RESEARCH



Evaluation and modeling of students' persistence and wheel-spinning propensities in formative assessments



Albert C. M. Yang^{1*} and Hiroaki Ogata²

*Correspondence: albertyang1994@gmail.com

¹ Department of Computer Science and Engineering, National Chung Hsing University, Taichung, Taiwan ² Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan

Abstract

Persistence represents a crucial trait in learning. A lack of persistence prevents learners from fully mastering their current skills and makes it difficult for them to acquire new skills. It further hinders the administration of effective interventions by learning systems. Although most studies have focused on identifying non-persistence and unproductive persistence behaviors, few have attempted to model students' persistence propensity in learning. In the present study, we evaluated students' persistence propensity in formative assessments by using an item response theory model with their attempt data. In addition, we modeled their wheel-spinning propensity. The students (N = 115) of first-level mathematics classes at a high school in Japan underwent the aforementioned formative assessments; their log data were collected. Persistence propensity was found to be correlated with frequency-related statistics, and wheel-spinning propensity was correlated with correctness-related statistics. However, persistence and wheel-spinning propensities were not correlated. A comparison of the students' scores with various persistence and wheel-spinning propensities revealed that both traits considerably influenced their academic performance. The present study provides insights into the use of attempt data to evaluate various characteristics crucial for learning, which are otherwise difficult to evaluate.

Keywords: Item response theory, Persistence, Wheel-spinning, Learner modeling, Formative assessment

Introduction

Persistence, or grit, is the ability of leaners to continue engaging in learning tasks until they master the relevant skills (Cloninger et al., 1993; Duckworth et al., 2007). Recent studies have reported that persistence is correlated with academic performance (Poropat, 2009), creativity (Prabhu et al., 2008), and long-term personal goals, such as studying later in life and future earnings (Datu et al., 2017; Duckworth et al., 2007). Educational psychology suggests that students who persist in their learning endeavors, embracing challenges as opportunities for growth, tend to achieve higher levels of skill proficiency (Dweck, 2006). This correlation is particularly evident in formative assessments, where persistent students demonstrate deeper engagement and higher rates of skill mastery



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativeCommons.org/licenses/by/4.0/.

(Black & Wiliam, 1998). Moreover, empirical studies underscore the importance of fostering persistence, as it enables learners to overcome obstacles and adaptively engage with complex material (Bandura et al., 1999). However, persistence may not always lead to successful outcomes. If a learner repeatedly attempts a task without achieving success, their subsequent efforts can become unproductive, resulting in a phenomenon known as wheel-spinning. Wheel-spinning refers to a behavior where students spend excessive time attempting to answer a question or solve a problem without making substantial progress (Beck & Gong, 2013). This behavior can hinder their learning process and impede their ability to move forward effectively. Wheel-spinning behavior indicates that learners continue to reattempt a task even when they recognize their inability to complete it. Beck et al. (2014) investigated whether affective factors influenced wheelspinning and identified a correlation between this behavior and gaming the system (e.g., guessing). Unproductive persistence may result in spending prolonged time and exhaustive effort on the challenges, thus leading to inefficient learning. Unlike wheel-spinning, in which learners persist unproductively, non-persistence refers to discontinuing a task without achieving the required mastery level. In addition to hindering the administration of effective interventions by learning systems, a lack of persistence prevents learners from fully mastering their current skills and makes it difficult for them to acquire new skills (Botelho et al., 2015).

