Exploring the Link Between Motivations and Gaming

Steven Dang
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15206
stevenda@cs.cmu.edu

Ken Koedinger
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15206
koedinger@cmu.edu

ABSTRACT

A student's ability to regulate their thoughts, emotions and behaviors in the face of temptation is linked to their task specific motivational goals and dispositions. Behavioral tasks are designed to strain a targeted resource to differentiate individuals through measures of their performance. In this paper, we explore how student behavior on differentially gamed learning material relates to estimates of student motivational goals and dispositions. We leverage observations of students in two different courses using an intelligent tutoring system over an entire academic year. We use a previously validated heuristic model of gaming detection to label instances of gaming. Each student's tendency to game is estimated separately using highly gamed and non-highly gamed sections of the course. Each estimate of student gaming is compared to pre-course self-reported measures of student motivations. Results indicate that in naturalistic settings, gaming on more challenging materials is less influenced by student motivations and potentially a result of adaptive learning behaviors. Similarly, student gaming estimates using only non-highly gamed material are significantly related to all targeted motivation measures. Implications and future directions are discussed.

Keywords

Motivation, Self-regulation, Gaming, Measurement, Intelligent Tutoring System

1. INTRODUCTION

Students in the US currently spend on average between 20-25 hours per academic year on standardized testing [29]. The largest cost of formal standardized tests is the cost of lost learning opportunities for students. Additionally, these formalized, high-stakes assessments also lead to a range of other systemic effects, such as reductions in topical coverage and cultures of teaching to the test, that result in negative impacts to student learning [1]. While the need for measurement of student performance at all levels is necessary for the continued improvement of educational institutions, there is a need to identify solutions that balance the need for information on institutional performance with the learning needs of the student.

Steven Dang and Kenneth Koedinger "Exploring the Link Between Motivations and Gaming" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 276 - 281

The increased prevalence of digital learning resources in schools has created an opportunity to explore an alternative solution to standardized testing. [26] demonstrated the viability of leveraging longitudinal observations of student performance in an intelligent tutoring system to match assessments of student mathematics aptitude from standardized exams. Similarly, [31], demonstrated how the educational design process can be extended to designing of educational games to produce game-based activities that produce valid assessments of student skill.

Student cognitive skill is only half of the formula for student success; they must also have the motivation to apply those skills diligently over time to achieve [11]. Currently there are inadequate instruments available for high stakes measurement of student motivational constructs [13].

Self-regulation related learner behaviors are linked to student motivations. The characteristics of the context when students demonstrate failures to self-regulate their learning behaviors can be informative of their motivational goals [28], their perceived value of the activity [15], and their beliefs about their self-efficacy [32]. Drawing on design principles of psychometric behavioral tasks, we believe we can identify contexts that sufficiently load on student self-regulation to measure student motivations. In this paper, we seek to explore the feasibility of leveraging observations of students' self-regulation as measured by gaming the system behaviors to measure student motivational goals and dispositions.

2. MOTIVATION

We define motivation as the orienting and invigorating impact on both behavior and cognition of prospective reward [9]. For this study, we focus on a set of well-defined goals and dispositions that have been shown to influence student motivation and achievement.

A student's interest in a domain will influence the subjective value any task from that domain. This perceived expected value from completing a task influences students' self-regulation decisions [15]. Students vary in their beliefs about their ability to successfully complete a task, their self-efficacy, and this difference in appraisal affects motivation to apply effort to a task [4]. Effort regulation describes the ability of students to motivate themselves and persevere on a task in the face of difficulty or failure [24]. Growth mindset captures the beliefs students have about the nature of intelligence and whether or not it is malleable [14]. As with self-efficacy, mindsets impact motivation through task appraisals. Student goals in academic tasks can be described using a two dimensional representation of mastery vs performance and approach versus avoidance [33]. Students with mastery approach goals set goals to learn any assigned knowledge and skills. Students with performance approach goals are motivated by a desire to perform better than their peers, while performance avoidant students are motivated by goals to avoid performing worse than their peers.

