Wheel-Spinning in a Game-Based Learning Environment for Physics

Thelma D. Palaoag[1(✉)], Ma. Mercedes T. Rodrigo[2],
Juan Miguel L. Andres[3], Juliana Ma. Alexandra L. Andres[2],
and Joseph E. Beck[4]

[1] University of the Cordilleras, 2600 Baguio City, Benguet, Philippines
tpalaoag@gmail.com
[2] Ateneo de Manila University, 1108 Quezon City, Metro Manila, Philippines
mrodrigo@ateneo.edu, alexandralandres@gmail.com
[3] Teachers College, Columbia University, New York, NY 10027, USA
jla2183@tc.columbia.edu
[4] Worcester Polytechnic Institute, Worcester, MA 01609, USA
josephbeck@wpi.edu

Abstract. We study wheel-spinning behavior among students using an educational game for physics. We attempted to determine whether students wheel-spin, and to build a wheel-spinning detector. We found that about 30 to 40 % of students are unable to successfully complete a level when attempting it 8 times or more, or when working on it for more than 160 s. We also found that past performance is predictive of wheel-spinning, and that persistence increases both the likelihood of success and of wheel-spinning. Finally, we found that wheel-spinning in this context is different from wheel-spinning exhibited in prior work in that it is relatively easy to detect and does not suffer from cold starts.

Keywords: Wheel-spinning · Physics playground · Persistence · Detector

1 Introduction

Persistence is a disposition, a habit of mind and action. In the context of this study, persistence is the ability to maintain an action, regardless of the person's feelings about achieving the task. It allows a person to keep taking action or pressing on even when he or she feels like quitting. [5] describes grit as tirelessness over the years to accomplish troublesome long-term objectives. Grit is portrayed with respect to stamina, stressing the part of exertion, interest, and passion in keeping focused on the objectives. In academic settings, these elements have a great impact on scholastic accomplishment or achievement.

While persistence is sometimes what distinguishes individuals who are successful and those who fail in an endeavor [5], a student's decision to persist or not is based on many factors. Early success or encouragement may result in the choice to persist in the face of difficult tasks. On the other hand, if a student's efforts are met with early failure, they may choose not to persist, opting instead give up or seek external help.

Persisting despite failure is not always productive or desirable. This type of persistence may in fact be wheel-spinning, a non-learning behavior coined by [3] referring to the failure to achieve mastery in a timely manner. Wheel-spinning denotes continuous, but futile effort. [4] found that wheel-spinning is probably related to knowledge deficits rather than boredom or other affective states. In this paper, we examine both persistence and wheel-spinning in the context of Physics Playground (PP), an educational game for physics. We attempt to (a) determine whether students wheel-spin within PP and, if they do, (b) build a detector for wheel-spinning.

2 Methods

2.1 Physics Playground

Physics Playground (PP) is a computer game designed to help high school students achieve a non-verbal conceptual understanding of how the physical world operates, characterized by an implicit understanding of concepts related to Newton's three laws: balance, mass, and conservation and transfer of momentum, gravity, and potential and kinetic energy. PP is described in detail in [1, 2].

Performance Metrics. Gold and silver badges are awarded to students who manage to solve a level. A gold badge is given to a student who is able to solve the level by drawing a number of objects equal to the particular level's par value (i.e., what the developers consider to be a reasonable number of objects needed to be drawn to solve the level). A student who solves a level using more objects will earn a silver badge. Many levels in PP have multiple solutions, meaning a player can solve the level using different agents.

2.2 Participant Profile

Data were gathered from 62 s year high school students, divided between one public school and one private school in Baguio City, Philippines. Participants ranged in age from 13 to 18 years old; 48 % were females and 52 % were males.

Thirty participants were from Bakakeng National High School (BNHS), located at Barangay Bakakeng Sur. The school has 7 instructional rooms and two 2 non-instructional rooms. With a total of 291 students, BNHS has a typical class size of around 42 students. Among the 30 participants, 15 came from honors sections, and 15 from regular sections.

The other 32 students were from University of the Cordilleras Grade and High School (UCGHS), located at Campo Filipino. 10 of the participants came from a star section, and the rest belonged to regular sections. Compared to other schools in the city, the UC grade school has maintained its tradition of academic excellence by winning in different interschool competitions.

