Procrastination and Gaming in an Online Homework System of an Inverted CS1

Jaemarie Solyst
Carnegie Mellon University
Pittsburgh, PA, United States
jsolyst@andrew.cmu.edu

Trisha Thakur
University of Toronto
Toronto, ON, Canada
trisha.thakur@mail.utoronto.ca

Madhurima Dutta
University of Toronto
Toronto, ON, Canada
madhurima.dutta@mail.utoronto.ca

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

Andrew Petersen
University of Toronto Mississauga
Mississauga, ON, Canada
andrew.petersen@utoronto.ca

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

ABSTRACT

Engaged preparation and study in combination with lectures are important for all courses but are particularly critical for online, hybrid, and inverted classrooms. Many instructors use online systems to deliver new course content and exercises, but students often delay assignments or game these systems (e.g., guessing on multiple-choice questions), often to the detriment of their learning. In an inverted CS1 course, many students self-reported high rates of gaming-the-system behavior, so we examine survey data to identify factors that contribute to engagement in these maladaptive behaviours. We supplement that analysis with interview data to gain a deeper understanding of the situation. We also implemented and evaluated a previously reported online intervention aimed at reducing gaming behavior. Unlike prior work, our intervention did not have a significant effect on guessing behavior. We discuss why the factors we identified might explain this result, as well as suggest future work to improve our understanding of gaming behaviours and inform the design of systems that encourage effective learning.

CCS CONCEPTS

• **Applied computing → E-learning**; • **Social and professional topics → CS1**; *Computing education*; *Computer science education.*

KEYWORDS

Gaming the system, online learning, inverted classroom, procrastination

ACM Reference Format:

Jaemarie Solyst, Trisha Thakur, Madhurima Dutta, Yuya Asano, Andrew Petersen, and Joseph Jay Williams. 2021. Procrastination and Gaming in an Online Homework System of an Inverted CS1. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21), March 13–20, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3408877.3432440

ACM ISBN 978-1-4503-8062-1/21/03…$15.00
https://doi.org/10.1145/3408877.3432440

1 INTRODUCTION AND RELATED WORK

With the rising popularity of hybrid and inverted classroom environments [14, 15], many instructors have turned to online systems to provide lesson content and practice exercises. In such online environments, effective student preparation before class meetings is the most important factor for determining whether the classrooms are more effective and engaging than traditional ones [9, 21]. The importance of preparation in classrooms using these pedagogies has led to efforts to complete lecture preparation more seriously [8].

However, it is often the case that students show maladaptive behaviors when it comes to using online systems, such as procrastinating [12] or "gaming the system" [3] to finish the material quickly. This lack of engagement and type of behavior leads to rushed or incomplete work [13] and has detrimental effects on learning [4]. We strive to better understand and characterize these maladaptive behaviors (procrastination and gaming) that undergraduate students display when using an online system for learning in an inverted introductory computer science (CS) course. Our study uses a mixed-method approach involving survey data, system logs, and supplemental interviews. Additionally, in an effort to reduce gaming behavior, we evaluate an intervention designed to reduce gaming behavior.

In this study, we asked the following research questions:

(1) What factors contribute to undergraduate CS1 students engaging in maladaptive behaviors (procrastination and "gaming the system") with online learning systems?
(2) What impact does embedding a lightweight prompt have on reducing gaming/guessing?

We believe that addressing these questions will contribute to characterizing reasons for undesirable student behavior within the context of online exercise systems for learning CS and will lend insight to designing future systems and interventions that encourage increased learning gains. In the next subsections, we'll review the literature on the two negative study behaviours we're investigating: procrastination and gaming the system. Then, we review previous interventions designed to reduce gaming behaviours.

1.1 Procrastination

Academic procrastination is nearly universal as Steel [28] found 80% of students engage in some form of procrastination and that 50% do so chronically. The impact of procrastination is high because it is linked to lower scores [32] and negative study behaviours, such as cramming and late submission [11]. Stewart et al. [29] link effective management of study time with academic success.

The roots of procrastination are complex and not simply a lack of time [27]. Task evasiveness [32] and lack of motivation [23] can be contributors. Furthermore, online environments, like the ones frequently used in inverted courses, are particularly at risk [17, 23].