3. RELATED WORK

A significant body of prior work has focused on assessing moment-by-moment motivation through detectors of affect [10] and engagement [12,18]. However, work analyzing the link between fine-grained behavioral measures and motivational goals and dispositions is much more limited.

[27] created a rational model of student affect that leveraged a range of individual attributes including Big 5 personality measures and achievement goals. This work established the value of students' achievement goals on predicting moment by moment motivations as inferred by affect.

Several researchers attempted to identify task specific behaviors that rationally should be linked to achievement goals. [20] attempted to relate help-seeking behaviors while using an ITS to achievement goals. Researchers expected mastery-oriented students to be more likely to use a glossary/index resource, while performance-oriented students might tend to ask for hints from the tutor instead. No significant relationship between self-reported achievement goals and help-seeking behaviors was found. However, task achievement goals as predicted by choice of help resources did relate to learning outcomes as would be predicted by achievement goal theory.

[25] expanded on this work and attempted to relate task choice, where descriptions of each task were closely linked to corresponding achievement goals, to self-reported achievement goals and learning outcomes. In this work, task achievement goals as inferred by task choice predicted learning outcomes for the lesson but did not align with self-reported achievement goals. However, self-reported achievement goals were more predictive of course outcomes. Researchers speculated that self-reported goals might reflect an average tendency to be motivated by particular goals over a range of tasks within the domain and thus explaining alignment with more aggregated measures such as course outcomes.

Gaming the system, a pattern of behavior where students abuse the design of the learning environment to answer a particular question, is a well-documented behavior that has been linked to poor learning outcomes [2]. In [3], the authors test the relationship between a range of student motivations and gaming the system behaviors across two different ITS's. The study results supported a link between gaming behaviors and some motivational measures but not others. One of the strongest results indicated that student's attitudes and interest towards the domain was related to observed gaming frequency. There was also strong support for a link between experiences of frustration and gaming as well as a lack of drive to motivate themselves on tasks in general as well as in the face of challenge. The results demonstrated mixed or weak support for a relationship with growth mindset and perceptions of the helpfulness of the ITS help resources. Interestingly, the researchers failed to identify a relationship between observed gaming and performance goals, though the performance goal measures were not drawn from validated achievement goal instruments. Furthermore, this study used strictly observed gaming frequencies. Subsequent work has identified the joint role of contextual and student factors in explaining gaming behaviors [19,21].

In this paper, we seek to answer two main research questions. Research Question #1: How does the relationship between gaming and measures of motivation differ when gaming estimates are derived from either raw observations of gaming or using random effects models that account for both student-level and contextual variation. Research Question #2: How does student performance on educational content with varying degrees of gaming frequency relate to their different motivational goals and dispositions?

4. THE DATASET

For this study, we used a dataset drawn from [16] that was collected as part of a year-long study [6] in a suburban middle school in a mid-atlantic state.

The students used the Carnegie Learning Cognitive Tutor software (CogTutor). The CogTutor software provides adaptive instruction based on a fine-grained skill representation of the domain. The application divides problems into steps that must be answered individually and each map to independent skills in the domain model. Student practice problems are selected according to whether they have demonstrated mastery of necessary skills. The instruction is also scaffolded, allowing students to request multiple levels of hints at every step of the problem, providing on-demand problem scaffolding that provides increasingly informative support to the students. The data logs generated by the software are transformed into the standard learning data format specified by [16] before being utilized in this analysis. This format specifies how long students spend on every interaction, whether the action was correct, incorrect, or a hint, and what skill is associated with a specific problem step. Each interaction is represented as a single student transaction in the dataset, which includes over 2M such transactions across all students analyzed.