2.3   Procedure

Participants were divided into batches of 15 to 17. Most batches of students played the game for 120 min; two batches played for only 90 min because they arrived at the testing venue late. As such, only the first 90 min of all the sessions were considered in this analysis.

2.4   PP Interaction Logs

We collected interaction log file data from 62 students but data from two students was corrupted. Hence, the succeeding analyses was based on data from 60 students only. These logs captured the following events:

– Menu Focus – an event that indicates the current playground and level,
– Level Start – an event that indicates that the player has begun playing a level,
– Level Restart – an event that indicates that the player triggered a level restart,
– Level End – an event that indicates that the player has finished playing a level,
– Time – an attribute of all events that pertains to the time the event was triggered,
– Object – an attribute of all events that pertains to the number of objects drawn, and
– Badges – an attribute of the Level End event that indicates what badge was given.

From these events, the following features we distilled: time spent on a level and number of restarts, both of which are considered by [6, 7] to be indicators of persistence. The number of restarts was considered to be equal to the number of attempts.

3   Wheel-Spinning

In prior work [3, 4], mastery was defined as three consecutive successful attempts at a skill. It was possible to adopt this as a criterion because the skills in the systems discussed in [3, 4] were traditional, structured tutoring systems with steps associated with defined skills. In PP, developers did not specify a mastery criterion. To answer our first research question, we did not attempt to associate specific skills with each level and we considered a level "mastered" once a student received either a gold or silver badge.

Per level, we noted how many attempts and how much time took for each student to achieve mastery. To compute the percentage of students who mastered a level, we counted the number of students who earned a gold or silver badge for a level, and then divided that number by the total number of students who attempted the level. We then computed for the average cumulative percentage of students who demonstrated mastery per number of attempt and over time.

The results are similar to those in [3, 4]. The cumulative percentage of students who achieve mastery plateaus. About 60 % of students master the skill after three attempts. After 8 attempts, there is almost no increase in the number of students who achieve mastery. Additional results show that about 60 % of students achieve mastery after about 80 s of working on a level. After 160 s, the number of students who master a level does not improve. It is therefore reasonable to conclude that at least some of these students were wheel-spinning.

4   Wheel-Spinning Detector

The first step in detecting wheel-spinning is to crisply define it for the PP context. Part of the difficulty in the task is that students can acquire either a gold or a silver badge. If a student acquires a silver badge after four attempts and 100 s of work, and struggles for 20 more minutes without earning a gold badge, was he wheel-spinning or not? The student made some progress, but the majority of the time was in engaged in behavior we would probably categorize as wheel-spinning.

   For our detector, we adopted the following definition of wheel-spinning:

1. All attempts after 15 min on a level were presumed to be wheel-spinning and were discarded, including any badges attained.
2. All attempts leading up to the first gold badge were not wheel-spinning, as the student made progress.
3. All attempts after the first gold badge were removed, as we were unsure how to score student performance. The student had already maxed out performance from the standpoint of the Tutor, and could have been experimenting for personal knowledge. There were 273 student-level attempts after receiving a gold badge.
4. All attempts leading up to the first silver badge were not wheel-spinning, as the student was making progress.
5. All attempts after the first silver badge that led to a gold badge, would be categorized as not wheel-spinning (this rule is a special case of rule #2).
6. All attempts after the first silver badge that did not lead to a gold badge were categorized as wheel-spinning.

Note that this framework categorizes some student effort on a problem as productive while later work can be wheel spinning. Reusing the example from the first paragraph, the students first four attempts would be considered not wheel-spinning, as it led to the student earning a silver badge. All subsequent attempts, up until a maximum of 15 min, would be labeled as wheel-spinning. For students who received a silver badge, there were 810 additional attempts at a level after. Unfortunately, only 50 of those attempts (6.2 %) were successful at making additional progress and receiving a gold badge. Surprisingly, students were more likely to achieve a gold badge if they had not already achieved a silver badge. One possible explanation is that stronger students realize they are not going to get a gold medal and so restart the level, while weaker students are happy to get any badge.

