The Ebb and Flow of Student Engagement

Measuring motivation through temporal pattern of self-regulation

Steven C. Dang
Carnegie Mellon University
stevenda@cs.cmu.edu

Kenneth R. Koedinger
Carnegie Mellon University
koedinger@cmu.edu

ABSTRACT

Effective teachers recognize the importance of transitioning students into learning activities for the day and accounting for the natural drift of student attention while creating lesson plans. In this work, we analyze temporal patterns of gaming behaviors during work on an intelligent tutoring system with a broader goal of detecting temporal trends in students' motivation. Findings demonstrate that observing gaming the system behaviors in the near beginning or end of a working session correspond with predictions made by self-regulation theories of ego-depletion and task-switching. Furthermore, analyses provide initial evidence these gaming behaviors are indicative of partial cognitive engagement and session-level influences on student motivation. These findings provide evidence for how temporal fluctuations in students motivations might be inferred through self-regulated behaviors like gaming the system, and how such information could inform better more intelligent tutoring systems that are responsive to cognitive and motivational dynamics during student work.

Keywords

Motivation, Self-Regulation, Measurement, Gaming the system, Ego-depletion, Task-switching, Intelligent tutoring system

1. INTRODUCTION

Many teachers can relate to the struggle of keeping an entire class engaged as the end of the day approaches. Some students may be listening raptly while other have started packing their belongings. Many teachers use class management techniques, such as specific activities in the beginning of class, in anticipation of the difficulties in ramping up the engagement of the entire class [9]. Student motivation appears to vary systematically over the course of a class period. Many good teachers adapt to this reality. It seems appropriate that intelligent tutoring systems should as well.

Student procrastination, the failure to engage in a task in a timely fashion, has a well-established link to student motivations [16]. The nature of the tasks that students have difficulty engaging themselves in can be revealing about their individual goals [15], their perceptions of the value of the task [5], and their beliefs about their abilities to complete the task [21]. Similarly, the context of what drives students to quit can be equally telling about the same facets of student motivation [20].

Measures of quitting and procrastination leverage the easily observable dichotomy of student engagement, but are there other within-task student behaviors that might similarly indicate motivation? Quitting and procrastination are evidence of students' failure to exercise their self-regulation. In these moments, students are failing to direct their attention towards a less desirable but beneficial learning task, and instead opting to engage in more desirable non-learning tasks. Applying this self-regulation lens, it may be possible to understand student motivation by identifying and analyzing other observable moments during student work where students engage in less desirable behaviors for learning.

1.1 Temporal Dynamics of Self-Regulation

Self-regulation is the capacity to control or direct one's attention, thoughts, emotions, and actions [27]. One of the leading models of self-regulation poses the construct as a reward-based decision-making process [2]. In this model, self-regulation is treated as a series of decisions that seeks to optimize some expected value based on anticipated rewards and costs. Motivation is defined as "the orienting and invigorating impact, on both behavior and cognition, of prospective reward" [2]. Through this theoretical lens, self-regulation decisions are a reflection of student's motivation.

For instance, solving an extra credit problem on the homework may likely push the student's grade from a B to an A for the year. However, the problem will likely take an hour to solve and the student may have to skip soccer practice to find time to complete the problem. Observing the student's choices and behaviors in these critical moments of self-regulation can reveal student's underlying motivation. Prior models of self-regulated learning behavior have focused on the cognitive facets of a given task: its difficulty level [4,7], its domain topic[10], its time cost[6], and its expected value to the student[19]. However, research on self-regulation point to temporal factors that influence decision making.

Task switching research indicates that the exercise of self-regulation imposes a cognitive cost. Once an individual chooses to engage in a task, they do not always appear to be applying themselves with full effort [8]. Additionally, when a person is forced to change tasks rapidly, they are not able to perform at the same level as those given more consolidated spans of time to perform on the same task [11]. These studies imply that students are likely to perform at a reduced capacity when initially beginning work to perform on a task upon initially beginning work,

Ego-depletion models of self-regulation posit that the ability to regulate attention over time may tend to deplete as some time-driven function of an internal and limited resource [23]. Thus,

Steven Dang and Kenneth Koedinger "The Ebb and Flow of Student Engagement: Measuring motivation through temporal pattern of self-regulation" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 61 - 68

motivation may also tend to wane over time leading to an eventual failure to self-regulate.

In this work, we seek to investigate whether these temporal properties of self-regulation are evident in the prevalence of student's failures to self-regulate.

2. Related Works

Measuring self-regulation related constructs is not a new concept in the intelligent tutoring system literature. Prior work has developed a range of models for detecting self-regulation related behaviors.

2.1 Off-task Detection

Some of the earliest work in this space identified off-task student behaviors by identifying large gaps of time between interactions in the log data of student interactions [26]. Inferences on student skill improvement, in addition to whether the students asked for help or attempted a problem correctly/incorrectly following a long gap between interactions determined whether students were off-task while idle.

[18] developed models of mind-wandering, when students' attention and thoughts move off-task, which enabled detection of off-task behavior over much shorter time spans. These models leveraged information from videos and human labels of short time segments to train a supervised model to classify when mind wandering occurs. The features fed into the model included a range of low-level image processing features, facial features, inferred emotions, and temporal features that describe the dynamics of facial features and emotions during a short time interval. [17] extended this work given user self-reports of mind-wandering and included body position information.