The dataset includes 189 students across 7 pre-algebra classes and 5 geometry classes. The population is predominantly white/Caucasian with only 2% of the sample being non-white. 56% of the students are female and 22% received either free or reduced lunch. The classes used the tutor for an entire academic year with an average of close to two class periods per week on the tutor. While the original dataset included 240 students across both of these courses, students with incomplete grade and survey responses were eliminated.

Additionally, some students and curricular sections were excluded due to having low observations in the data. The median student was observed during at least 40 sessions. However, 18 students were eliminated because EDA indicated these students as different from most others. The excluded students were observed in less than 20 class sessions and completed on average 780 total interactions with the system over the course of the year. By contrast, the median student completed about 15k transactions over the course of 40 sessions on average. These students were excluded from analysis because they appear unengaged and/or unmotivated, but there are so few observations of their behavior that drawing any conclusions from limited data is more prone to unobserved confounds.

Similarly, transactions from 31 sections are excluded from the dataset because they were observed with less than 6 students completing any work in the section. These sections are excluded because such sections might be measurements of only the fastest working or highest achieving students, thus introducing a bias to observations of gaming within those sections.

4.1 Motivational Measures

In addition to fine-grained student log data, several survey measures were collected at the beginning of the course to measure students' pre-course motivational goals and dispositions. Each scale utilized was drawn from well-validated instruments. Survey measures include scales for interest in math[17], self-efficacy [5], effort regulation [24], growth mindset[8] and achievement goals [33]. Questions from each scale are include in Figure 1 below. Each question was answered using a 5-point Likert rating, and responses for each scale were summed to represent students' motivation along each dimension.

Interest in Math
1. Math is practical for me to know
2. Math helps me in my daily life outside of school
3. It is important to me to be a person who thinks mathematically
4. Thinking mathematically is an important part of who I am
5. I enjoy the subject of math
6. I like math
7. I enjoy doing math
8. Math is exciting for me
Self-Efficacy
1. I am confident that I will do well in math class
2. I expect to do well in math
3. I am confident that I can learn future math concepts
4. Considering the difficulty of this course, I think I will do well in mathematics in the future
5. I am confident that I will do an excellent job on future math problems.
Effort Regulation
1. I often feel so lazy or bored when I do homework for math class that I quit before I finish what I planned to do.
2. I work hard to do well in math class even if I don't like what we are doing.
3. When class work is difficult, I give up or only study the easy parts.
4. Even when math class assignments are dull and uninteresting, I manage to keep working until I finish.
Growth Mindset
1. You have a certain amount of intelligence and you really can't do too much to change it.
2. Your intelligence is something about you that you can't change very much.
3. You can learn new things, but you can't really change your intelligence.
4. No matter who you are, you can change your intelligence a lot.
5. You can always greatly change how intelligent you are.
6. No matter how much intelligence you have, you can always change it quite a bit.
Achievement Goals
1. My aim is to completely master the material presented in this unit.
2. In this unit, I am striving to do well compared to other students.
3. In this unit, my goal is to learn as much as possible.
4. In this unit, my aim is to perform well relative to other students.
5. In this unit, my goal is to avoid performing poorly compared to others.
6. I am striving to understand the content of this unit as thoroughly as possible
7. My goal is to perform better than the other students in this unit
8. In this unit, I am striving to avoid performing worse than others.
9. In this unit, my aim is to avoid doing worse than other students.

Figure 1. Motivational survey inventory

4.2 Gaming the System Behaviors

We used the heuristic model of gaming behaviors introduced by [22] as this model appeared to produce better kappa on unseen data from across multiple systems including the CogTutor. Using this model, individual transactions were labeled according to a taxonomy that captures a range of relevant behaviors such as thinking before a hint request, spending time reading hint requests, and variations of guessing behaviors. Transactions are labeled as gaming if they are a member of a set of subsequent transactions that matches one of the thirteen identified patterns by [22] and shown in Figure 2. The patterns encode two primary types of gaming: guessing and hint abuse. Guessing patterns include placing the same answer incorrectly into multiple available answer slots and answering the same question rapidly with very small changes in the answer across attempts. Hint abuse patterns include not stopping to think about multiple subsequent errors before requesting help and rapidly requesting hints to seek a bottom-out hint, which in the CogTutor environment is simply the answer to the problem step given as the second or third hint.