1.2 Gaming the System

Gaming the system can be defined as "attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material" [5]. Gaming behaviour has been observed as either "help abuse" (e.g., clicking the hint button immediately before trying to solve the problem on their own) or "trial and error" (e.g., guessing on multiple-choice (MC) questions until the system registers the correct answer) [3, 6]. Gaming behavior has detrimental effects on learning outcomes both in the short-term [4] and long-term [22], though some students (e.g., who already understand the material and want to finish a task quickly) may not be very negatively impacted. Whether attributes of the exercises or the students are more causal of gaming is not yet settled. Walonoski et al. concluded that gaming behaviours result from a combination of low prior knowledge and task difficulty [30]. In contrast, Muldner et al. suggest student traits are more important than the difficulty of the task [19]. Most recently, Dang et al. concluded gaming is a learned behavior where student motivation and task complexity interact [7].

1.3 Reducing Gaming Behaviors

Some interventions and designs have been successfully deployed in virtual learning systems to reduce gaming behavior [16]. For example, delaying or limiting the number of hints students may receive in a short amount of time encourages students to try a problem before asking for a hint [1, 10], and random selection of questions after each submission can discourage guessing [24]. Other interventions focus on reflective approaches to reducing gaming, such as using visualization [31], or, as we tried in this study, reminding students of the cost of gaming. This approach of providing just-in-time messaging about gaming has been shown to decrease gaming behaviour in some settings, particularly when combined with self-monitoring activities [2, 25]. However, many interventions discouraging gaming behaviour are ineffective or become so after the students learn how they work [20]. Many seek to treat the effects of gaming behaviours, rather than addressing the root cause, which has been shown to be ineffective [3]. As a result, a large design space of interventions to address gaming behaviours remains open to exploration.

2 METHODS

To pursue our research questions, we conducted a survey and supplemental interviews and implemented an intervention on an online homework system (PCRS) as a randomized A/B test. This study analyses data from an undergraduate, inverted-classroom, CS1 course that uses PCRS at a large, research-intensive North American university. We collected data during the second 12-week-long semester of the academic year.

2.1 CS1 Course Setup

1219 undergraduate students from the inverted CS1 course participated in this study, and 1097 (90%) consented for their data to be used in this research (M = 437, F = 488, *Other* = 172). Of the consenting students, 662 did not have any prior experience in programming, and 511 took CS1 to fulfill their course requirements (M = 0.49, SD = 0.50).

PCRS is an online course content and exercise system for novice programmers. As part of the course, students were expected to complete weekly assigned sections, including a "Prepare" section due on Monday mornings, consisting of videos on topics in CS1 and MC questions, and a "Perform" section due on Friday afternoons, consisting of current content covered previously and included both MC questions and open-ended Python programming questions. Prepare questions were for pre-lecture preparation while Perform questions tested students' knowledge post-lecture.

To receive full marks on these sections, full completion of the Prepare section was required (worth 0.5% of the entire grade each), while both completion and correctness were required in the Perform section (worth 1% of the entire grade each). Students had unlimited attempts in all problems in PCRS for both Prepare and Perform.

2.2 Survey Description and Analyses

We conducted two surveys in the same CS1 course. One of them was released in the middle of the semester after the midterm, and the other was released at the end of the semester. Students were given additional 2% course credit for completing both surveys; those who completed only one of them received 0%. In this study, we only look at data from the mid-semester survey to understand their guessing and gaming behavior using PCRS. The survey asked students to what extent they agreed with the following statements about learning behaviors in PCRS on a seven-point Likert scale (from strongly agree to strongly disagree), as shown in Table 1.

Table 1: Survey questions related to guessing, gaming, and reflective behaviour.

Code	Question
G1	I tend to guess on PCRS, before I even think about the answer.
G2	I repeatedly enter answers on PCRS over and over, until I get it right, without stopping in between to understand.
G3	After I get the correct answer or a working solution on PCRS, I spend 10 seconds or more thinking about what I learned, or why it was correct.

G1 was used as a measurement of guessing behavior in PCRS, whereas G2 was used as a measurement of gaming behavior since it describes systematic trial-and-error, which is one of the common features of gaming behavior [6]. In G3, we asked whether students reflect while solving problems in PCRS. This question tells us if students also display reflective behavior in PCRS, a behavior that is often viewed as contrasting to gaming behavior [6]. It may provide insight into potential relationships between gaming/guessing behavior and reflection.

In addition, the mid-semester survey contained questions about students' language background, fixed and growth mindset, attitude towards PCRS, and opinion of CS, which can be found in Table 2.

Students answered questions from the table (except for Q7) on a Likert scale of 1-7 (strongly agree to strongly disagree). To remove

Table 2: Demographic survey questions.