2.2 Persistence and Quitting

[4] developed a model of student persistence by analyzing patterns of behavior that included observed student actions contingent on properties of the problems being worked and the student's skill on those problems. In this work, two types of students emerged, where the authors posited that trait level differences in students' capacity for sustained attention lead to differences in learning strategies and persistence during problem solving.

[3] designed a game-based measure of trait level persistence and validated the measure against other existing survey and standard psychometric behavioral tasks. The measure looked at average time on unsolved versus solved problems given a wide range of difficulty levels.

In [7], the authors built models of quitting an educational game. They leverage many features including features of each level of the game, the current state of game progress of the student, and the time in the current level. The final model that emerged from the supervised machine learning process were focused around actions of the student and the state of progress and counts of actions at each level across and within attempts at the level, thus not including any of the limited temporal features given at model training time.

[10] attempted to predict when students would quit reading a given passage. In this work, the authors used semantic features of the reading passages, the recent context of what passage is being read, which passages have been read recently, and both current page and total reading time. Total reading time, a similar proxy to ego-depletion, was found to be a significant contributor to models of quitting with respect to the first page of a passage. The authors also implicitly investigated the role of task switching by predicting quitting at the beginning of a new passage compared to some other new page within a passage. While some of the data supports a differential impact of task switching and time on quitting, the authors do not explicitly explore how quitting behaviors vary over time.

2.3 Gaming the System

With intelligent tutoring systems that provide scaffolding supports through progressively informative hints and feedback, another behavior tends to arise called "gaming the system" [24]. These behaviors have been identified using information about a series of recent actions such as time spent or the number of recent hint requests and errors, and the characteristics of the problems worked, such as problem section and difficulty in those interactions [12]. Extensive work has attempted to determine what drives gaming behaviors. While some initial work determined that problem context better explained gaming behaviors over trait-like individual propensities to game [25], later work presented the opposite result using a different intelligent tutoring system [22]. A large multi-environment analysis was conducted that compared the types of gaming behaviors observed across urban, suburban, and rural contexts using three different intelligent tutoring systems [13]. The study found that across tutoring environments, students displayed different predominant gaming behaviors, which implies that the lure of certain types of gaming may be different given tutoring environment or problem-type affordances. Similarly, within tutoring environments, students from areas of different population density (eg: rural versus urban) display different predominant patterns of gaming. These differences point to how variation in work environment may have differential anticipated costs to gaming, while the variation within environment but across geographic regions point to possible cultural and thus motivational differences.

2.4 Research Questions

Prior work has developed extensive models of self-regulation behaviors that demonstrate the importance of cognitive, contextual factors, and local temporal factors for influencing student's self-regulation decisions. However, these models have not investigated how self-regulation behaviors might vary systematically over time and how such trends relate to student learning. In this work, we seek to investigate whether the within-session temporal properties of self-regulation are evident in student behaviors and whether these temporal trends are predictive of similar negative impacts on student learning.

Models of the cognitive cost of task switching imply that self-regulation related behaviors such as gaming the system are more likely to occur in the beginning of a work session. Similarly models ego-depletion imply that self-regulation related behaviors such as gaming are more likely to occur after students have been working for some time. We propose to investigate whether models of task-switching and ego-depletion are evident in some changes over time of the probability of gaming the system, a behavioral instance of self-regulation. We then investigate whether lower cognitive engagement as predicted by task-switching theory co-occurs with gaming the system. We follow this with an analysis to determine if failures in self-regulation during critical time periods are indicative of session-level motivation.

3. The Dataset

We utilize an observational dataset [1] including 214 students across 22 classrooms using the Carnegie Learning Cognitive Tutor (CT) in Pre-Algebra, Algebra 1, and Geometry. The tutor was used approximately two class-periods per week for a full school year. The dataset includes over 2.3M user transactions covering 55, 33, and 26 curricular units divided into 173, 98, and 44 sections across the three courses respectively.

The CT leverages computational cognitive models to provide adaptive problem selection and hint support and correctness feedback to the students. Problems are broken down into a multi-step process, which allows the system to identify independent skills and trace skill improvement over a fine-grained skill model of the domain. On each step, the system is able to provide multiple levels of hint support, with the final level containing the answer to the problem step. The system logs all interactions with the system including problem attempts, hint requests, response accuracy, and problem step time. In this study, transactions for all students over the course of an entire academic year are utilized.

3.1 Measuring Gaming the System

We leverage the model of gaming developed by [14] to annotate transactions as gamed. This model identifies a set of patterns of transactions that experts identify as gamed patterns. A student is determined to be gaming at some time if a series of transactions matches an identified transaction. For instance, a common pattern is when students enter the same or a very similar answer into multiple places without answering correctly, effectively guessing where a calculation result belongs without understanding the organization of the problem. Another common pattern is when students ask for help without taking much time to consider the problem, followed shortly after by an incorrect input. In this case, the student appears to be using the help facility to get an answer but is not taking enough time to use the information provided to derive an answer. The dataset consists of 4.1% of transactions as being labeled as part of a gaming behavior, where the majority of students are labeled as gaming between 3.2 to 4.3% of all observed transactions.

