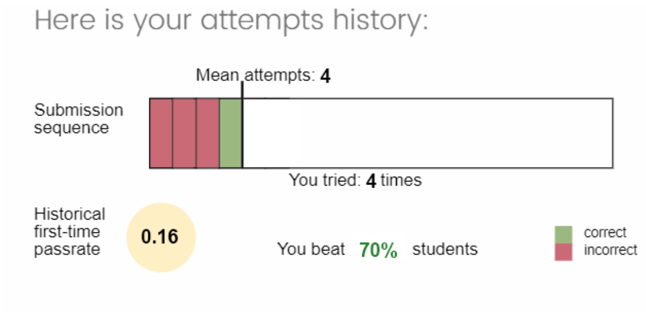

Figure 2. V2 shows the attempt times of peers by mean attempts and first-time pass rate from last semester in order to convey that difficult problems take considerable effort for other students to solve and that it is unnecessary to give up quickly by resorting to gaming the system.

attempts and the first-time pass rate to indicate the difficulty level of this problem. Additionally, we tell what percentage of students a student "beats" to make this design more persuasive. We use the red and green bars to represent incorrect and correct submissions, respectively (G3). To resolve the problem that no first-time pass rate exists when no one submits their answer in the exercise, we use the data of this exercise from the last semester. E.g., in Fig. 2, the historical first-time rate is only 0.16, and the number of mean attempts is four, indicating that this problem is quite difficult. Although this student solved the problem on the fourth attempt, he/she is better than or as good as 70% of students with more trials. In addition, if the student tries to guess the correct answer by submitting many times, he/she will beat fewer students, which persuades him/her to think carefully about being competitive in the class (G2).

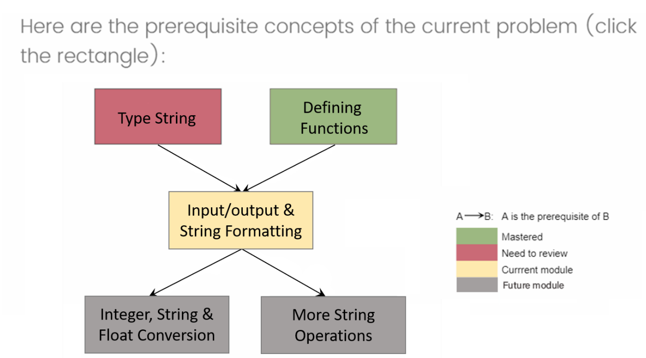

Figure 3. V3 shows one example of the prerequisite relationship among learning concepts in order to convey that learning concepts are connected and that the negligence of one concept may hinder the mastery of later concepts due to the cumulative nature of the course.

V3(Fig 3) is designed to convey that learning concepts are connected and that the negligence of one concept may hinder the mastery of later concepts due to the cumulative nature of the course (G1). Each rectangle represents a learning concept, and the arrow from A to B means A is the prerequisite of B. We constructed the concept hierarchy based on the prerequisites relationship labeled by the instructors for each question. The background color of the rectangle represents how much the student has mastered the learning concept. The green one means students' average score on that learning concept is high enough, indicating the student has mastered the concept, while the red one means that the corresponding learning concept needs to be reviewed. Furthermore, the yellow one means the current learning module, and the grey ones require an understanding of the current learning concept. The reason we use this structure is that existing research reported that the tree structure is easy to understand [29] (G3). We attempted to use this concept graph to show the connections between learning concepts and the importance of the current one. We also use red rectangles to persuade students to review previous learning concepts for solving the current problem if they can not solve them correctly (G2). These rectangles are clickable and can direct students to the corresponding pages.

EVALUATION

To answer RQ3, we deployed our interventions (reflective nudges) on the research platform we designed for. This serves as a technology probe for understanding users' needs and experiences in a real-world setting to inform the proper design of technologies. We analyzed data from students' submission records, online questionnaires, and in-depth interviews to gain insights into the effectiveness of our approach to gaming reduction, information conveyance, reflection promotion, and visualization understanding.

Experiment Design

This part describes the deployment setting, online questionnaires, and the procedure of post-study interviews.

Deployment Setting We launched our reflective nudges on the previously mentioned online programming homework platform, which has been used by a university-level introductory programming course with 205 students from various departments. Since gaming behavior is affected by many contextual factors such as time (beginning of the semester, midterm period, etc.) and problem attributes (topic, difficult level, etc.) [37], we selected problems from the last module of the course to ensure the relative consistency of these factors (time and topics). Also, it might be easier for us to observe the effect as a certain number of students are likely to game the system due to the time pressure in that period.