In the context of a learning task, students can exhibit one of three persistence states: productive persistence, wheel-spinning, or non-persistence. In educational contexts, it is essential to differentiate between productive persistence and wheel-spinning behavior to optimize learning outcomes. Productive persistence refers to a learner's consistent and effective effort, marked by resilience and adaptive strategies, leading to meaningful skill development and mastery (Zimmerman, 2002). In contrast, wheel-spinning is characterized by continuous effort without significant progress, often due to ineffective learning strategies or lack of appropriate support (Beck & Gong, 2013). This distinction is crucial as it guides educators in tailoring their interventions; while productive persistence should be encouraged, wheel-spinning requires a shift in approach or additional support to overcome learning stagnation. The ability to identify these behaviors is key for educators, as it impacts not only the efficacy of instructional strategies but also students' motivation and confidence (Pintrich, 2003). Effective identification and response to these behaviors can lead to improved educational outcomes, making it a vital aspect of teaching and learning. Recent studies have aimed to predict students' non-persistence and wheel-spinning behaviors at an early stage in a learning task (Beck & Gong, 2013; Wang et al., 2020). These studies have explored methods to identify and anticipate when students might exhibit these unproductive or disengaged behaviors, allowing educators to intervene and provide timely support. Persistence propensity represents the tendency of learners to exhibit persistent behavior during the learning process. Research studies have emphasized the value of modeling persistence propensity for predicting student engagement and success in educational contexts (Credé & Phillips, 2011). These models provide insights into the role of motivation, self-regulation, and task characteristics in shaping individuals' persistence (Pintrich, 2003; Wolters, 2003). By considering these factors, educators can design interventions that promote adaptive persistence behaviors and cultivate a growth mindset, leading to improved learning outcomes and academic performance (Duckworth et al., 2007; Dweck, 2006). Such modeling efforts contribute to creating supportive learning environments that empower students to persist and overcome challenges, ultimately enhancing their educational experiences and long-term success.

Although various approaches and log data have been explored for modeling persistence, few studies have specifically modeled learners' persistence propensity using psychometric models. The successful application of various psychometric models, such as item response theory (IRT), for modeling the latent abilities of students in computerized adaptive testing (Choi & McClenen, 2020; Desmarais & Baker, 2012; Melesko & Novickij, 2019) has promoted the application of relevant models in the evaluation of persistence propensity (Zhang et al., 2021). In addition, the definition of persistence and the granularity of data for measurement may vary across learning contexts. Zhang et al. (2021) measured item-level persistence propensity and defined the trait as the propensity to reattempt an item after an incorrect response. Thus, the modeling of persistence propensity with a different granularity of log data is essential. In the present study, an IRT model was used to evaluate the topic-level persistence propensity of high school students through formative assessments in which students were assessed on their learned topics. In addition, their wheel-spinning propensity was measured; this parameter has not been extensively studied previously. Wheel-spinning propensity indicates the tendency of students to exhibit wheel-spinning behavior in the assessments. Subsequently, we investigated the correlation between the aforementioned traits and various attempt statistics to observe whether the latent traits reflect the corresponding behaviors and explored whether two latent traits were correlated. Further, the influence of the traits on long-term academic performance was evaluated. Overall, the following research questions were addressed in the present study:

- *RQ1* Can the item response theory model be used to model students' persistence and wheel-spinning propensities?
- *RQ2* What are the influences of persistence and wheel-spinning propensities on students' academic performance?

Literature review

Identifying non-persistence and wheel-spinning behaviors

Learners tend to disengage or abandon a given task for the following reasons: they have not mastered the prerequisite skill, the task is extremely difficult, the learner fails to manage time effectively, and they do not find the task sufficiently engaging (Kizil-cec & Halawa, 2015). Therefore, the identification of a non-persistence behavior and its underlying reasons is necessary. The definition of non-persistence behavior varies across learning contexts. For example, non-persistence behaviors were defined differently in two studies focusing on ASSISTments learning environments. Botelho et al. (2019) considered learners to be non-persistent if they abandoned a problem set without achieving the mastery level, whereas Kai et al. (2018) defined non-persistence as attempting <10 problems for a given problem set. A common approach for evaluating non-persistence behavior is to analyze process data, such as the amount of time spent on and the number of attempts made for a challenging task (Ventura et al., 2013; Wang et al., 2020).

However, these attempts may not be all productive; in addition, the evaluated persistence behavior may be accompanied by the wheel-spinning behavior. Therefore, differentiating wheel-spinning behavior from productive persistence behavior is crucial. Wheel-spinning behavior may have slightly different interpretations and definitions depending on the specific learning context. Beck and Gong (2013) defined wheel-spinning as the inability to achieve a predefined mastery level even after attempting \geq 10 problems within a problem set. Similarly, Kai et al. (2018) defined wheel-spinning as the inability of students to correctly solve three consecutive problems in a total of 10 attempts or to demonstrate the retention of a newly gained skill in future assessments.