3.2 Aligning Session Time

The data described above only includes transactions after eliminating certain transactions from the original dataset. In order to see temporal patterns, data was excluded from short sessions with length in the bottom 5% of all student session lengths, which was determined to be about 5 minutes. The resulting observed student sessions ranged from 5 minutes to 58 minutes, with a median length of 32 minutes.

One difficulty in measuring ego-depletion with observational data is in controlling for differences in the depleting effects of context. In ego-depletion studies, the task is controlled for and thus can be ruled out to explain observed differences in behavior. In intelligent tutoring contexts. The adaptive instruction will provide variably challenging and types of content and may differentially deplete students across the experiences within the same period of time. To overcome this issue, we leverage the insight that when two students begin working, they might be in similar states relative to their internal thresholds for self-regulation. We also assume that when two students stop working, they are in comparable states. If these two students stop working at different times, it implies similar start and finish attention states, but different depleting effects of context that were experienced over time. In order to account for these differences in uncontrolled contextual factors, we created an additional time measure that aligned individual student transactions within sessions by the percentage of the session time that has elapsed. This alignment facilitates comparison of transactions relative to the start and end of a session, scaled to the session length.

4. Modeling the Effect of Time

Theories of self-regulation imply different models of the effect of time on self-regulation. Attentional shift models posit a cognitive cost of task switching. These costs may cause some tasks to seem more difficult near the beginning of a session. Ego-depletion models imply a reduction of a limited capacity to self-regulation resource over time. These models suggest students may eventually find it difficult to continue in a task and signs of fatigue, such as gaming, may be revealed by an increased tendency to engage in gaming behaviors before finishing working. To test these model implications, we compare five random effect logistic regression models to determine how self-regulation may vary over the course of a session.

We introduce M1 as the baseline model for comparison to determine if any temporal models are significantly more predictive than current best practices as suggested by prior gaming research. This model includes random effects for both student and curricular section to control for the previously established impacts of student and context on student's tendency to game. The remaining four subsequent models similarly control for student and contextual factors while introducing additional factors representing temporal effects.

To define the remaining four models, time is represented along two dimensions. In the first dimension, time is represented as either time elapsed since the student began working or percentage of total working time elapsed, as described section 3.2. Time elapsed models represent the default model informed by both ego-depletion and task switching theories. Percentage of time elapsed models test the hypothesis that such a representation better captures motivation as temporally relative to the most informative moments of student behavior. In the second dimension, time is represented linearly or quadratically. Linear models allow only one main temporal effect to be captured by the model, either a constant increase or decrease in motivation over the course of a session. Quadratic models can capture different effects at the start and end of the session that differ from each other and the middle of the session. All temporal variables are normalized over the full dataset for model interpretation.

M4.1: ***Baseline*** – Baseline model for comparison controlling for differences in student's tendency to game and contextual factors across curricular sections, such as average difficulty, that influence gaming.

Eq 4.1: Gaming ~ (1|Student) + (1|Section)

M4.2: ***Linear Session Time*** – Extending the baseline model M4.1 by adding a linear term for time-elapsed since the student has begun working

Eq 4.2: Gaming ~ time-elapsed + M4.1

M4.3: ***Linear Percent Time*** – Extending the baseline model M4.1 by adding a linear term for proportion of session time elapsed as a percentage of total time observed working.

Eq 4.0: Gaming ~ pct-time-elapsed + M4.1

M4.4: ***Quadratic Session Time*** – This model extends model M4.2 by adding a quadratic term

Eq 4.0: Gaming ~ time-elapsed$_2$ + M4.2

M4.5: ***Quadratic Percent Session Time*** – In addition to the random effects in Eq 4.1, this model tests the hypothesis that students self-regulation resources are

Eq 4.0: Gaming ~ pct-time-elapsed$_2$ + M4.3

4.1 Comparing Models

Table 1. Comparing models of temporal trajectories of student gaming behaviors

Model	BIC	AIC	LogLik
M4.1	434741	434703	-217348
M4.2	434682	434632	-217312
M4.3	434668	434619	-217305
M4.4	434454	434392	-217191
M4.5	454503	434441	-217215

The results of fitting each of the five models are shown in Table 1, including model performance as assessed by AIC, BIC, and log-likelihood. In general, all models with temporal factors outperform the baseline model, M4.1. This implies that temporal information has a significant effect on student's self-regulation behaviors. Additionally, both quadratic models, M4.4 and M4.5, are significantly better than their linear counterparts (Chisq = 179 ($p<0.001$) for M4.2 vs M4.4, and Chisq = 242 ($p<0.001$) for M4.3 vs M4.5). Likewise, M4.4 and M4.5 are significantly better than baseline with Chisq = 315 ($p<0.001$) and Chisq = 266 ($p<0.001$) respectively.