In particular, the last course module has four multiple-choice questions (P1-P4), and students have to finish them within two weeks. We applied our interventions to the last two questions and used the students' submission behavior on the first two questions to describe their gaming patterns before adding our interventions. We divided the students into four groups evenly (44) by randomizing their ids (three experimental groups, each receiving one of the three visualizations during the problem-solving process, and one control group without visualization intervention), but each group had some students who did not log in the system to do the exercises and thus the final numbers were uneven (39 students in the control group; 37, 44, and 38 students in V1, V2, V3 group, respectively).

Fig. 4 shows the system interface. The reflective nudges were embedded below the system feedback (Fig. 4C) and showed up each time students clicked the "submit" button (Fig. 4B). This setup aims to promote the intended reflection after submissions, based on the theory that the short-lived nature of priming effects demands the intended persuasive outcome to follow its corresponding priming stimulus closely in time [45].

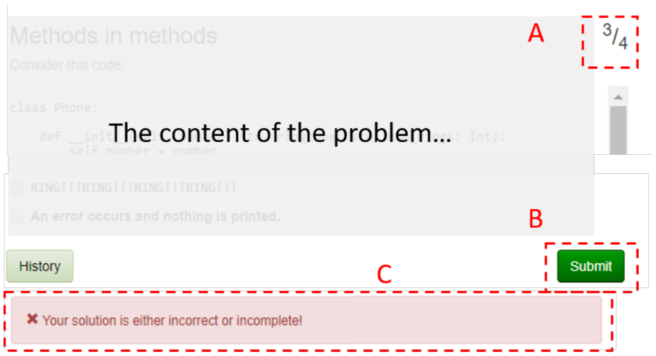

Figure 4. The interface of the online programming homework platform (A: Partial score of the current submission; B: "Submit" button; C: Result of the current submission, correct or incorrect).

Questionnaire Upon finishing all the questions, participants in the non-control groups were invited to fill out a post-study online questionnaire (optional) derived from a previous study [47] and the requirements by the instructors that we were not supposed to disturb the process with a long and open-ended questionnaire. These questions, listed below, are to measure the efficacy of the proposed gaming interventions in terms of information conveyance (Q1), reflection promotion (Q2 and Q3), and visualization understanding (Q4) on a 7-point Likert scale (from "1. Strongly disagree" to "7. Strongly agree"). Note that Q1 has three versions (Q1-1, Q1-2, and Q1-3) corresponding to three types of reflective nudges, respectively. **Q1-1** This visualization clearly shows that 'gaming the system may cause a longer time to review before the exam to ensure the final performance'. **Q1-2** This visualization clearly shows 'peers also spend considerable efforts on the difficult problems and gaming the system may cost more attempts than peers'. **Q1-3** This visualization clearly shows 'a learning concept is related to other learning concepts, so gaming the system on the questions may impair the performance on the others'. **Q2** This visualization helps you reflect on your perception of "gaming the system" behavior. **Q3** This visualization helps you reflect on your question-answering behavior. **Q4** It is easy to understand this visualization.

In-depth Interviews Since four questions with 7-point scales may not gather all information, we conducted post-study interviews with students who had received the interventions to gather reasons behind their questionnaire ratings and suggestions on these intervention designs. Altogether, we interviewed eight students (S1-S8, six males, age: 24 ± 3.48, with backgrounds in computer science, political science, psychology, and economics), whom we recruited on-site after office hours and sent recruitment emails to the course mailing list. As nearly no new insight emerged after the sixth interviewee, we did not conduct further recruitment. We asked about which interventions they received and their ratings again since we could not cross-check their anonymous survey results. Four of them received V2, two V1, and two V3. All procedures were approved by the local university's Institutional Review Board. Each interview lasted about 60 mins and students were compensated with $8.

The interview proceeded as follows. (1) We collected the participant's consent by signing the consent form. (2) We provided a 5-min introduction of the project and then collected their responses to general questions such as "did you finish the last two questions in week 12?", "which visualization did you receive?". We also showed the platform to help interviewees to recall their experiences with the visualization designs. (3) We went over all the three designs with each interviewee and asked Q1 - Q4 for the elaboration on their ratings. For example, if a participant gives a rating 5 (Somewhat Agree) in Q2, we would ask them which part(s) of the visualization helped on their reflection on gaming behavior and how. In addition, we added a fifth question to gain insights into room for improvement: **Q5** How can it be improved to help you promote your reflection on your "gaming the system" behavior?