Modeling persistence propensity with item response theory

Because persistence is essential for learning, the modeling and evaluation of this trait have been attempted recently. Persistence propensity has previously been assessed using self-reported questionnaires (Datu et al., 2017), on the basis of the time spent on gamebased assessments (Ventura & Shute, 2013) and challenging tasks (Ventura et al., 2013), and by enumerating the number of attempts made by leaners before answering a question correctly (Wang et al., 2020). Another modeling approach is to use psychometric models, such as IRT. IRT models are used to estimate a range of latent traits that cannot be observed directly using survey or learning logs. Scoular and Care (2020) used an IRT model to identify student behavior patterns in collaborative problem-solving tasks. Zhang et al. (2021) used a tree-based IRT model to evaluate the latent ability and resilience of students during assessments allowing multiple attempts. IRT models were initially used to assess students' latent abilities and item difficulties using a set of the responses of all students to assessment items. Compared with conventional approaches, such as considering test scores, IRT models consider item properties (i.e., difficulty and discrimination) when estimating the latent ability of leaners. The idea is that the probability of a correct response to an item is a function of the students' latent traits and the item (Kolen, 1981; Reckase, 1997). This method is commonly used in computerized adaptive testing (Jia & Le, 2020). One advantage of using IRT models over the direct use of attempt data for evaluating persistence propensity (e.g., number of topics persist) is that the former considers the differences in task properties (e.g., repeatability or difficulty for learners to persist). When two students persist on the same number of tasks, the student who persists on the tasks that most students tend to quit on is considered to be more persistent than the other. A limited number of studies have been conducted to evaluate persistence propensity using IRT or other psychometric models. In addition, persistence propensity can be measured using various data granularities. For instance, it can be assessed through metrics such as the number of attempts made on a particular question, the number of questions attempted within a specific topic, or the overall count of assignments completed throughout a course.

Methods

Context

We obtained learning data (July 2021–February 2022) from the students in three firstlevel mathematics classes at a high school in Japan. This study included a total of 115 students. All classes used the same textbooks and were conducted in physical classrooms. The same instructors delivered the learning content, and the lesson and method of instruction for each topic were identical across each classroom setting. Students could access various learning materials and exercises through an ebook reading system, Book-Roll (Flanagan & Ogata, 2018). The students' learning behavior data on all components were stored in an LRS system for subsequent analysis. The students were assessed after they learned each topic. The exercises for assessment included open-text mathematical questions from their textbook, which were also uploaded on BookRoll to obtain students' attempt data. An example of the question: *Consider a set A that contains all the even numbers between 1 and 20, and a set B that contains all the multiples of 3 between 1 and 20. Create a new set C, which is the union of sets A and B. Determine the number of elements in set C, and explain the process you used to find this number.* When solving the exercise problems, the students could answer directly in their textbook or on BookRoll. However, they were encouraged to attempt the exercises on BookRoll.

Figure 1 illustrates the experimental process. Before the experiment, a mathematical examination involving the mathematical knowledge they learned was conducted. During the experiment, the students were instructed to attempt the aforementioned exercises; this helped us evaluate their knowledge level after they learned each topic. All exercises were already uploaded on BookRoll to ensure that the students could attempt the exercises at any given time. Their performance on these exercises did not affect their final grades, and they could repeatedly attempt the exercises to assess their knowledge levels. At the end of the experiment, another examination was conducted to evaluate the levels of mathematical knowledge gained by the students. These scores represented their final grades. Thus, the first examination (pretest) was conducted to measure the levels of the students' prior knowledge, whereas the second examination (posttest) was conducted to evaluate their learning performance. The scores on both examinations ranged from 0 to 100.