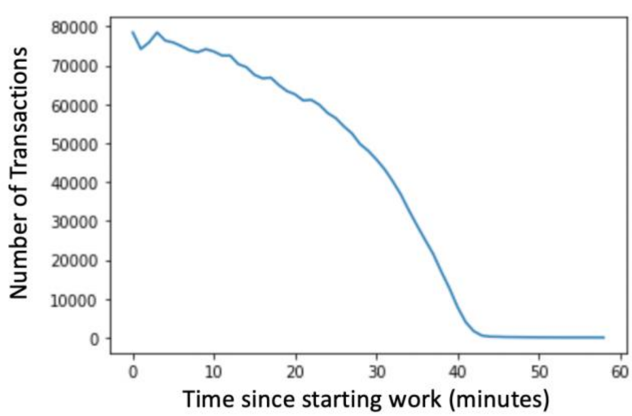

Figure 1. Number of observations over time in session

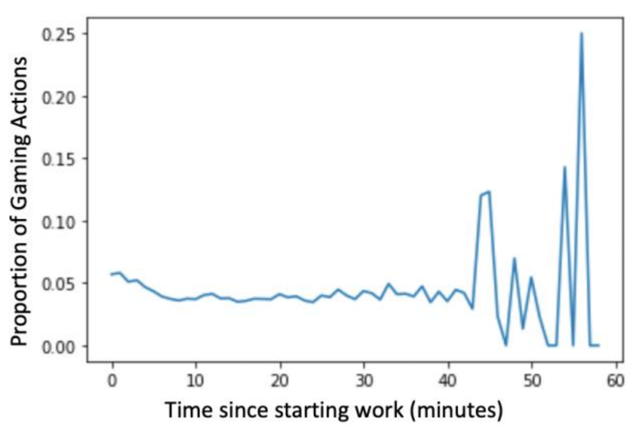

Figure 2. Proportion of Gaming Actions by minute

Exploratory plots of proportion of gaming the system transactions over the session support these interpretations. Figure 2 and 4 plot the proportion of transactions identified as gaming the system behaviors across the session over minutes passed or proportion of total session time respectively. As expected from the quadratic fit models, each figure shows an increased proportion of gaming behaviors near the start and end of sessions.

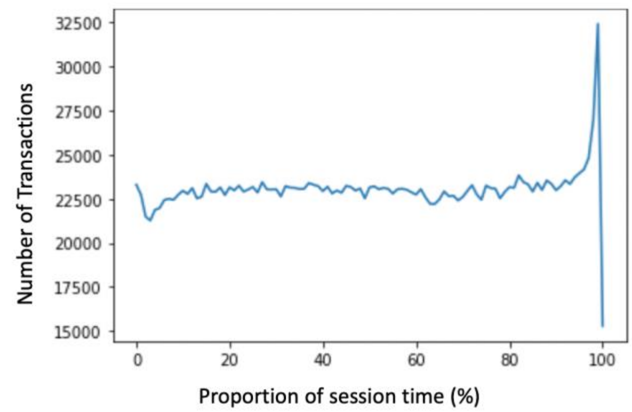

Figure 3. Number of observations over proportion of session time

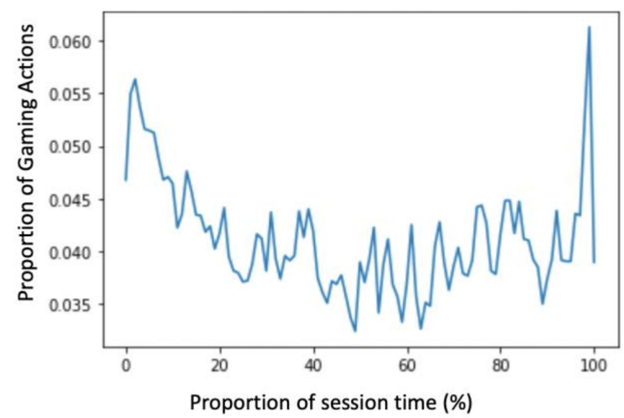

Figure 4. Proportion of Gaming Actions by Proportion of Session Time Passed

A closer look at the data in Figure 1 reveals that there is a large student participation drop-off near the 43 minute mark. While whole class sessions seem to regularly measure about 60 minutes, students' login and logout times are quite staggered such that 99% of observed student sessions are less than 43 minutes in length. Only 82 out of more than 9800 sessions are observed where students worked continuously for between 43 and 60 minutes. Furthermore, analyzing gaming averaged over each minute of the hour, Figure 2, shows that this dramatic reduction in data is associated with very large and volatile estimates of average students gaming per unit time. Because of the low amount of data observed in the last 17 minutes of sessions longer than 43 minutes, it is hard to draw stronger conclusions about whether students are much more likely to display gaming behaviors if they are able to stay on task longer than 43 minutes, or if the volatility is due to random sampling bias.

A closer inspection of data in Figure 3 also shows some peculiar variability in data at the start and end of sessions. Because session time is divided evenly across the proportion of sessions, there is no a-priori reason to believe students have more or less frequent transactions at any time in the session. The small decrease in quantity of transactions near the start of sessions implies students take longer on average to complete actions near the start of work. The large spike of activity near the end implies students are taking less time per action shortly before stopping work. In both cases, the data sparsity issue seen in Figure 1 is not likely driving the changes in proportion of gaming seen in Figure 4. The small decrease in activity near the start is associated with the start of a broader downward trend in proportion of gaming behaviors that continues even after activity frequency flattens. The sudden increased frequency of transactions near the end of sessions is associated with a comparable spike in prevalence of gaming the system behaviors. However, because some gaming behaviors are defined by rapid actions in succession, this relationship is expected.