Code	Question
Q1	I find it challenging to read and listen in English. For example, in reading text and listening to videos in PCRS.
Q2	You can learn new things, but you can't really change your basic CS ability.
Q3	No matter how much intelligence you have, you can always change it quite a bit.
Q4	I don't work on homework until just before it is to be handed in.
Q5/Q6	CS is interesting./CS is fun.
Q7	Many CS1 students have no prior programming experience, while others have some. Rate the amount of prior programming experience that you had before you began the course.

noise and look at the broader sentiment expressed, answers were grouped into three main categories, 1-3 was labelled as *Disagree* (-1), 4 was labelled as *Neutral* (0) and 5-7 was labelled as *Agree* (1).

We also took into account students' self-rated prior knowledge of programming (Q7 in Table 2) on a scale from 1 - no experience to 9 - a lot of experience. These ratings were grouped into three larger categories, such that ratings 1-3 are considered as *low prior* (-1), 4-6 as *some prior* (0), and 7-9 as *high prior* (1).

These questions are a subset of the survey, which were chosen with previous literature in mind on what student traits correlate to guessing, gaming, or reflective behaviors. Q1 helps us gain insight into student understanding of the content and problem. Q2 and Q3 are related to fixed and growth mindset respectively [3]. Q4 references procrastination. Q5 and Q6 are related to the students' personal views of CS [3]. Q7 reports students' prior experience.

2.2.1 Qualitative coding scheme. Additionally, to better understand why students did not start the assigned exercises earlier, we qualitatively coded open-ended responses to two survey questions:

B1: What's the biggest barrier to starting Prepare earlier?
B2: What's the biggest barrier to starting PCRS Perform earlier?

The coding scheme we created primarily uses macrothemes identified in Schraw, Wadkins & Olafson's grounded theory of academic procrastination [26] that are relevant to this context. With consideration to the CS1 course, student population, and the system we worked with, we also made several additions to the themes (italicized) as described in Table 3. Moreover, macrothemes with less than 1% of the student responses corresponding to them were pooled into a new *Miscellaneous* macrotheme. Responses that reported no barriers to starting earlier were coded as *No Barrier*, while responses that yielded no insight (e.g., submitted a blank response or just mentioned "procrastination") were coded as *No Information.*

Students who reported more than one "biggest barrier" received more than one code. E.g., if a student's response to B1 is "other coursework, and very boring", it would be coded as *Other Commitments* and *Self.* Two researchers agreed on the coding scheme after independently looking at 10% of the data and then explaining the reasoning for each code for every survey response to reach a consensus. After agreeing on which codes were appropriate for which types of responses, the researchers independently coded all student responses (including the discussed 10% again) and then calculated inter-rater reliability and Cohen's kappa for each of the two survey questions from the independently coded data.

Table 3: Coding scheme macrothemes.

Macrothemes	Themes
Task	Low background knowledge Task difficulty
Self	Interest Organizational skills *Mental health* *Disability*
Maladaptive aspects of procrastination	Laziness Fear of failure Postponement of work
Deadlines	Lack of due dates for assignments *Prefer to learn material after class* *Due date does not fit well with schedule*
Lack of incentives	Low intrinsic motivation; high self-efficacy Low impact of procrastination on grades
Other commitments	*Other academic or job work beyond CS1* *Leisure time or* *Social and athletic activities*

Miscellaneous	Instructor
	  Clear expectations
	  Well-organized
	  Tests & graded materials
	Unclear direction
	  Ill-defined course material
	  Unclear criteria for grading
	Adaptive Aspects
	  Cognitive efficiency
	  Peak work experience
	External Logistics
	  Internet/Wifi issues
	  Cannot find a quiet space to work

No Barrier	*No barriers to starting earlier*
No Info	*Blank response* *Response too vague to code*

2.3 Supplementary interviews

At the end of the CS1 course, four students were recruited via email for semi-structured interviews regarding multiple aspects of the course. Students were compensated with $7.50 CAD for a 20-30 minute interview, during which they consented to be recorded. For the scope of this study, we only report on students' answers to:

(1) Do you ever guess on PCRS multiple-choice problems?
(2) If so, why do you guess on PCRS?