Results and Analysis

We report the students' submission data, questionnaire ratings, and interview feedback in the following four aspects: indication of potential gaming reduction, information conveyance, reflection promotion, and visualization understanding.

Potential Gaming Reduction Although the goal of this study is not to systematically validate the behavior-changing effect of our designs, we still inspect students' behavior to check the potential influence on behavior. We identified all possible gaming behavior in students' submission records based on two distinctive patterns extracted students' self-reports as presented in Section 3 (Context Understanding of Gaming): 1) "submitting the answer many times within a very short period of time", and 2) "submitting an empty solution first, followed by another answer shortly afterward according to the partial score returned by the system". Then we manually labeled submissions records of five problems from the previous semester by two authors. By tuning the thresholds of time duration between two submissions and submission times to distinguish gaming from non-gaming based on labeled data, we then calculated the ratio of gaming students for each problem accordingly.

	P1	P2	P3	P4
V1-time spending(37)	0.30	0.65	0.11	0.08
V2-attempt number(44)	0.34	0.63	0.16	0.07
V3-prerequisite graph(38)	0.37	0.63	0.21	0.16
Baseline1-control group(39)	0.26	0.59	0.21	0.23
Baseline2-last semester(138)	0.32	0.65	0.21	0.16
First-time pass rate	0.26	0.09	0.71	0.56

Table 2. Gaming proportions change over time on four problems, among three experimental groups (V1, V2, and V3) and two Baselines (control group and last-semester group). The number of students in each group is shown in the parentheses. The last row is the first-time pass rate of these four problems in this semester to indicate the difficulty levels.

Table 2 shows the changes in the proportion of students with gaming behaviors across problems P1 - P4 among three experimental groups (V1-time on problems, V2-number of attempts, and V3-prerequisites graph) and one control group (Baseline1). We also added another baseline (Baseline2) – historical submission records on these four problems from the previous semester. The first-time pass rate of each problem is also provided for reference. Problems P1 and P2 were released just before the deployment of our interventions on P3 and P4.

As shown in Table 2, while the gaming portion over the four problems fluctuated with the problem's level of difficulty inferred from the first-time pass rate (P1-P4: 0.26, 0.1, 0.71, 0.56), one can identify some evidence that the proportion of gaming students decreased in V1 and V2 compared with Baselines. These two groups had a larger or similar percentage of students who gamed on P1 and P2 compared with the Baselines. After launching V1 and V2, the rate of gaming dropped to 11% and 16% on the first problem P3, smaller than 21% in the control group and previous semester. By the second problem P4, the rate of gaming further declined to 8% and 7%, much lower than 20% in Baseline1. We used z-test to evaluate whether there are any significant differences of gaming students between V1, V2, V3 with Baselines on P3 and P4. In particular, we found that on P4, V1 has a marginally significant difference ($p = 0.07 < 0.1$) compared with Baseline1 and that V2 has a significant difference compared with Baseline1 ($p = 0.04 < 0.05$), while V3 has no significant difference compared with both Baselines. Generally, it seems that the exposure to V1 and V2, but not V3, might have a potential impact on students' actual practice. We further explore potential explanations in the post-study interviews.

	Q1-Information conveyance	Q2-Reflection on gaming	Q3-Reflection on question-answering	Q4-Easy to understand
V1	4.6(1.7)	4.3(1.8)	4.3(1.8)	3.7(1.9)
V2	5.2(1.5)	4.7(1.2)	4.4(1.3)	5.1(1.2)
V3	5.8(0.7)	5.0(1.2)	5.3(1.1)	4.7(1.8)

Table 3. Means and standard deviations (in parentheses) for ratings of agreement (1-strongly disagree, 4-neither agree nor disagree, 7-strongly agree) on questions: Q1. information conveyance, Q2. reflection on gaming, Q3. reflection on question-answering, and Q4. easy to understand by conditions (V1, V2, and V3).

Information Conveyance Tab. 3 shows the means and standard deviations of students' ratings on Q1-Q4 in the optional online survey. The numbers of valid responses we received in V1, V2, V3 are 11, 11, and 12, respectively. Survey question Q1 is used to test whether the visualization clearly conveys the information, reasons not to game, to students (design goal G1). Results show that the average scores are all above 4 (neither disagree nor disagree) across V1, V2, and V3, confirming that our designs deliver the message clearly to some extent.