Transactions are rolled-up into student steps, where each student step encapsulates metadata about all the transactions associated with a problem step until a correct answer is reached. Each student step is labeled as gamed if any transaction associated with the step was also labelled gamed. The resulting student step data was utilized to calculate student and content gaming frequencies.

The overall dataset included 3.5% gamed student steps. These numbers align reasonably well with gaming frequencies observed in prior work on CogTutor data. [2] found students gaming the system about 3% of the time based on in-classroom human observations. [23] found a slightly higher overall gaming frequency of 6.8% in their dataset utilizing the same detection model as used here. However, this deviation isn't so different that it is due to significant unobserved differences in the populations.

incorrect → [guess] & [same answer/diff. context] & incorrect
incorrect → [similar answer] [same context] & incorrect → [similar answer] & [same context] & attempt
incorrect → [similar answer] & incorrect → [same answer/diff. context] & attempt
[guess] & incorrect → [guess] & [diff. answer AND/OR diff. context] & incorrect → [guess] & [diff. answer AND/OR diff. context & attempt
incorrect → [similar answer] & incorrect → [guess] & attempt
help & [searching for bottom-out hint] → incorrect → [similar answer] & incorrect
incorrect → [same answer/diff. context] & incorrect → [switched context before correct] & attempt/help
bug → [same answer/diff. context] & correct → bug
incorrect → [similar answer] & incorrect → [switched context before correct] & incorrect
incorrect → [switched context before correct] & incorrect → [similar answer] & incorrect
incorrect → [similar answer] & incorrect → [did not think before help] & help → incorrect (with first or second answer similar to the last one)
help → incorrect → incorrect → incorrect (with at least one similar answer between steps)
incorrect → incorrect → incorrect → [did not think before help request] & help (at least one similar answer between steps)

Figure 2. Patterns of Gaming

The CogTutor content is organized hierarchically into multiple units. Each unit consists of several sections that themselves have multiple skills to be learned. Each section has problems that are divided into highly granular steps which each are associated with at least one skill. We chose to group observations at the section level to capture differences across the curriculum with sufficient resolution while having sufficient observations across students to make reasonable estimates of gaming frequency. The data included 206 sections with a mean gaming frequency of 1.95% and a standard deviation of 1.7%. A number of sections were found to have 0 observed gaming, while there was one extreme outlier section with a frequency of 12.12%.

Unlike in prior work, [2], no students were found to have never gamed throughout the year. The average student was observed gaming 3.66% of the time with a standard deviation of 1.16%. The minimum observed gaming frequency for students was 1.98% while the maximum observed was 11.95%.

4.3 Comparing measures of gaming

In this study, we generate four estimates of student gaming and compare these estimates of student gaming frequency to each motivational measure using partial correlations. In the partial correlations, we control for gender, ethnicity, and free/reduced lunch status.

$$(1) \qquad \theta_{ObservedGaming} = \frac{x_{NumGamed}}{N_{TotalSteps}}$$

$$(2) \qquad P(Gamed) \sim (1\ |Student\ ) + (1|Section)$$

$$(3) \qquad \theta_{Gaming} = e^{\theta_{student}}$$

To investigate RQ1, we calculate student gaming using frequencies calculated using only raw observations for each student as shown in Eq 1. We also predict gaming on each step using a random effects model with a random effect for student and section as shown in Eq 2. The model is fit over all observed student steps and the student gaming is found by calculating the exponential of the fitted random intercept, $\theta_{student}$, for each student as shown in Eq 3. To investigate RQ2, we divide the data into two subsets, hard sections and non-hard sections. We set an 80% quantile cutoff of 3.06% gaming frequency for each section to identify non-highly gamed sections. There were 164 non-highly gamed sections and 41 highly gamed sections. Again, the random effects model in Eq 2 was used for each data subset to estimate student gaming.