2.4 Guessing Intervention

We designed and deployed a supportive intervention on an MC question in PCRS, aimed at reminding students not to guess. Students saw the following message after submitting the MC exercise on the first question of a problem set shown in Figure 1: *"It is completely normal to feel the urge to guess on PCRS problems, but remember, guessing on PCRS problems until you get the answer prevents you from learning and can affect you in the final exam. Try to think for at least 30-60 seconds before you choose an answer!"*

Consider this function call:

```
print("this \n is the newline character in Python")
```

What is printed by this function call?

```
☐ "this \n is the newline character in Python"
☐ this \n is the newline character in Python
☐ "this
   is the newline character in Python"
☐ this
   is the newline character in Python
```

Figure 1: The pre-test problem given to the students. A post-test problem of similar difficulty was given after two weeks.

We tested the intervention's impact via randomized A/B testing on two treatments: students who received the intervention ("Message") and those who didn't ("No Message"). Two indicators of guessing behaviour were chosen based on past literature [3]: the number of submissions made by students on the MC problems and the average time (in seconds) between consecutive submissions. The logic behind these heuristics is that a reduction in guessing would lead to a reduction in the number of submissions as well as an increase in the average time taken between submissions.

The pre-test MC problem took place two weeks before the intervention, serving as a baseline for guessing indicators without the intervention, and the post-test was in the same week's problem set as the intervention. Of the students who completed the problem with the intervention in addition to the pre-test and post-test problems (N = 664), 168 students were shown the intervention message and 496 students were not.[1]

3 ANALYSIS

3.1 Coded survey analyses

Figure 2 displays the codes identified from the responses to B1 and B2. Many responses did not contain information that answered the survey question (28.8%) and were coded as "No Information." Inter-rater reliability was high, with 90% agreement rates between the two coded data sets for both questions regarding barriers to starting Perform (Cohen's kappa = 0.88) and Prepare (Cohen's kappa = 0.89) earlier. Above 0.8 is considered "almost perfect" agreement [18].

The majority of the relevant responses were due to students' other academic or work-related commitments (25.3%), which instructors have relatively little control over, especially since very few students reported reasons related to having lack of incentives (1.0%). It suggests students generally viewed the exercises as important and appropriately weighted in the marking scheme. Significantly fewer students reported no barriers to starting earlier (9.3%).

A similar number (9.4%), coded as Self, was related to not being interested in the task, forgetting about the task, or a lack of self-organization. It is unclear why the Self macrotheme comes up more in Prepare (10.6%) than Perform (8.1%). However, forgetfulness (i.e., lack of organization) often came up as a theme underlying the Self macrotheme. This may also be related to the Deadlines code, where the Prepare (11.8%) and Perform (8.6%) exercises had different weekly due dates; Prepare was due on Monday morning, and Perform was due on Friday afternoon. More students may have found the Prepare deadline to be more inconvenient or forgettable (given the weekend) than the Perform deadline.

15.6% of responses for Perform barriers were coded as Task, while it was only 9.1% for Prepare barriers. We believe this difference is because of the differing nature of Prepare and Perform. Prepare exercises consist of only MC questions, while Perform has open-ended Python coding questions (students write code) and sometimes MC questions. Students mentioned that the open-ended coding questions were sometimes very difficult.

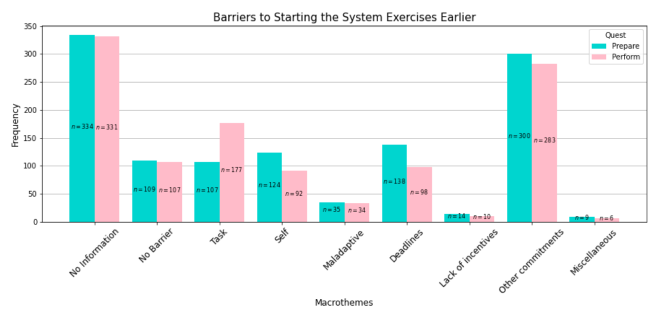

Figure 2: Macrothemes identified in the short answer responses to "What's the biggest barrier to starting PCRS [Prepare/Perform] earlier?"

3.2 Quantitative survey analyses

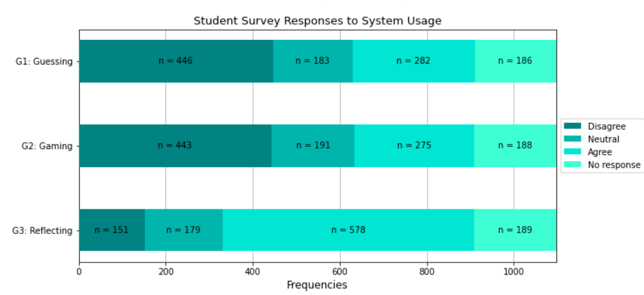

Figure 3: Number of students expressing (dis)agreement with Table 1 survey questions about Guessing, Gaming, and Reflective behaviors.