In the interview, six out of eight interviewees generally thought V1 communicates with clarity the idea that if students spend less time on solving questions seriously, they will need more time to review before the exam. *"The big blue block clearly shows a longer review time. It serves as a warning."* – S1, male, 25. *"Obvious to know to guess more will meet trouble in the exam period."* – S5, male, 28. *"It gives a clear idea of how students should structure their time."* – S6, male, 29. S4 said that students might fail to notice the blue block (exam review), which would hinder their understanding of the information. Additionally, S3 suggested there should be gaming text explicitly shown adjacent to the visualization.

All participants believed that V2 allows users to calibrate their performance against peers straightforwardly, indicating that there is no need to rush on difficult problems. They felt it was evident that *"guessing can not make you better than peers"*– S6, male, 29. *"And the visualization helped students focus on the difficulty level by showing the attempts information, the average amount of effort required, and first-time pass rate."*– S8, female, 25.

Six out of eight students thought V3 clearly conveyed that gaming the system on the questions might impair the performance on others because one learning concept is related to other learning concepts. *"It shows you, in a different way, why guessing isn't going to lead you spending less time by showing you how the questions themselves are connected correspondingly."* – S6, male, 29. *"We're just looking on how to improve learning behavior, which is good. Trying to eliminate guessing behavior is how you improve learning."* –S7, male, 24. Two participants (S1 and S8) mentioned that relating the visualization with "guessing" behavior was not easy.

In general, these three visualizations convey the corresponding information to a certain degree. It is somewhat effective to use the designed visualization to communicate reasons not to game with students.

Reflection Promotion Our designs try to promote students' reflection on gaming – about the negative effect it entails and alternative strategies instead of gaming (design goal G2). We asked two questions regarding this matter in our survey and interviews: (Q2) revisitation of one's perception of gaming, and (Q3) introspection of one's current question-answering behavior. Average survey ratings of Q2 and Q3 are all above 4 (neither agree nor disagree), which indicates that our designs can encourage students to review the ramifications of gaming to some extent.

In the post-study interviews, six out of eight students gave comments on how V1 stimulated their reflections. Apart from spending more time on thinking to save time for exam reviews, two students found it good for hard-working students to persist. E.g., *"if a student who is trying to work hard but haven't got the correct answer and they see they fall in this category (good students' in V1). I think it's very good because I think it's even more helpful than for the other students. So they don't feel frustrated. It is consistent with all these other successful students. So they know that success is coming. They just have to keep doing what they're doing."* – S7, male, 24. They also gave certain suggestions for removing some redundant information in the design to improve the effectiveness of promoting reflection; e.g, *"remove the 20s"*(S7) and *"not necessary to differentiate between correct and incorrect submission"* (S5).

As for V2, six students suggested that they had a deeper understanding of the negative consequences brought by gaming and were clearer about the next steps they would take instead of gaming. They discussed how different components of the design led them to take gaming more seriously. *"The first-time pass rate helps reflect on how other students treat this problem since it's very, very unlikely for someone totally guessing to get around the first."* – S7, male, 24. *"The average number pushes people to reflect on their own position compared with* *the class and whether they should work harder."* – S8, female, 25. *"'Beat' and 'first-time pass rate' make me feel like I will be more cautious in the future and spend more time on learning and verifying."* – S3, male, 23. The interviewees also pointed out some potential drawbacks of the V2 design. S4 compared it with V1 and said that *"mean attempts are not so alarming as the exam review bar."* Additionally, S1 and S7 were concerned that *"low self-esteem or hard-working students might get hurt by seeing this (their attempts more than the mean attempts)."*

Four students thought V3 got them to revisit their belief about gaming behavior. *"It's kind of hint and helps you think the question twice."* – S1, male, 25. However, seven students felt that *"it is more reflective on learning than gaming"* – S7, male, 24. In the interview, the students mentioned more specifically that *"it helps reflect and think on how concepts are connected"* – S1, male, 25. *"One more towards your actually reflecting on the course content and things that you need to understand to move forward... get them (other students) to review before question/exam."* – S5, male, 28. The interviewees also noted that V3 assisted them in identifying space to improve. *"It's focused very clearly on the exact skills that they need to do well, telling people where are missing and reinforcing people wanting to actually spend time on the work, guiding them on how to get to that feeling (rewarding)."* – S6, male, 29.

Overall, gaming directly relates to people's submission behavior (emphasis of V1 and V2 design), but its association with higher-level learning concepts behind the homework problems (the core of V3 design) is not as apparent. Students thought things related to exams, time, and peers' performances can more easily be linked with gaming behavior. This observation might as well explain the less obvious gaming behavior reduction in V3 on problems P3 and P4.

Visualization Understanding Our third design goal (G3) is to make the designs intuitive. Q4 is designed to verify whether the visualizations are easy to understand. From the rating, the mean scores of V2 and V3 are above 4 (neither agree nor disagree), but below 4 for V1.