For RQ1, we expect student gaming estimates from the random effects model to better correlate with motivation because the model takes into account variance in gaming due to sections, which may not be observed for all students, as well as accounting for statistical noise due to sampling of a rare event.

For RQ2, we investigate the hypothesis informed by design principles of psychometric behavioral tasks. Measuring a targeted construct requires straining the resource and identifying a metric upon which to differentiate subject performance. Therefore, we expect estimates of student gaming using only highly gamed sections will have a more significant relationship with motivational variables compared to data without highly gamed sections.

5. RESULTS

The results of the partial correlation analysis are shown in Tables 1.1 and 1.2. The first row of both tables present evidence contrary to the results from [3]. Prior research found correlations with math interest, effort regulation, and growth mindset using only averages of observed gaming. However, in this dataset, only interest in the subject is related to gaming behaviors, and no other motivational measure has a significant correlation with student's gaming frequency.

On the other hand, the second row reflects correlations with student gaming estimated using a random effects model fitted with all of the data. In general, more motivational measures are correlated with these gaming estimates than those derived from the raw observations, which supports the hypothesis for RQ1. Comparing these results to [3], there are no direct measures of frustration, however it is possible that self-efficacy mediates whether student's experience of frustration explaining the correlation. Growth mindset is found to be marginally significant, which further bolsters the previous mixed evidence for a link between mindsets and average student gaming.

There are two cells where these correlations do not seem to agree with prior research. Effort regulation is expected to be correlated both as a matter of face validity as well as because prior research found a relationship between gaming and students' drive to persevere on academic work.

The link between achievement goals and gaming are mixed. In [3], the authors assessed performance goals using questions such as, "If you had your choice, what kind of extra-credit projects would you most likely do". It is unclear how this question maps to achievement goals, however, performance approach goals are not significant as might be extrapolated from prior work. On the contrary, mastery approach and performance avoidance goals are correlated with gaming. This relationship is rationally derived from the theory on self-regulation and motivation, but not predicted by specific prior work. Overall, the random effects model yielded a significant relationship to more motivational constructs than gaming estimates from raw observations.

Table 1.1 Correlations with Motivation Measures

Data Subset	Math Interest	Self Efficacy	Effort Reg.	Growth Mindset
Observed	-0.22**	-0.10	-0.11	0.00
All	-0.17*	-0.16*	-0.11	-0.14(.)
High Gaming	-0.19*	-0.14(.)	-0.10	-0.11
Low Gaming	-0.16*	-0.19*	-0.16*	-0.14*

(.) –0.10>p≥0.05, * - p<0.05, ** - p<0.01, *** - p<0.001

Table 1.2 Correlations with Achievement Goals

Data Subset	Mastery Approach	Performance Approach	Performance Avoidance
Observed	-0.03	-0.01	-0.05
All	-0.20**	-0.10	-0.15*
High Gaming	-0.14(.)	-0.08	-0.11
Low Gaming	-0.25***	-0.14(.)	-0.21**

(.) –0.10>p≥0.05, * - p<0.05, ** - p<0.01, *** - p<0.001

The results from estimating gaming using only highly gamed sections, row #3, are contrary to what is expected. Many of the correlations that appear when using all of the data, are weakened or not significant when using only the hardest questions. While the loss of significance with some constructs could be an artifact of random sampling from the full dataset, this does not explain the results seen in the bottom row. When estimating gaming using only non-highly gamed sections, correlations arise with every available motivational construct as seen in the fourth row. This is an unlikely consequence of sampling from the population and supports the idea that student gaming performance on highly gamed questions is introducing additional noise to the available signal in the rest of the data. Thus, the evidence points towards student gaming behaviors in the non-highly gamed sections as being more informative of student motivations than behaviors in the highly-gamed sections where self-regulation is under greater strain.