Figure 3 lists the responses to the questions in Table 1 about gaming and reflective behaviours. Surprisingly, more than half the respondents indicated that they did pause after submitting questions to reflect on the answers, and only about a third of students agreed that they participated in guessing or gaming behaviours.

[1]We did not include students in the analysis who completed only a subset of these three problems.

To further understand the factors that might contribute to gaming and reflective behaviours, we performed a Spearman's Rank Correlation Coefficient between the questions in Table 1 and the demographic questions in Table 2. We tried to minimize the chances of spurious results in our analyses by conducting a literature review first to decide which survey questions to analyse. We omit discussion of the results for question G2 since G1 and G2 behaviors showed very similar correlations with each of the categories. However, G1 and G3 behaviours were often very different; for example, students might have both guessed (at times) but reflected after ($r(906)$ = -0.1, $p < 0.001$).

3.2.1 Fluency in English. We found a correlation between students with comfort with English (Q1) and guessing behavior, $r(908) = 0.28$, $p < 0.001$. A lower proportion (26.4%) of students who report being comfortable with English guess as compared to those who report having difficulty with English (51.8%). A similar trend is visible with not reflecting on the answers to questions after solving them, $r(905) = 0.11$, $p < 0.001$. We learn that students who find learning in English challenging will resort to guessing/gaming more often.

3.2.2 Fixed Vs Growth Mindset. We use Q2 and Q3 to investigate student mindset. The results are aligned, so we focus on Q2. There's a significant positive correlation between Q2 and guessing behavior ($r(906) = .24$, $p < 0.001$). Students reporting a fixed mindset have a greater tendency to guess (42.9%) as compared to those reporting a growth mindset (26.4%), but the relationship with reflection is virtually non-existent ($r(905) = .08$, $p = 0.01$). These results confirm that a student's mindset is related to their guessing behaviour.

3.2.3 Procrastination. We examined Q4, which asked about procrastination to identify relationships between guessing, gaming, or reflecting behavior and students' tendency to procrastinate and work on their homework close to the deadline. We find that 20.2% of the students guessed when they work ahead of the deadline in contrast to the 48.6% working at the deadline, $r(909) = 0.29$, $p < 0.001$. There's almost no correlation between students who reflect and work close to the deadline, $r(906)$ = -0.08, $p = 0.01$.

3.2.4 Appreciation in CS. Since Q6 refers to "fun" which likely encapsulates both interesting and enjoyable, while in Q5 "interesting" may just be interesting, we will be focusing on Q6 only. We see no substantial relationships between students finding CS fun and their tendency to guess ($r(907)$ = -0.1, $p < 0.001$) (a negative correlation means that student's appreciation for the subject led to a reduction in their guessing habits). However, unlike the other relationships, interest in the subject seems to be correlated with increased reflection after solving problems ($r(905) = 0.23$, $p < 0.001$).

3.2.5 Prior Programming Experience. There are no statistically significant relationships between guessing ($p = 0.55$) or reflection ($p = 0.25$) and prior experience, which suggests it is possible that some of the observed gaming behaviour may be "non-harmful" (e.g., experienced students are simply trying to finish the task quickly) [6]. However, the data suggests that the majority of students (65.8%) with no prior experience in computer programming often reflect back on their answers. Similarly, 62.9% of students without prior knowledge do not resort to guessing their answers in PCRS.

3.3 Supplemental Information from Interviews

All four students reported that they guessed in MC questions in circumstances. Three of the four students said they guessed due to time constraints and other activities, which resulted in leaving the assigned PCRS content until very close to the deadline, one mentioning that they ended up guessing because they were "really caught up in [their] other schoolwork" and almost "forgot about it." These insights support the results of the survey, which found that other academic deadlines and forgetfulness were both common reasons for not starting the work earlier. They also link these issues with guessing behavior: students may find themselves in positions where they feel that guessing is the only option to obtain credit.

Of the four, only one student mentioned that they did not often guess and started the exercises early. They stated they guessed when they did not know how to solve the problem but also noted that they reflected after to learn the content. This suggests that guessing and reflection behaviours are not mutually exclusive.