According to the interviews, four out of eight students thought V1 was not readily understandable. The main reasons identified are that there are *"too many components"* (S1, S3, and S4), that *"fonts are small"* (S7), and that it is *"not clear where you should start reading"* (S7). S5 further explained, *"it's near the exam and students do not want to spend extra time on extra information."* Six out of eight students found V2 rather straightforward and intuitive. V2 is easier to understand than V1 because *"there are only two colors"* (S4), *"the graphics match very well"* (S6), and *"the green and red colors are a good stimulus, like the traffic light in the psychology area"* (S2). Students thought V3 had a clear layout with the tree structure and the only change they brought up was that the legends could be bigger (S4).

In sum, V2 and V3 are rather easy to grasp for students in the real-world deployment, while V1 is somewhat complex to comprehend fully in a short period. This calls our attention to the fact that real-world deployment is stricter about the expressiveness and simplicity of visual design.

Summary We have some suggestive data that V1 and V2 may have better gaming reduction effects compared with the control groups on the two problems we deployed. The online rating and in-person interviews show that all three visualizations (V1, V2, and V3) convey reasons not to game to a certain extent. They all encourage students to review the ramifications of gaming behavior to some degree. In particular, V1 and V2 promote reflecting more on gaming itself, whereas V3 seems to lead students to think more about their question-answering behavior. V2 and V3 are easier to understand than V1.

DISCUSSION

This section first summarizes a set of design considerations of reflective nudges in online learning. Further, it discusses the generalization and limitations of the current work.

Design Considerations

We list five design considerations for reflective nudge designs from both information and visualization perspectives.

Color is effective. As for visual design, among all visual encodings (e.g., size, shape, etc.), color was mentioned the most and consistently by the students with a short exposure to different visualizations and its persuasiveness was not addressed by previous research related to learning. They identified several usages of colors to make a visual nudge more effective. Colors can show the contrast between "good" and "bad". *"(For V3) I like red as the need to review and green as you master it. Red is like to stop. Green is like to continue to work. Then when you see green, you go to the next one. I think the color scheme works. I think this is very clear"* – S6, male, 29. *"(For V3) the green and red color are good stimuli, like the traffic light in the psychology area"* – S2, female, 19. *"(For V1) so let's say if it takes more than 20 seconds, and it's (space between two submissions) yellow. If it's less than 20 seconds, it shows a different color."* – S7, male, 24. Moreover, the color highlight can make the corresponding information more visually salient. *"Use more bight color to emphasize the exam review."* – S3, male, 23; S4, male, 20. One thing to keep in mind is that one should not use too many colors in the persuasive visualization [32]. *"Because the other one (V1) has like a few more colors like red, green, yellow and blue. Here, there's (V2) only red or green. The message is more clear."* – S7, male, 24.

Perceived authenticity increases persuasiveness. Enhancing the perceived authenticity of data can improve the persuasiveness. In our interview, five out of eight students asked for revealing the data source (e.g., which semester) in V2 and V3 to make the message more compelling. *"(In V2), show it explicitly that the data (historical first-time pass rate) is from *** course from 2018 winter semester. People will be more sensitive."* – S5, male, 28. *"(In V3), maybe you can add that it's suggested by the instructor."* – S3, male, 23. This result is in line with the conclusion in the previous research that highly educated people were more likely to value the source of data visualization [40].

Time-related information is important. Time management is a critical issue in online learning [34, 28], especially when one has committed to the rigorous demands of a college education. We learned that saving time is the top excuse for gaming among the university students we had interviews with at the needs-finding stage. According to the post-study feedback, one possible reason why V1 may have a relatively good effect on gaming reduction is that it shows students how to manage their time more effectively. Recent studies on learning dashboards also show that students cared more about the graphical information related to time expenditure [28], and they appreciated being aware of how individuals spend their time [42]. Therefore, designers should integrate the time-related information to motivate students' learning behaviors.

Connecting to peers is worth considering. Showing information related to peers can stimulate students to reflect. Four interviewees liked V2 because it made them feel that they were working alongside other students, which aligns with the "relatedness" rule (i.e. doing the learning activities helps them feel more connected to others) of motivating students [22]. Showing the average performance, the first-time pass rate of the class (V2), and the examples of good and struggling students (V1) are all possible ways to help students better position themselves among peers. *"So you'll feel connected to this ('beat *** students'). This number will directly remind where you stand in the class."* – S5, male, 28. However, one should be aware that while showing such information seemed to benefit most of the interviewees, it may have certain side effects on students who work hard but are not good at this problem as discussed in Results and Analysis. One thing we could do to mitigate possible negative effects of having peer information is to personalize the visualizations according to students' learning state (gaming or hard-working). Furthermore, showing information related to peers' learning habits (e.g., time spent on solving problems) instead of emphasizing performance may make a nudge less aggressive, e.g., in V1, no students interviewed commented that there are any side effects.