   To predict wheel-spinning, we used the following features:

1. The number of prior attempts the student has made to solve this level.
2. The cumulative amount of time (before this attempt), in seconds, the student has spent on this level.
3. Looking at past performance for this student, the probability he receives a gold medal in less than 5, 10, and 15 min (i.e., 3 features).
4. Similarly, based on past performance, the probability this student has received a silver medal in 5, 10, and 15 min (i.e., 3 features).
5. Whether the student has already earned a silver medal on this level.

To train the model, we used data from 14,232 student attempts at solving a level. We performed 10-fold cross validation, ensuring that each student's data was entirely within one of the folds. We trained a logistic regression model both for its interpretability, and for consistency with [3] in comparing the predictability of wheel-spinning. We experimented with a variety of temporal thresholds (5, 10, and 15 min) in case relatively stronger student performance, indicated by earning medals within 5 min, would be a stronger negative influence on wheel-spinning. However, the probability the student would receive a silver medal within 15 min was the best predictor. Spending additional time on the level unsurprisingly increased the probability of wheel-spinning. Interestingly, earning a silver medal on the level increased the likelihood of wheel-spinning. One explanation is that obtaining a gold medal is fairly difficult, and students attempting it were likely to get stuck.

The detector had fairly strong performance with an AUC of 0.853, and achieved 82.9 % correct in its predictions (vs. a base rate of 73.4 % correct).

The detector does a fairly good job at quickly detecting which students are unlikely to make progress on the current problem. As the student works longer within the problem, the tutor gains some additional information in terms of how long the student has spent, and whether or not he has earned a silver badge on the level. As a practical matter, performance when a student initially starts on a problem is just over 82 %, noticeably higher than the baseline of 73.4 % for guessing majority class. The detector does not suffer from a cold start problem.

5   Future Work, Contributions, and Conclusions

Next steps for this work include better analysis of the objects created by the student and level restart behavior. A better understanding of how students interact with the game will aid both detection of wheel-spinning and other pedagogical interventions.

This paper contributes to the ITS literature in at three ways. First, it demonstrates that the incidence of wheel-spinning is about the same within a game-based learning environment as it is in more traditional intelligent tutoring systems. About 30 to 40 % of students require additional intervention in order to help them towards mastery. Second, it shows that past performance is predictive of wheel-spinning and persistence. While increasing likelihood to succeed at a level, past performance also increases the probability of wheel spinning. Third, we identified that wheel-spinning in PP is different compared to wheel-spinning exhibited in ASSISTments and the Scatterplot Tutor [4]. Wheel-spinning in PP is relatively easy to detect, and does not suffer from the cold start problem seen in other work. Therefore, augmenting the tutor with an intervention to discourage students from wasting time should be straightforward.

In conclusion, this paper presents a first attempt at determining whether wheel-spinning behavior exists in Physics Playground. PP is an open-ended environment, and differs greatly from traditional ITS where wheel-spinning analyses have been done previously. We found that wheel-spinning exists, and that its emergence is non-random as it is predictable with our classifier. Determining how to utilize this detector is our next step.

References

1. Andres, J.M.L., Rodrigo, M.M.T.: The Incidence and persistence of affective states while playing Newton's playground. In: 7th IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (2014)
2. Andres, J.M.L., Rodrigo, M.M.T., Baker, R.S., Paquette, L., Shute, V.J., Ventura, M.: Analyzing student action sequences and affect while playing physics playground. In: International Workshop on Affect, Meta-Affect, Data and Learning (AMADL 2015), p. 24 (2014)
3. Beck, J.E., Gong, Y.: Wheel-spinning: students who fail to master a skill. In: Lane, H., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS, vol. 7926, pp. 431–440. Springer, Heidelberg (2013)
4. Beck, J., Rodrigo, M.T.: Understanding wheel spinning in the context of affective factors. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 162–167. Springer, Heidelberg (2014)
5. Duckworth, A.L., Peterson, C., Matthews, M.D., Kelly, D.R.: Grit: perseverance and passion for long-term goals. J. Pers. Soc. Psychol. **92**(6), 1087–1101 (2007)
6. Shute, V.J., Ventura, M., Kim, Y.J.: Assessment and learning of qualitative physics in Newton's playground. J. Educ. Res. **106**(6), 423–430 (2013)
7. Shute, V., Ventura, M.: Stealth Assessment: Measuring and Supporting Learning in a Video Games. MIT Press, Cambridge (2013)