3.4 Submission Analysis

We analyzed the number and timing of submissions to both a pre-test and post-test question to determine if a message reminding students that guessing is detrimental to learning is effective. Two multiple linear regressions were performed. In both cases, the pre-test behaviour (either number of submissions or time between submissions) and the treatment condition were independent variables, and the post-test behaviour was the dependent variable.

Before the analysis, we removed students who spent more than ten minutes between submissions, as we considered these students to have stopped working (the student could have returned to watch the video related to the question again and resubmitted within ten minutes). This resulted in 655 student data points.

Table 4 contains the results of the regression performed on submission counts. The model was unable to account for a significant amount of the variance (only 1.8%, $F(2, 652) = 6.115$, $p = 0.00234$). While the post-test number of submissions are linearly related to the pre-test number of submissions ($\beta = 0.0487$, $p < 0.05$), the treatment had no significant effect ($\beta = -0.1076$, $p = 0.239$).

Table 4: Results of the regression performed on submission counts.

	coef	std err	t	P>|t|
Intercept	1.5287	0.052	29.436	0.000
Pre-test submission counts	0.0487	0.015	3.327	0.001
Message	-0.1076	0.091	-1.179	0.239

Table 5 shows the results of the regression performed on the time between submissions. The results of the regression indicate there was no measurable effect of the pre-test average time between submissions ($\beta = -0.0260$, $p = 0.406$) and the intervention message ($\beta = -2.3580$, $p = 0.067$) on the post-test average time between submissions. The model accounted for 0.6% of variance in the post-test average time between submissions, $F(2, 652) = 1.994$, $p = 0.137$.

Since the intervention message had no significant effect on the post-test number of submissions and the average time between consecutive submissions, we found no evidence to suggest that the intervention message had any impact on preventing students from

Table 5: Results of the regression performed on average time between submissions.

	coef	std err	t	P>|t|
Intercept	4.6503	0.655	7.096	0.000
Pre-test time b/w submissions	-0.0260	0.031	-0.831	0.406
Message	-2.3580	1.286	-1.834	0.067

guessing. It is possible that students read the message but did not put it to practice or perhaps students did not read the message at all, as it was not for any credit. Furthermore, the low variance that could be explained by the model indicates that other undetected factors have a more significant effect. This experiment demonstrates that the obvious approach does not necessarily work as desired, so additional work will need to be done to identify the factors that influence student submission behaviour.

4 DISCUSSION

In this study, we first sought to explore the factors that contribute to undergraduate CS1 students engaging in maladaptive behaviors, specifically "gaming the system" and procrastinating.

We looked at the coded open-ended responses to identify the biggest barriers to working on PCRS exercises sooner. The majority of students identified at least one barrier, with the most common being "other commitments". Other major barriers included the time of the deadline and the difficulty of the exercises. Although instructors cannot typically reduce the many commitments students have, the differences of Prepare and Perform barriers in the Self code (specifically, forgetfulness) and Deadline code (specifically, day and timing) lend insight into how important the deadline time is.

We see from the interviews and survey correlations that the two maladaptive behaviors we focused on may be related for some students. Those who procrastinate on starting often end up gaming the system to finish before the deadline. However, starting the exercises late was not the only contributing factor. The correlations we found suggest that having a growth mindset is correlated with less gaming, which is in line with past work [3], and that students who do not find CS fun (i.e., who may lack motivation) are also at increased risk for gaming. This also corroborates past work [3]. Discomfort with speaking and listening in English was also correlated with gaming behaviors; it makes sense that a student would lean on gaming behavior if they struggle to comprehend the question.

However, a number of our correlation analyses results diverged from the existing literature. In particular, we found no relationship between guessing/gaming and prior experience, which contradicts the past finding that students with low prior knowledge are more likely to game [4]. We speculate this may be because students with high prior knowledge also reported gaming the system, but perhaps for different reasons. Students with high prior knowledge may engage in "non-harmful" guessing due to the knowledge they already have [6] trying to save time, while students with low prior knowledge may game due to not knowing how to solve the problem.

We also found that guessing (a behaviour we labeled maladaptive) and reflection (which is beneficial) are not mutually exclusive. Some students who fall back on guessing when they do not know how to solve the problem still spend time reflecting on the correct answer. We did not detect any correlation between guessing and reflection in the data, and our qualitative data sources include some students who guessed and reflected. This suggests that there are many reasons why students guess (e.g., time constraints) and that not all guessing is necessarily "maladaptive," since it may be, in fact, an adaptive strategy for obtaining hints.