Taking the model comparisons and exploratory data analysis together, this evidence supports the interpretation that there are non-monotonic differences in gaming the system behaviors between the start, middle, and end of sessions.

Table 2. Model coefficients for M4.4 and M4.5

Term	M4.4 - ß	Term	M4.5 - ß
Intercept	-4.215	Intercept	-4.217
Percent time elapsed	-0.265	Time elapsed	-0.283
(Percent time elapsed)^2	0.231	(Time elapsed)^2	0.252

Comparing the two quadratic models, M4.4 is the best fit model by all 3 measures, BIC, AIC, and Log Likelihood. The model details can be seen in Table 2. The variance in gaming attributable to curricular sections is 0.87. This translates to average gaming attributable to tutor context level factors to range between 0.23% and 8.4% for 95% of sections. The variance attributable to students is much smaller, 0.088. This translates to average gaming attributable to trait-level student factors to range between 0.82% to 2.57%. An inspection of the model coefficients shows that the model predicts the average gaming level at the start of a session, P(gaming|t=0), is 4.1%. Average gaming at the end of the session, P(gaming|t=60 minute), is 18.7%. The quadratic model reaches a minimum observed gaming of 1.3% at 23 minutes into the session.

An 18.7% average probability of gaming after working for 60 minutes appears to be very high given that gaming only occurs overall in the dataset in about 4.5% of all actions. As discussed in the previous exploratory data analysis, the very high gaming proportion observed in the last 17 minutes of sessions is potentially related to the increased volatility created from estimates drawn from small amounts of data. These estimates spike upwards as high as 25%, which corresponds with the dramatic difference between start and end gaming predicted by M4.4. Therefore, the model is reflecting this same artifact of the data.

Inspecting M4.5, the model predicts that gaming is more likely in the start and end of the session. The average probability of gaming decreases to 1.35% by the time the student has worked 67% of the total time. According to the model, we are 3.34 times more likely to observe students game the system near the start of work than near their peak level of focus. Likewise, it is 1.32 times more likely to observe gaming the system in the moments shortly before students stop work. This model appears to make less dramatic predictions that are more inline with expectations based on overall average frequencies of gaming while not reflecting the same uncertainties as M4.4.

These results support the hypothesis that self-regulation processes have an impact on the average occurrence of gaming the system behaviors over the course of a work session. Students in this data appear to experience decreased motivation near the start of work as would be predicted by the cognitive costs of task switching. Likewise, students appear to show some decreased motivation before stopping work as predicted by ego-depletion theories.

5. Leveraging Gaming for Prediction

The previous analysis has demonstrated that observing instances of weaker self-regulation, such as gaming the system behaviors, support a view of student's dynamic self-regulation capacities over time as predicted by ego-depletion and task-switching theories. This raises the natural question of exactly what observing such lapses in self-regulation implies about a student's internal capacities.

5.1 Gaming Indicates Cognitive Effort

If students are not observed to game the system early in a session, we expect that student motivation is likely higher around this time despite the brief slightly negative impact of task switching. This greater motivation allows students to bring greater cognitive resources to the work relative to days when gaming is observed near the start. When comparing assistance rates in the beginning of a session, the proportion of questions either answered incorrectly or with a request for help on first attempt, a student who is more cognitively engaged should be less likely to make errors or ask for help. Likewise, similar patterns should be associated with assistance rates near the end of students work.

We compared the assistance rates for sessions where a student is observed gaming in the first 10% of the session time (the first 3 minutes for the median session) to assistance rates where no gaming is observed in the first 10% of the session time. To calculate the assistance rate, the raw student transactions are aggregated by problem-step. The outcome of each step is determined by the first attempt at the step. The step is labeled as gaming the system if any of the aggregated transactions are labeled as gaming. Because patterns of gaming generally involve either incorrect or help-seeking behaviors, steps that were labeled as gaming the system are removed before calculating the proportion of incorrect and help-request steps to overall steps observed in the portion of the session.

The assistance rates in the start of sessions are shown in Figure 5 and were found to be significantly lower ($t=-15.22$, $p < 0.001$). The average assistance rate where gaming is observed is 30% (sd=25) while the average rate when gaming is not observed is 21% (sd=26). Similarly, Figure 6 shows boxplots for assistance rates in the last 10% of sessions. Rates were found to be to be significantly lower ($t=-11.6$, $p<0.001$) with the average session where gaming is observed having a rate of 25.3% (sd=22) compared to the average non-gaming session having a rate of 18.6% (sd=24).

This simple analysis does not take into account factors such as question difficulty. It is possible that if students are working on difficult content near the start, then they are more likely to make errors and request hints. It also implies that more challenging material may impact how students evaluate the likelihood of prospective reward given their perceived abilities. This may lead students to believe that applying effort is unlikely to result in experiencing the reward or attempting to apply effort may have greater depleting effects that impact future actions. In either case, it is possible that more challenging material instead of task-switching or ego-depletion explains the relationship between increased assistance score and gaming behaviors near the start and end of work. However, these tests do provide compelling evidence for a possible impact of decreased cognitive engagement on some practice opportunities that can inform future modeling work.