Ensuring a good grasp of information is critical. According to the post-study interviews, students tend to have a strict requirement of intuitiveness in a real-world setting. *"It takes me 4-5 seconds to understand, but it needs to reduce down to 2-3 seconds (for V1)."* – S3, male, 23. It is not a simple job to make informative and persuasive visualizations that can be grasped within several seconds when the information is complex or its amount is large. There are two possible ways to ensure effective conveyance of information in such a case. The first approach is to present the information at various levels and allow interactive exploration. We can hide the details and ensure that the most critical message is visually salient and intuitive to comprehend in a very short time. If users are interested in learning more, they can bring up further information interactively. For example, in V3 which aims to show that learning concepts are connected, we only display the names of the learning concepts first. If users want to learn more about the actual contents, they can click on the rectangle to open the corresponding page. The second method is to attract and hold users' attention so that they stay with the visualization for a long enough time to process all the information. For instance, if the information is abstract or complex, we can apply techniques like animation as suggested by S5, *"maybe you can use animation to show how good students and struggling students spend their time to attract students (for V1)."*

Generalizability to Other Platforms

Our designs (V1, V2) can be easily generalized to other learning platforms that allow multiple submissions. The reason is that V1 and V2 only use the submission records to calculate the time between consecutive submissions and the number of attempts. Online learning systems always keep such records for users to track their progress, and thus V1 and V2 can be readily employed on them. V3 may need the prerequisites relationship among instructional units, which can either be labeled by instructors or inferred according to submission records [43]. For learning platforms that do not allow multiple submissions, our designs (V1, V2) can be easily adjusted to replace the submission with other learning behavior such as seeking help, inquiring hints, or watching videos during the learning process.

Limitations and Future Work

First, we only deployed our visualization interventions on two problems in one week. While our goal is to use our information visualizations as a technology probe for understanding users' needs and experiences in a real-world setting to inform the proper design of reflective technologies, the time is too short to test whether our designs have any significant effect on reducing gaming behavior. Second, we did not personalize which visualization to display according to student attributes or problem properties, e.g., whether the current problem is difficult or not for a specific student, whether the student is working hard or not. We randomly divided the students into four groups in order to collect a relatively equal amount of feedback for each information visualization design since the current study focused on exploring what kinds of information and designs are more effective in promoting student's reflection in general. In the future, we plan to deploy the information visualizations for a longer time to systematically validate the behavior-changing effects in a personalized manner as well as their influence on the learning outcomes. Furthermore, we plan to explore more reflective nudge designs such as incorporating interaction or animation techniques.

CONCLUSION

In this work, we proposed reflective nudges, a new way of promoting students' reflection on why gaming is not encouraged, using persuasive visualizations. Specifically, we identified three common gaming contexts and involved students and instructors in co-designing context-specific persuasive visualizations. We deployed our information visualizations on a real online programming homework platform. Through embedded surveys and in-person interviews, we found some evidence that the designs can promote students' reflection on gaming and got suggestive data that two of the visualizations may reduce gaming compared with the control group. We presented design considerations on reflective nudges in online learning.

Acknowledgements

This work is partially sponsored by Innovation and Technology Fund (ITF) with No. ITS/388/17FP and Xiaojuan Ma is the corresponding author.

REFERENCES

[1] Elena Agapie, Gene Golovchinsky, and Pernilla Qvarfordt. 2013. Leading people to longer queries. In *Proceedings of the 2013 SIGCHI Conference on Human Factors in Computing Systems*. Citeseer, 3019–3022.

[2] Stephen J Aguilar and Clare Baek. 2019. Motivated Information Seeking and Graph Comprehension Among College Students. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. ACM, 280–289.

[3] Carlos Alario-Hoyos, C Delgado Kloos, Iria Estévez-Ayres, Carmen Fernández-Panadero, Jorge Blasco, Sergio Pastrana, and J Villena-Román. 2016. Interactive activities: the key to learning programming with MOOCs. *Proceedings of the European Stakeholder Summit on Experiences and Best Practices in and Around MOOCs, EMOOCS* (2016), 319.