For our second research question — exploring the possible impact of a lightweight prompt to reduce guessing and gaming behavior — we found no particular changes in student behavior. This contradicts earlier work in the area of intelligent tutoring systems that found such messages to be effective [2, 25]. However, it is possible that some students did not read the message, so it is unclear whether the content or the delivery method caused the message to be ineffective. It is also possible that the intervention was ineffective because the factors which contributed to gaming behaviour in our context (often a lack of time or difficulty with language comprehension) were not addressed. This experiment demonstrates that the obvious approach does not necessarily work as desired, so additional work will help identify the factors that influence submission behaviour.

4.1 Threats to Validity

We identified several potential threats to the validity of our study. First, the survey data that establishes who gamed the system or reflected is self-reported. The students reported behavior may not fully align with their actual behavior, as they may have answered as they believed we wanted them to answer. We corroborated this data with a small number of interviews held after the semester had ended; a larger number held closer to the events being investigated (during the term) would have been desirable. Finally, the intervention we deployed was lightweight. We cannot be certain that the students in the treatment group saw or fully read the intervention text.

5 CONCLUSIONS AND FUTURE WORK

Our work presents information specific to guessing, gaming, procrastination, and reflection behaviors in a hybrid CS1 context, helping to characterize students' (mis)use of online systems. Some of our findings are in the same vein as the existing literature (e.g., fixed mindset and increased gaming behaviors [31]), while others suggest differences (e.g., lower prior experience are correlated to higher gaming rates [4]). While several factors we identified are outside the control of instructors, we found that students more easily forget deadlines after the weekend, and we further suggest providing additional support to English language learners.

We also identified a complex relationship between gaming, procrastination, and reflection, with procrastination potentially leading to increased gaming and a lack of relationship to reflection suggesting that not all gaming behaviors are necessarily unhealthy.

The intervention we designed and implemented illustrates that results presented in one context may not always translate to another application environment. Confirmation as to why this intervention failed to work as expected remains a future direction, as does identification of an intervention that does work. Although our intervention did not have a noticeable effect on students' guessing, we believe that it would be worth exploring different delivery methods and message content to reduce gaming behaviors. Additionally, to better understand in more granularity why students in CS1 game, more in-depth interviews could be conducted.

6 ACKNOWLEDGEMENTS

The authors would like to thank members of the Intelligent Adaptive Interventions lab for contributing to the creation of the mid-semester survey, as well as Angela Zavaleta Bernuy for helping to conduct interviews for this study. We would also like to thank all of the CS1 students who consented to having their data be in this study and the instructors for letting us run our intervention in their course. This work was partially funded by Office of Naval Research grant number #N00014-18-1-2755 to Joseph J. Williams.

REFERENCES

[1] 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).

[2] 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.

[3] 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.

[4] Ryan Shaun Baker, Albert T Corbett, Kenneth R Koedinger, and Angela Z Wagner. 2004. Off-task behavior in the cognitive tutor classroom: when students" game the system". In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 383–390. https://doi.org/10.1145/985692.985741

[5] Ryan SJD Baker, AMJA De Carvalho, Jay Raspat, Vincent Aleven, Albert T Corbett, and Kenneth R Koedinger. 2009. Educational software features that encourage and discourage "gaming the system". In *Proceedings of the 14th international conference on artificial intelligence in education*. 475–482.

[6] Mihaela Cocea, Arnon Hershkovitz, and Ryan SJD Baker. 2009. The impact of off-task and gaming behaviors on learning: immediate or aggregate?. In *Proceeding of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*. IOS Press, 507–514.

[7] Steven Dang and Ken Koedinger. 2019. Exploring the Link between Motivations and Gaming. *International Educational Data Mining Society* (2019), 276–281.

[8] Alex Edgcomb, Frank Vahid, Roman Lysecky, and Susan Lysecky. 2017. Getting Students to Earnestly Do Reading, Studying, and Homework in an Introductory Programming Class. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (Seattle, Washington, USA) *(SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 171–176. https://doi.org/10.1145/3017680.3017732

[9] David Gross, Evava S Pietri, Gordon Anderson, Karin Moyano-Camihort, and Mark J Graham. 2015. Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE—Life Sciences Education* 14, 4 (2015), ar36.

[10] 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), 88–95.