5.1 Exploring High-gamed sections

One possible explanation for this counterintuitive result is that gaming in highly gamed sections could be due to poorly designed content instead of cognitively challenge questions. To quantify average section difficulty, we calculate the ratio of the number of steps where student's first transaction is either a hint request or an error to the total number of student steps observed for each section (the assistance score). The highly gamed sections do in fact appear to be more difficult sections. The average highly-gamed section has 18.3% assistance steps with a standard deviation of 5.2%. The easiest section in this subset was observed with 9.9% assistance. By contrast, the average non-highly-gamed section consisted of 9.6% assistance steps with a standard deviation of 4.1%. Therefore, students required less overall assistance on the majority of non-highly-gamed sections than the easiest of the highly-gamed sections.

If the content, was in fact challenging to students, perhaps the average questions were beyond the abilities of students to answer. If students are unmotivated to try on difficult problems, then students should be most likely to game on the first encounter with a particular skill. As experience with a skill goes up, perceived difficulty should go down, thus monotonically reducing the probability of gaming with each practice opportunity. However, if a skill is beyond a student's ability and the student is sufficiently motivated to attempt to learn, then the student should be less likely to game on the first encounter of a skill than on subsequent encounters. In this case, after first encountering a skill, students are able to assess that the skill is too far beyond their abilities. On subsequent encounters with the skill, it is an adaptive behavior to abuse help resources or apply other gaming strategies to acquire the answer. Prior work has similarly found that not all gaming is harmful to learning [2], for instance students may be using hints as a form of worked example [30].

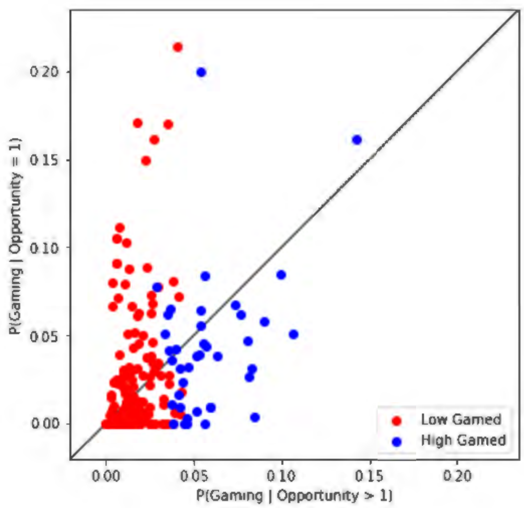

Figure 3. Gaming on 1st opportunity vs subsequent

In Figure 3, the observed gaming on 1st opportunities and all subsequent opportunities are compared for each section. Students in most highly gamed sections, in blue, are more likely to game of subsequent opportunities than the first. 70.7% of highly gamed sections share this characteristic as compared to 46.9% of non-highly gamed sections. This evidence, in additional to the lack of correlation between gaming on high-gamed sections and student's effort regulation supports the rational that gaming is sometime a somewhat desirable adaptive behavior for students and observations of gaming and in these sections should be treated differently than in other sections. In fact, the ratio of gaming on first opportunity to gaming on subsequent opportunities might be a valuable measure to incorporate into future motivational measurement models.

6. DISCUSSION

In this study, we demonstrate that leveraging random effects models to cope with statistical noise in observations of student's tendency to game on any given section better estimates student's gaming as related to their motivational goals and dispositions. Additionally, we provide initial evidence towards a measurement model of student's motivational goals and dispositions by leveraging observations of gaming. Results indicate a relationship between gaming and motivation that involves an interaction with the difficulty level and prior experience with a problem.