Fig. 1 Experimental process

Modeling persistence and wheel-spinning propensities using an IRT model

Various IRT models are used for relevant purposes. The inclusion of several parameters in an IRT model indicates the possibility of evaluating numerous item properties. For example, a two-parameter logistic model considers both difficulty and discrimination of items, whereas a one-parameter logistic model considers only one parameter by treating the discrimination parameter as a constant. Furthermore, the Rasch model, which is a simple model, neglects the effects of discrimination and assigns the parameter to one. Although the inclusion of several item properties may offer a high amount of information regarding students' latent traits, the calculation time may increase substantially with increasing sample size. To reduce calculation time, we selected the Rasch model to evaluate the students' latent traits. In the Rasch model, the probability of a learner (s) responding to an item (i) correctly is calculated using the following formula:

$$P_{si}(\theta_s, b_i) = \frac{exp(\theta_s - b_i)}{1 + exp(\theta_s - b_i)} \tag{1}$$

 θ_s is the latent ability of s, and b_i is the estimated difficulty of i. In the present study, students were assessed after they learned a given topic. Their attempt data were used to model their persistence and wheel-spinning propensities. Because the latent traits were modeled on the basis of the students' frequency of persistence and wheel-spinning behaviors, thus indicating their propensity to exhibit the behaviors during assessments. The number of items varied across topics. The number of items was 1–10 for most topics but exceeded 20 for a few. The students' attempt data on each topic were converted into two nodes: persistence and wheel-spinning. The persistence node—Np(s,t)—represented the persistence behavior of a student (s) on a topic (t); by contrast, the wheel-spinning node—Nws(s,t)—represented the wheel-spinning behaviors vary across learning contexts. Figure 2 shows the process of determining the value of persistence and wheel-spinning nodes for a topic in the present study. Considering that a small number of items were developed for most topics during the experiment, students were considered to exhibit



Fig. 2 The process of determining the value of persistence and wheel-spinning nodes for a topic

wheel-spinning behavior if they attempted > 1 items but failed in approximately 33.3% of the attempted items on a topic. Thus, Np(s,t) is 1 if s persists on t, otherwise 0; similarly, Nws(s,t) is 1 if s persists but exhibits wheel-spinning behavior on t, otherwise 0. Because students can repeatedly attempt an item, we used the result of the last attempt to determine their success on the item. Notably, Nws(s,t) is unknown if s does not persist on t because only students who does not give up the topic can exhibit wheel-spinning behavior. Subsequently, persistence and wheel-spinning propensities were evaluated using an IRT model based on their corresponding nodes for each topic. A high persistence propensity suggests that the student often attempts > 1 item on a topic, and high wheel-spinning propensity indicates that the student often fails on 33.3% of the attempted items on a topic. The model calculates topic difficulties for each latent trait, which indicates the difficulty experienced by the student to exhibit the corresponding behaviors on the topic. Few students persist and wheel spin on a topic with high persistence difficulty and one with high wheel-spinning difficulty, respectively.

Preprocessing

A total of 8849 attempt logs were collected, involving 147 topics and 896 items. Because the numbers of attempted topics and items varied across students, preprocessing was performed to address the sparsity present in the log data before modeling the latent traits. To ensure sufficient attempt data and an adequate number of students, we selected topics that were attempted by at least 12 (i.e., median number of student attempts) students and were represented in at least two items. In addition, we only included students who attempted more than 10 selected topics. After preprocessing, a total of 7933 attempts remained, which involved 66 topics, 416 items, and 99 students.

Data analysis

In the present study, preprocessed log data were input in the Rasch model developed using Girth, an IRT package for Python (https://pypi.org/project/girth/). The model-estimated latent traits (i.e., persistence and wheel-spinning propensities) and topic properties ranged between -5.99 and +5.99. The maximum value of the estimated persistence propensity indicated that the student persisted on every attempted topic, and the maximum value of persistence difficulty indicated that no students persisted on the topic. The maximum value of estimated wheel-spinning suggested that the student always exhibited wheel-spinning behavior when persisting on a topic, and the maximum value of wheel-spinning difficulty suggested that no students exhibited wheel-spinning behavior on the topic. We applied Spearman correlation analysis to investigate the relationship between the latent traits and various attempt statistics calculated using the complete set of attempt data. These statistical parameters included the students' average number of attempts per topic, percentages of attempted topics and items, and percentages of the correct response across attempts and the eventual correct response. The percentages of topics on which the students persisted or exhibited wheel-spinning behavior were also calculated. Next, the Spearman correlation coefficient between persistence propensity and wheel-spinning propensity and that between persistence and wheel-spinning difficulties were calculated. The students were divided into four groups on the basis of their latent traits (Table 1). Groups A, B, C, and D, respectively, comprised students with high