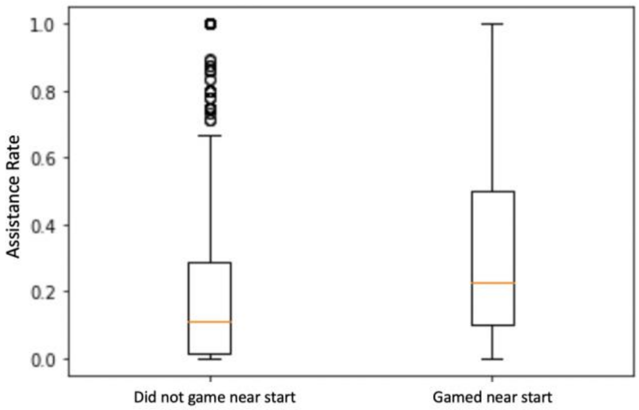

Figure 5. Comparing assistance rate at the start of sessions

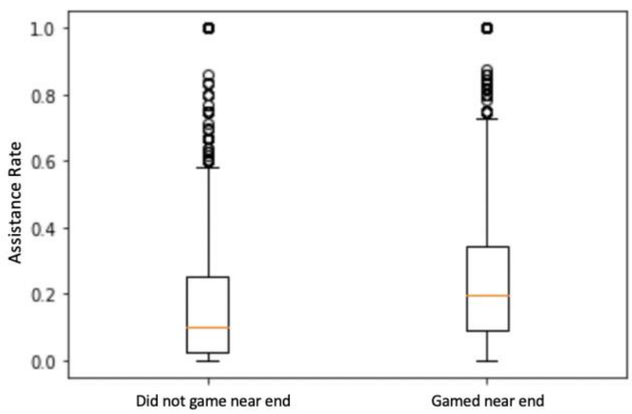

Figure 6. Comparing assistance rate at the end of sessions

5.2 Gaming Indicates Motivation Levels

Student's day-to-day average motivation level is affected by factors in the school, in the classroom, and in the student's life more broadly. A death in the family, a fight with a significant other, or a poor grade in another class might be weighing on a student's mind while that begin working. These factors may have a negative effect on student's ability to self-regulate throughout the entire session. If this is the case, these factors will act in combination with the additional impacts of task-switching or ego-depletion at the start and end of the session to impact a student's capacity to self-regulate. Thus, observing gaming the system behaviors at the start or end of a session may also be informative about a student's more general motivational level. In this section, we analyze gaming behaviors throughout the session using information about whether students gamed at the beginning or end of a session to improve predictions of gaming in the rest of the session.

Gaming at the start and end are defined the same as in the previous section. In the data, 29.7% of sessions are observed with gaming at the start while 32.0% of sessions have gaming at the end. Together 49.9% of sessions have instances of gaming the system in the start or end, while only 11.8% of sessions are observed with gaming in the start and end of the session. While gaming near the start or end might be indicative of session level motivational impacts, in this analysis we test whether seeing any gaming at the start or end is sufficiently informative or if start and end are differently informative.

To perform this analysis, we use the best model from the Section 4 analysis, M4.4 the quadratic percent-time-elapsed model. This model will control for the variance due to student and tutor contextual factors, removing concerns about confounds such as gaming at the start may be due to generally more difficult material that makes gaming more likely throughout the session. We compare models that add main effects for whether gaming was observed at the start or at the end as well as linear and quadratic interaction effects. The models are elaborated as follows:

M5.1: Baseline Quadratic Model – the baseline model from Section 4 analysis for comparison.

Eq 5.1: Gaming ~ pct_elapsed + pct_elapsed$_2$ + (1|Stu) + (1|Sect)

M5.2: Gaming at start/end main effect – M5.1 with a binary indicator variable of whether gaming is observed near the beginning of the session and a binary indicator variable of whether gaming is observed near the end of the session

Eq 5.2: Gaming ~ M5.1 + g_start + g_end

M5.3: Combined Gaming at start or end main effect – M5.1 with a binary indicator of whether gaming is observed at either the beginning or the end of the session

Eq 5.3: Gaming ~ M5.1 + g_start_end

M5.4: Gaming at start and end with linear interactions – M5.4 elaborates on top of M5.2 adding linear interactions with time.

Eq 5.4: Gaming ~ M5.2 + g_start:pct_elpsed + g_end:pct_elpsed

M5.5: Gaming at start and end with quadratic interactions – M5 elaborates on top of M5.4 adding interactions with quadric time terms.

Eq 5.5: Gaming ~ M5.4 + g_start:pct_elpsed$_2$ + g_end:pct_elpsed$_2$

Comparing M5.2 and M5.3, we see that including separate main effects for gaming at the start and gaming at the end leads to better models rather than combining the information into a single indicator of whether there were any self-regulation failures at either the start of the end of the session. This particular result is worth further investigation to understand how and why self-regulation at the start of a session is differently indicative of student motivation levels compared to gaming at the end of the session.

The results in Table 3 indicate the best fit model is M5.5, the model with start/end gaming information and interactions with linear and quadratic terms. This model is significantly different from the baseline quadratic model (Chisq=49.42, p<0.001) and establishes the informativeness of gaming in the start or end of a session on student's motivation levels through the time that students are working. Details about the model are given in table 4.