[11] Robert M Klassen, Lindsey L Krawchuk, and Sukaina Rajani. 2008. Academic procrastination of undergraduates: Low self-efficacy to self-regulate predicts higher levels of procrastination. *Contemporary Educational Psychology* 33, 4 (2008), 915–931.

[12] Katrin B. Klingsieck, Stefan Fries, Claudia Horz, and Manfred Hofer. 2012. Procrastination in a distance university setting. *Distance Education* 33, 3 (2012), 295–310. https://doi.org/10.1080/01587919.2012.723165 arXiv:https://doi.org/10.1080/01587919.2012.723165

[13] Cornelius J. Koch and Martin Kleinmann. 2002. A stitch in time saves nine: Behavioural decision-making explanations for time management problems. *European Journal of Work and Organizational Psychology* 11, 2 (2002), 199–217. https://doi.org/10.1080/13594320244000120 arXiv:https://doi.org/10.1080/13594320244000120

[14] Kate Lockwood and Rachel Esselstein. 2013. The Inverted Classroom and the CS Curriculum. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education* (Denver, Colorado, USA) *(SIGCSE '13)*. Association for Computing Machinery, New York, NY, USA, 113–118. https://doi.org/10.1145/2445196.2445236

[15] Mary Lou Maher, Celine Latulipe, Heather Lipford, and Audrey Rorrer. 2015. Flipped classroom strategies for CS education. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. 218–223.

[16] John M Malouff and Nicola S Schutte. 2019. The Efficacy of Interventions Aimed at Reducing Procrastination: A Meta-Analysis of Randomized Controlled Trials. *Journal of Counseling & Development* 97, 2 (2019), 117–127.

[17] Barbara Woods McElroy and Bruce H Lubich. 2013. Predictors of course outcomes: Early indicators of delay in online classrooms. *Distance Education* 34, 1 (2013), 84–96.

[18] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica* 22, 3 (2012), 276–282.

[19] 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. https://doi.org/10.1007/978-3-642-13388-6_23

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

[21] Abelardo Pardo, Dragan Gašević, Jelena Jovanovic, Shane Dawson, and Negin Mirriahi. 2018. Exploring student interactions with preparation activities in a flipped classroom experience. *IEEE Transactions on Learning Technologies* 12, 3 (2018), 333–346.

[22] 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*. 177–184.

[23] Glenda C Rakes and Karee E Dunn. 2010. The Impact of Online Graduate Students' Motivation and Self-Regulation on Academic Procrastination. *Journal of Interactive Online Learning* 9, 1 (2010), 78–93.

[24] Joaquin Rodriguez, Gabriele Piccoli, and Marcin Bartosiak. 2019. Nudging the classroom: Designing a socio-technical artifact to reduce academic procrastination. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. 1–10.

[25] Ido Roll, Vincent Aleven, Bruce M McLaren, and Kenneth R Koedinger. 2007. Can Help Seeking Be Tutored? Searching for the Secret Sauce of Metacognitive Tutoring.. In *AIED*, Vol. 2007. 203–10.

[26] Gregory Schraw, Theresa Wadkins, and Lori Olafson. 2007. Doing the things we do: A grounded theory of academic procrastination. *Journal of Educational psychology* 99, 1 (2007), 12–25. https://doi.org/10.1037/0022-0663.99.1.12

[27] Laura J Solomon and Esther D Rothblum. 1984. Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of counseling psychology* 31, 4 (1984), 503–509.

[28] Piers Steel. 2007. The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological bulletin* 133, 1 (2007), 65–94.

[29] Martyn Stewart, Tim Stott, and Anne-Marie Nuttall. 2016. Study goals and procrastination tendencies at different stages of the undergraduate degree. *Studies in Higher Education* 41, 11 (2016), 2028–2043.

[30] Jason A Walonoski and Neil T Heffernan. 2006. Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In *International Conference on Intelligent Tutoring Systems*. Springer, 382–391.

[31] Meng Xia, Yuya Asano, Joseph Jay Williams, Huamin Qu, and Xiaojuan Ma. 2020. Using Information Visualization to Promote Students' Reflection on" Gaming the System" in Online Learning. In *Proceedings of the Seventh ACM Conference on Learning@ Scale*. 37–49. https://doi.org/10.1145/3386527.3405924

[32] M Betul Yilmaz. 2017. The Relation between Academic Procrastination of University Students and Their Assignment and Exam Performances: The Situation in Distance and Face-to-Face Learning Environments. *Journal of Education and Training Studies* 5, 9 (2017), 146–157.