Group	Persistence propensity	Wheel- spinning propensity	Description
A	High	High	The students showed a tendency to attempt more than one item on the selected topics but often failed on one-third of them
В	High	Low	The students showed a tendency to attempt more than one item on the selected topics and rarely failed on one-third of them
С	Low	High	The students only attempt more than one item on a few selected topics and often failed on one-third of them
D	Low	Low	The students only attempt more than one item on a few selected topics but rarely failed on one-third of them

Table 1	Persistence and	wheel-spinning	propensities of the	study groups
---------	-----------------	----------------	---------------------	--------------

Table 2 Example logs of the study groups

	Standard and general types			Equation of tangent to a circle			Common point of two circles		
	Attempt	Correct	Incorrect	Attempt	Correct	Incorrect	Attempt	Correct	Incorrect
S1 (A)	5	3	2	2	0	2	12	4	8
S2 (B)	3	3	0	1	0	1	8	8	0
S3 (C)	1	1	0	1	0	1	6	3	3
S4 (D)	1	1	0	1	0	1	4	3	1

The boldfaced value in the attempt column represents that the student quit on the topic, whereas the boldfaced value in the incorrect column represents that the student exhibited wheel-spinning behavior on the topic

persistence propensity and high wheel-spinning propensity, high persistence propensity but low wheel-spinning propensity, low persistence propensity but high wheel-spinning propensity, and low persistence propensity and low wheel-spinning propensity. The median of persistence (M=0.04) and wheel-spinning (M=0.16) propensities were used as the threshold for assigning groups. Note that although the non-persistence behavior and wheel-spinning behavior on a topic are mutually exclusive, one can possess low persistence propensity and high wheel-spinning propensity simultaneously, as the latter represents the tendency of students to exhibit wheel-spinning behavior when they persist on a topic. Table 2 presents the example logs of the aforementioned groups. We compared the pretest and posttest scores of the groups to evaluate the influence of the latent traits on the academic performance of the students.

Results

Distribution of estimated persistence and wheel-spinning propensities

Figure 3 presents the distribution of estimated persistence and wheel-spinning propensities and persistence and wheel-spinning difficulties. Higher estimated latent traits indicate the higher frequency with which the student exhibits the corresponding behaviors in the assessments, while higher estimated topic properties represent that fewer students exhibit the related behaviors on that topic. Most estimated persistence (M=0.04) and wheel-spinning (M=0.16) propensities were between -2 and +2, suggesting that only a small portion of students exhibited high persistence or wheel-spinning propensities. The persistence difficulties for most topics exceeded 0 (M=2.12), indicating that many students attempted only a single item on the selected topics and quit. The wheel-spinning difficulties for most topics also exceeded 0 (M=1.04), implying that most students



Fig. 3 Distribution of persistence and wheel-spinning propensities (left) and persistence and wheel-spinning difficulties (right)

Estimated trait	Avg attempts per topic (Mean = 2.52)	Percentage of attempted topics (Mean = 0.22)	Percentage of attempted items (Mean = 0.10)	Percentage of the correct response across attempts (Mean = 0.67)	Percentage of the eventual correct response (Mean = 0.70)	X ^a (Mean = 0.53)	Y ^b (Mean =0.20)
Persis- tence	0.51***	0.94***	0.97***	0.16	0.11	0.55***	0.03
Wheel- spinning	- 0.20*	-0.12	-0.13	-0.88***	- 0.92***	- 0.16	0.91***

X^a, Percentage of topics across attempts on which the students persisted; Y^b, Percentage of topics across attempts on which the students exhibited wheel-spinning behavior. *p < .05; **p < .01; ***p < .001

did not exhibit wheel-spinning behavior when persisting on the topics. Furthermore, the maximum wheel-spinning difficulty of 5.99 was noted for some topics, indicating that the students who persisted on these topics exhibited no wheel-spinning behavior.