Table 3. Comparing Gaming Predictions using Start/End Gaming

Model	AIC	BIC	LogLik
M5.1	434441	434503	-217295
M5.2	422316	422403	-211151
M5.3	427322	427397	-213655
M5.4	419913	420045	-209958
M5.5	418266	418402	-209122

The variance accounted for by section and student level random effects are reduced in comparison to the baseline quadratic model reported in Section 4. The variance attributable to student factors was found to be 0.0789, which translates to an average gaming level of 0.64% to 1.91% for 95% of students. The variance attributable to section level factors was found to be 0.7527, which translates to an average gaming frequency of 0.20% to. 5.79% for 95% of sections. This implies that a significant fraction of observations of gaming that were previously explained by section-level factors appears to now be explained by motivational factors indicated by gaming at the start or end of a session.

Table 4: Coefficients for start/end gaming with quadratic interaction terms

Term	ß
Intercept	-4.489
Percent time elapsed	1.129
(Percent time elapsed) ^{2}	-1.251
Gamed at start	0.301
(Gamed at start) * Percent time elapsed)	-1.480
(Gamed at start) * Percent time elapsed) ^{2}	1.170
Gamed at end	0.356
(Gamed at end) * Percent time elapsed)	-0.490
(Gamed at end) * Percent time elapsed) ^{2}	0.900

Table 5 contains the predicted gaming attributable to the main effect terms in model M5.5. The first column describes average predicted gaming at the start of work. The third column describe average predicted gaming at the end of work. Because the model includes quadratic terms, the second column is included to describe the optimum (minimum or maximum) probability of gaming throughout the session. The fourth column describes the odds ratio the chance of gaming at the start relative to the optimum point. The fifth column describes the odds ratio of the chance of gaming at the end compared to gaming at the optimum point. The complexity of the model can make it challenging to interpret, however there are some important trends indicated by the model. If gaming is observed only in the start of a session, gaming is most likely to occur similarly near the start and will reduce over the course of the session as evidenced by the odds of gaming being greatest at the start relative to the end. Likewise, observing gaming only at the end of the session implies that students tend to be well regulated near the beginning of the session and will appear to fatigue over the session until near the end where the odds fall slightly. When students are not observed gaming at the start or end, there is a corresponding low probability of observing gaming near the start and end. However, over the course of the session, the model predicts that these students become more likely to have slightly reduced motivation until the latter half of the session where attention on the time pressure of the end of class might increase motivation through the end of class. In the limited sessions where students are observed gaming at the start and end, the model predicts a much greater propensity to game throughout, with a 53% chance in the start and a 5% chance near the end.

Table 5: P(Gaming) Main effect predictions given start/end gaming observations

Context	Game (t=0)	Game (t=opt)	Game (t=100)	Start Odds	End Odds
No Gaming start or end	0.35%	1.43%	0.21%	0.24	0.15
Start Gaming	2.14%	2.14%	0.66%	1	0.31
End Gaming	0.18%	2.10%	1.71%	0.086	0.81
Start + End Gaming	53.1%	1.72%	5.1%	30.9	2.98

Taken together, these results support the conclusion that gaming at the start and end of work are indicative of session-level motivational factors influencing student behavior. It also provides initial evidence for separable constructs indicated by gaming at the start versus at the end. Each of these constructs appears to have different degrees of impact on underlying student motivation factors and the resulting decision processes that lead to observable behaviors.

6. Discussion

We have treated gaming the system behaviors as indicators of student's self-regulation. Task switching and ego-depletion theories of self-regulation predict a temporal pattern to student's abilities to self-regulate over the course of a class period. Predictive model comparisons are supportive of the hypothesis that both task switching and ego-depletion are evident in the patterns of student behaviors over each class session. Further analysis indicates that observations of self-regulation behaviors in the start and end of class might be indicative of both temporally immediate degrees of cognitive engagement as well as more session or day-level influences on motivation.

Open questions remain about how student models could operationalize task switching or ego-depletion. The work presented, uses information about the full student session to represent time, though such information is not available to real-time models. This raises the question of how should student's prior behaviors inform a predictive models of student ability to task switch or ego deplete? To what degree do students display consistency in their ability to task switch quickly or manage ego-depletion more effectively across sessions? Over the course of months or years? To what degree are these capacities independent or can correlations be attributable to other latent motivational causes?

We believe these findings highlight the importance of leveraging student models that incorporate temporal variables in the design of learning activities. Problem selection algorithms may want to be biased for lower challenge or greater interest to overcome negative effects of task switching. Similarly, activities may want to incorporate changes in the rhythm of the activity in order to periodically re-engage student attention as it wains over time. This work exposes an unexplored design space for how educational activities could incorporate temporal effects of student motivation to better enable student learning.

In this work, we introduce the importance of considering temporal factors in addition to content-related cognitive factors to more effectively support students' motivational trajectories within a work session. These findings extend the rich body of work on modeling student motivational and cognitive processes with self-regulated learning. Students are not machines, and they do not always jump immediately into tasks full throttle or have the endurance to work as long as they are asked. Hopefully, a future that recognizes these dynamics can take intelligent tutoring systems one step closer to emulating the capabilities of effective teachers.