Several correlations with gaming estimates appear contrary to prior research and merit further analysis. The significant correlation with both mastery approach and performance avoidance disagrees with the results found by [3]. This disagreement could be due to the independence of achievement goals from each other, where gaming may be driven by an aggregate motivation of all achievement goals. More analysis is necessary to bridge this seeming contradiction and understand how patterns of gaming across problems of varying difficulty and prior experience might support an interpretation of gaming as indicative of different achievement goal profiles.

Gaming frequency was leveraged as a proxy measure for a range of unencoded difficulty factors. While this includes factors such as poor classroom instruction or a poorly designed cognitive model, it also encapsulates difficulty of individual problem-steps. A natural next step would be to investigate how more detailed student skill models might improve estimates of perceived difficulty and corresponding enrich model understanding of the nuances of why students are gaming and how this relates to different motivational goals.

Prior work has also shown that on longer time-scales, motivational goals are not necessarily stable [7]. In this study, we looked for relationships between pre-course motivations and in-course gaming behaviors. For students with fluctuations in achievement goals or self-efficacy, the contexts in which such students tend to game or disengage from the lesson in other manners might similarly change. Further analysis is necessary to investigate whether variations in gaming over time are similarly reflective of variations in motivational goals and dispositions over time.

Furthermore, the study included a fairly large body of students, but the observations were still limited to a single school in a particular region of the country with limited ethnic and socio-economic status diversity represented in the sample. Such factors are known to be correlated with variations in the types and frequencies of gaming behaviors observed in the population [21]. As such, we exercise caution in extrapolating these relationships beyond this demographic group without further validation.

Nonetheless, the results presented in this work lay the groundwork for further investigation into measurement models of motivational goals and dispositions that leverage an understanding of the contexts that strain students' self-regulation. Such unobtrusive measurement models hold the keys to a future where schools can better utilize instructional time that is currently occupied by standardized test and test-specific preparation while still receiving the student, and class-level performance measures necessary to support continuous improvement.

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] Wayne Au. 2016. High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. 36, 5: 258–267.

[2] Ryan Shaun Baker, Albert T Corbett, Kenneth R Koedinger, and Angela Z Wagner. 2004. Off-task behavior in the cognitive tutor classroom. 383–390.