[4] Vincent Aleven. 2001. Helping students to become better help seekers: Towards supporting metacognition in a cognitive tutor. *German-USA Early Career Research Exchange Program: Research on Learning Technologies and Technology-Supported Education, Tubingen, Germany* (2001).

[5] Basak Alper, Nathalie Henry Riche, Fanny Chevalier, Jeremy Boy, and Metin Sezgin. 2017. Visualization literacy at elementary school. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5485–5497.

[6] John R Anderson, Albert T Corbett, Kenneth R Koedinger, and Ray Pelletier. 1995. Cognitive tutors: Lessons learned. *The journal of the learning sciences* 4, 2 (1995), 167–207.

[7] Ivon Arroyo, Kimberly Ferguson, Jeffrey Johns, Toby Dragon, Hasmik Meheranian, Don Fisher, Andrew Barto, Sridhar Mahadevan, and Beverly Park Woolf. 2007. Repairing disengagement with non-invasive interventions. In *AIED*, Vol. 2007. 195–202.

[8] David P Ausubel and Mohamed Youssef. 1965. The effect of spaced repetition on meaningful retention. *The Journal of General Psychology* 73, 1 (1965), 147–150.

[9] Ryan Baker, Jason Walonoski, Neil Heffernan, Ido Roll, Albert Corbett, and Kenneth Koedinger. 2008. Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research* 19, 2 (2008), 185–224.

[10] Ryan Shaun Baker, Albert T Corbett, and Kenneth R Koedinger. 2004a. Detecting student misuse of intelligent tutoring systems. In *International conference on intelligent tutoring systems*. Springer, 531–540.

[11] Ryan Shaun Baker, Albert T Corbett, Kenneth R Koedinger, and Angela Z Wagner. 2004b. Off-task behavior in the cognitive tutor classroom: when students game the system. In *Proceedings of the 2004 SIGCHI conference on Human factors in computing systems*. ACM, 383–390.

[12] Ryan Shaun Baker, Ido Roll, Albert T Corbett, and Kenneth R Koedinger. 2005. Do performance goals lead students to game the system?. In *AIED*. 57–64.

[13] Joseph Beck and Ma Mercedes T Rodrigo. 2014. Understanding wheel spinning in the context of affective factors. In *International conference on intelligent tutoring systems*. Springer, 162–167.

[14] Susan Bull, Blandine Ginon, Clelia Boscolo, and Matthew Johnson. 2016. Introduction of learning visualisations and metacognitive support in a persuadable open learner model. In *Proceedings of the sixth international conference on learning analytics & knowledge*. ACM, 30–39.

[15] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 503.

[16] Sorathan Chaturapruek, Thomas S Dee, Ramesh Johari, René F Kizilcec, and Mitchell L Stevens. 2018. How a data-driven course planning tool affects college students' GPA: evidence from two field experiments. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM, 63.

[17] Ran Cheng and Julita Vassileva. 2005. Adaptive Reward Mechanism for Sustainable Online Learning Community.. In *AIED*. 152–159.

[18] Ryan SJ d Baker, Albert T Corbett, Kenneth R Koedinger, Shelley Evenson, Ido Roll, Angela Z Wagner, Meghan Naim, Jay Raspat, Daniel J Baker, and Joseph E Beck. 2006. Adapting to when students game an intelligent tutoring system. In *International Conference on Intelligent Tutoring Systems*. Springer, 392–401.

[19] Pierre Dragicevic and Yvonne Jansen. 2017. Blinded with Science or Informed by Charts? A Replication Study. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 781–790.

[20] Stephen E Fancsali. 2013. Data-driven causal modeling of" gaming the system" and off-task behavior in Cognitive Tutor Algebra. In *NIPS workshop on data driven education*.

[21] George Farkas. 2003. Cognitive skills and noncognitive traits and behaviors in stratification processes. *Annual review of sociology* 29, 1 (2003), 541–562.

[22] Larry Ferlazzo. 2015. *Building a community of self-motivated learners: Strategies to help students thrive in school and beyond*. Routledge.

[23] Jeffrey Johns and Beverly Woolf. 2006. A dynamic mixture model to detect student motivation and proficiency. In *Proceedings of the national conference on artificial intelligence*, Vol. 21. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 163.

[24] E Joseph. 2005. Engagement tracing: using response times to model student disengagement. *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology* 125 (2005), 88.

[25] Mera M Kachgal, L Sunny Hansen, and Kevin J Nutter. 2001. Academic procrastination prevention/intervention: Strategies and recommendations. *Journal of Developmental Education* 25, 1 (2001), 14.

[26] Daniel Kahneman. 2011. *Thinking, fast and slow*. Macmillan.