7. ACKNOWLEDGMENTS

The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150008 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

8. REFERENCES

[1] Bernacki, M.L. and Ritter, S. Dataset 613 in DataShop. Retrieved from https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=613.

[2] Botvinick, M. and Braver, T. 2015. Motivation and Cognitive Control: From Behavior to Neural Mechanism. Annual Review of Psychology. 66, 1 (Jan. 2015), 83–113.

[3] Cho, K. et al. 2015. Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. Computers & Education. 86, c (Aug. 2015), 224–235.

[4] Dumdumaya, C.E. et al. 2018. Identifying Students' Persistence Profiles in Problem Solving Task. In Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP'18). 281–286.

[5] Eccles, J.S. 2005. Subjective task value and the Eccles et al. model of achievement-related choices. Handbook of Competence and Motivation. 105–121.

[6] Flake, J.K. et al. 2015. Measuring cost: The forgotten component of expectancy-value theory. Contemporary Educational Psychology. 41, C (Apr. 2015), 232–244.

[7] Karumbaiah, S. and Shute, V. 2018. Predicting Quitting in Students Playing a Learning Game. *In Proceedings of the 11th International Conference on Educational Data Mining* (Jul. 2018).

[8] Kool, W. et al. Decision making and the avoidance of cognitive demand. Journal of Experimental Psychology: General. 139, 4, 665–682.

[9] Marzano, R.J. et al. 2003. Classroom management that works: Research-based strategies for every teacher.

[10] Mills, C. et al. 2014. To Quit or Not to Quit: Predicting Future Behavioral Disengagement from Reading Patterns. In *Proceedings of the International Conference on Intelligent Tutoring Systems.* Springer, Cham. 19–28.

[11] Nieuwenhuis, S. and Monsell, S. 2002. Residual costs in task switching: Testing the failure-to-engage hypothesis. Psychonomic Bulletin Review. 9, 1 (Mar. 2002), 86–92.

[12] Paquette, L. and Baker, R.S. 2017. Variations of Gaming Behaviors Across Populations of Students and Across Learning Environments. In *Proceedings of the 18th International Conference on Artificial Intelligence in Education* (Cham, Jun. 2017), 274–286.

[13] Paquette, L. et al. 2015. Cross-System Transfer of Machine Learned and Knowledge Engineered Models of Gaming the System. In *Proceedings for the International Conference on User Modeling, Adaptation, and Personalization* (Apr. 2015), 183–194.

[14] Paquette, L. and de Carvalho, A.M.J.A. 2014. Towards Understanding Expert Coding of Student Disengagement in Online Learning. In *Proceedings for the Annual Meeting of the Cognitive Science Society* (2014), 1126–1131.

[15] Scher, S.J. and Osterman, N.M. 2002. Procrastination, conscientiousness, anxiety, and goals : Exploring the measurement and correlates of procrastination among school-aged children. Psychology in the Schools. 39, 4 (May 2002), 385–398.

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

[17] Stewart, A. et al. 2017. Face Forward: Detecting Mind Wandering from Video During Narrative Film Comprehension. In *International Conference on Artificial Intelligence in Education* (Cham, Jun. 2017), 359–370.

[18] Stewart, A. et al. 2016. Where's Your Mind At?: Video-Based Mind Wandering Detection During Film Viewing. In *Proceedings for the 2016 Conference on User Modeling, Adaptation, and Personalization* (New York, New York, USA, Jul. 2016), 295–296.

[19] Wigfield, A. and Eccles, J.S. 2000. Expectancy–Value Theory of Achievement Motivation. Contemporary Educational Psychology. 25, 1 (Jan. 2000), 68–81.

[20] Wolters, C.A. 2004. Advancing Achievement Goal Theory: Using Goal Structures and Goal Orientations to Predict Students' Motivation, Cognition, and Achievement. *Journal of Educational Psychology*. 96, 2 (2004), 236–250.

[21] Wolters, C.A. 2003. Understanding procrastination from a self-regulated learning perspective. *Journal of Educational Psychology*. 1, 95 (2003), 179.

[22] Muldner, K. et al. 2011. An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. In *Proceedings for the Conference on User Modeling and User-Adapted Interaction* (Jan. 2011), 99–135.

[23] Schmeichel, B.J. 2007. Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psycholgy General*. 136, 2 (2007), 241–255.

[24] Baker, R.S. et al. 2004. Detecting Student Misuse of Intelligent Tutoring Systems. In *Proceedings of the International Conference on Intelligent Tutoring Systems* (Berlin, Heidelberg, Aug. 2004), 531–540.

[25] Baker, R.S. 2007. Is gaming the system state-or-trait? Educational data mining through the multi-contextual application of a validated behavioral model. In *Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling* (2007).

[26] Baker, R.S. et al. 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* (New York, New York, USA, 2004), 383–390.

[27] Mcclelland, M.M. and Cameron, C.E. 2011. Self-Regulation in Early Childhood: Improving Conceptual Clarity and Developing Ecologically Valid Measures. *Child Development Perspectives*. 6, 2 (Jul. 2011), 136–142.