[3] Ryan Baker. 2008. Why Students Engage in "Gaming the System" Behavior in Interactive Learning Environments. Why Students Engage in "Gaming the System" Behavior in Interactive Learning Environments 19, 2: 185–224.
[4] Albert Bandura. 1997. Self-Efficacy: The Exercise of Control. Macmillan.
[5] Albert Bandura. 2006. Guide for Constructing Self-Efficacy Scales. In *Self-Efficacy Beliefs of Adolescents*. 307–337.
[6] Matthew L Bernacki and Steven Ritter. 2013. Hopewell 2011-2012. Dataset 613 in DataShop. Retrieved from https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=613.
[7] Matthew L Bernacki, Vincent Aleven, and Timothy J Nokes-Malach. 2014. Stability and change in adolescents' task-specific achievement goals and implications for learning mathematics with intelligent tutors. *Computers in Human Behavior* 37: 73–80.
[8] Lisa S Blackwell, Kali H Trzesniewski, and Carol Sorich Dweck. 2007. Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child development* 78, 1: 246–263.
[9] Matthew Botvinick and Todd Braver. 2015. Motivation and Cognitive Control: From Behavior to Neural Mechanism. *Annual Review of Psychology* 66, 1: 83–113.
[10] Rafael A Calvo and Sidney D'Mello. 2010. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing* 1, 1: 18–37.
[11] Martin V Covington. 1992. *Making the Grade*. Cambridge University Press.
[12] Sidney D'Mello, Ed Dieterle, and Angela Duckworth. 2017. Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning. *Educational Psychologist*: 1–20.
[13] Angela L Duckworth and David Scott Yeager. 2015. Measurement matters assessing personal qualities other than cognitive ability for educational purposes. 44, 4: 237–251.
[14] Carol Dweck. 2007. Mindset: The New Psychology of Success.
[15] Jacquelynne S Eccles. 2005. Subjective task value and the Eccles et al. model of achievement-related choices. *Handbook of competence and motivation*: 105–121.
[16] Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J. 2010. A Data Repository for the EDM community: The PSLC DataShop. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press.
[17] Lisa Linnenbrink-Garcia, Amanda M Durik, AnneMarie M Conley, Kenneth E Barron, John M Tauer, Stuart A Karabenick, and Judith M Harackiewicz. 2010. Measuring Situational Interest in Academic Domains. *Educational and Psychological Measurement* 70, 4: 647–671.
[18] Caitlin Mills, Sidney D'Mello, Nigel Bosch, and Andrew M Olney. 2015. Mind Wandering During Learning with an Intelligent Tutoring System. In *Artificial Intelligence in Education*. Springer, Cham, Cham, 267–276.
[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: 99–135.
[20] Christine Otieno, Rolf Schwonke, Ron Salden, and Alexander Renkl. 2013. Can Help Seeking Behavior in Intelligent Tutoring Systems Be Used as Online Measure for Goal Orientation? In *Proceedings of the Annual Meeting of the Cognitive Science Society* 35, 35.
[21] Luc Paquette and Ryan S Baker. 2017. Variations of Gaming Behaviors Across Populations of Students and Across Learning Environments. In *International Conference on Artificial Intelligence in Education* 374-286.
[22] Luc Paquette, Adriana de Carvahlo, Ryan Baker, and Jaclyn Ocumpaugh. 2014. Reengineering the Feature Distillation Process: A case study in detection of Gaming the System. In *Proceedings of the Secenth International Conference for Educational Data Mining*. London, UK.
[23] Luc Paquette, Adriana M J A de Carvalho, and Ryan S Baker. 2014. Towards Understanding Expert Coding of Student Disengagement in Online Learning. In *Proceedings of the 36th Annual Cognitive Science Conference, 1126-1131*.
[24] Paul R Pintrich. 1991. A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ).
[25] J E Richey, Matthew L Bernacki, and D M Belenky. Relating a Task-Based, Behavioral Measure of Achievement Goals to Self-Reported Goals and Performance in the Classroom. In *Proceedings of the Annual Meeting of the Cognitive Science Society 36,36*.
[26] Steve Ritter, Ambarish Joshi, Stephen E Fancsali, and Tristan Nixon. 2013. Predicting Standardized Test Scores from Cognitive Tutor Interactions. In *Proceedings of the Sixth International Conference on Educational Data Mining.*
[27] Jennifer Sabourin, Bradford Mott, and James C Lester. 2011. Generalizing Models of Student Affect in Game-Based Learning Environments. In *Affective Computing and Intelligent Interaction* (2nd ed.). Springer, Berlin, Heidelberg, Berlin, Heidelberg, 588–597.
[28] Steven J Scher and Nicole M Osterman. 2002. Procrastination, conscientiousness, anxiety, and goals. *Psychology in the Schools* 39, 4: 385–398.
[29] Council of the Great City Schools. 2015. Student Testing in America's Great City Schools: An Inventory and Preliminary Analysis. 1–164.
[30] Benjamin Shih, Kenneth R Koedinger, and Richard Sheines. 2008. A Response Time Model for Bottom-Out Hints as Worked Examples. 117–126.
[31] Valerie J Shute. 2011. Stealth assessment in computer-based games to support learning. *Computer games and instruction* 55: 22, 503-524.
[32] Christopher A Wolters. 2003. Understanding procrastination from a self-regulated learning perspective. *Journal of Educational Psychology*, 95: 179.
[33] 2000. An Achievement Goal Theory Perspective on Issues in Motivation Terminology, Theory, and Research. *Contemporary Educational Psychology* 25, 1: 92–104.