[27] Maria M Klawe and others. 1998. Designing game-based interactive multimedia mathematics learning activities. In *Proceedings of UCSMP International Conference on Mathematics Education*.

[28] Lisa Lim, Shane Dawson, Srecko Joksimovic, and Dragan Gašević. 2019. Exploring students' sensemaking of learning analytics dashboards: Does frame of reference make a difference?. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. ACM, 250–259.

[29] Lim Kian Long, Lim Chien Hui, Gim Yeong Fook, and Wan Mohd Nazmee Wan Zainon. 2017. A Study on the Effectiveness of Tree-Maps as Tree Visualization Techniques. *Procedia Computer Science* 124 (2017), 108–115.

[30] Rikke Magnussen and Morten Misfeldt. 2004. Player transformation of educational multiplayer games. *Proceedings of other players* (2004).

[31] Kasia Muldner, Winslow Burleson, Brett Van de Sande, and Kurt VanLehn. 2011. An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. *User modeling and user-adapted interaction* 21, 1-2 (2011), 99–135.

[32] Gerald M Murch. 1984. Physiological principles for the effective use of color. *IEEE Computer Graphics and Applications* 4, 11 (1984), 48–55.

[33] R Charles Murray and Kurt VanLehn. 2005. Effects of Dissuading Unnecessary Help Requests While Providing Proactive Help.. In *AIED*. Citeseer, 887–889.

[34] Ilona Nawrot and Antoine Doucet. 2014. Building engagement for MOOC students: introducing support for time management on online learning platforms. In *Proceedings of the 23rd International Conference on world wide web*. ACM, 1077–1082.

[35] Curtis G Northcutt, Andrew D Ho, and Isaac L Chuang. 2016. Detecting and preventing "multiple-account" cheating in massive open online courses. *Computers & Education* 100 (2016), 71–80.

[36] Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. 2014. The persuasive power of data visualization. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2211–2220.

[37] Luc Paquette, Ryan S Baker, and Michal Moskal. 2018. A system-general model for the detection of gaming the system behavior in CTAT and LearnSphere. In *International Conference on Artificial Intelligence in Education*. Springer, 257–260.

[38] Abelardo Pardo. 2018. A feedback model for data-rich learning experiences. *Assessment & Evaluation in Higher Education* 43, 3 (2018), 428–438.

[39] Zachary A Pardos, Ryan SJD Baker, Maria OCZ San Pedro, Sujith M Gowda, and Supreeth M Gowda. 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 (2014), 107–128.

[40] Evan M Peck, Sofia E Ayuso, and Omar El-Etr. 2019. Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 244.

[41] Maria Ofelia Pedro, Ryan Baker, Alex Bowers, and Neil Heffernan. 2013. Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In *Educational Data Mining 2013*.

[42] Ronald Pérez-Álvarez, Mar Pérez-Sanagustín, and Jorge J Maldonado-Mahauad. 2017. NoteMyProgress: Supporting learners' self-regulated strategies in MOOCs. In *European Conference on Technology Enhanced Learning*. Springer, 517–520.

[43] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein. 2015. Deep knowledge tracing. In *Advances in neural information processing systems*. 505–513.

[44] Leena Razzaq, Mingyu Feng, Neil T Heffernan, Kenneth R Koedinger, Brian Junker, Goss Nuzzo-Jones, Michael A Macasek, Kai P Rasmussen, Terrence E Turner, and Jason A Walonoski. 2007. A web-based authoring tool for intelligent tutors: blending assessment and instructional assistance. In *Intelligent Educational Machines*. Springer, 23–49.

[45] Joseph Seering, Tianmi Fang, Luca Damasco, Mianhong'Cherie' Chen, Likang Sun, and Geoff Kaufman. 2019. Designing User Interface Elements to Improve the Quality and Civility of Discourse in Online Commenting Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 606.

[46] James Turland, Lynne Coventry, Debora Jeske, Pam Briggs, and Aad van Moorsel. 2015. Nudging towards security: Developing an application for wireless network selection for android phones. In *Proceedings of the 2015 British HCI conference*. ACM, 193–201.

[47] Meng Xia, Mingfei Sun, Huan Wei, Qing Chen, Yong Wang, Lei Shi, Huamin Qu, and Xiaojuan Ma. 2019. PeerLens: Peer-inspired Interactive Learning Path Planning in Online Question Pool. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 634.

[48] Cindy Xiong, Joel Shapiro, Jessica Hullman, and Steven Franconeri. 2019. Illusion of Causality in Visualized Data. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 853–862.