Correlation between latent traits and various attempt statistics

To answer RQ1, we compared the students' latent traits estimated using partial attempt data with various statistics in the complete dataset. As shown in Table 3, the estimated persistence propensity was positively correlated with the average number of attempts per topic (r=0.51; p<0.001), percentage of the attempted topics (r=0.94; p<0.001), and percentage of attempted items (r=0.97; p<0.001); this finding indicated that persistence propensity was correlated with frequency-related statistics. Furthermore, the estimated wheel-spinning propensity was correlated with the correct response percentage across attempts (r=-0.88; p<0.001) and the eventual correct response (r=-0.92; p<0.001); this finding indicated that wheel-spinning propensity was correlated with correctness-related statistics. Persistence and wheel-spinning propensities were significantly correlated with the number of topics on which the students persisted (r=0.55; p<0.001)



Fig. 4 Correlation between latent traits (left) and topic properties (right)

Group	Ν	Adjusted mean	Standard deviation	F	n2	Post Hoc (t test)
A	23	53.59	11.77	4.17**	0.09	A>C
В	27	60.58	10.96			B > A B > C B > D
С	27	47.65	14.25			
D	22	58.78	14.64			D>A D>C

 Table 4
 Results of one-way analysis of covariance of the posttest scores of the groups

*p < .05; **p < .01; ***p < .001

and the number on which the students exhibited wheel-spinning behavior (r = 0.91; p < 0.001), respectively. Thus, the latent traits modeled using the IRT model reflected the likelihood of the corresponding behaviors.

Correlation between latent traits and topic properties

We further investigated the correlation between the estimated traits and topic properties. As shown in Fig. 4, no correlation (r = -0.15) was evident between the two latent traits, which suggests that the two traits were independent of each other. A strong weak correlation ($r = 0.29^*$) was noted between the two topic properties, implying that students merely exhibited wheel-spinning behavior on topics most students did not persist on.

Influence of latent traits on students' academic performance

To answer RQ2, we divided the students into the aforementioned four groups and compared their posttest scores. Groups A, B, C, and D, respectively, comprised 23, 27, 27, and 22 students. One-way analysis of covariance (ANCOVA) was used to evaluate the influence of persistence and wheel-spinning propensities on the academic performance of the students. In this analysis, the covariate, independent variable, and dependent variable were the pretest score, group, and posttest score, respectively. The results of Levene's test (F=0.80; p > 0.05) revealed that the posttest scores varied consistently across groups. The analysis of the interaction effect between the covariate and the independent variable indicated that the regression coefficients exhibited intragroup homogeneity (F=0.45; p > 0.05). The results of the ANCOVA (Table 4) revealed a significant intergroup difference (F=4.17; p < 0.01) after covariate adjustment, which suggested that the students' latent traits markedly influenced their posttest scores. The adjusted mean of the posttest scores of group B (adjusted mean = 60.58; standard deviation [SD] = 10.96) was considerably higher than that of groups A (adjusted mean = 53.59; SD = 11.77), C (adjusted mean = 47.65; SD = 14.25), and D (adjusted mean = 58.78; SD = 14.64); this finding indicated that the students who persisted but rarely exhibited wheel-spinning behavior outperformed those who exhibited this behavior or quit. The adjusted mean of the posttest scores of group D was substantially higher than that of groups A and C, revealing that students exhibiting low persistence propensity outperformed those exhibiting high wheel-spinning propensity and those exhibiting both propensities. Thus, persistence and wheel-spinning propensities markedly influenced academic performance.

Discussion and conclusion

We used an IRT model to evaluate students' persistence and wheel-spinning propensities in formative assessments and investigated the correlation between the latent traits and various attempt statistics. The study contributes to the relevant literature by providing insights into the evaluation of psychometric traits (e.g., persistence and wheel-spinning propensities) that are essential to learning. Persistence propensity was correlated with frequency-related statistics (e.g., average number of attempts per topic and percentages of attempted topics and items), whereas wheel-spinning propensity was correlated with correctness-related statistics (e.g., percentages of correct response across attempts and the eventual correct response). These findings are consistent with those reported by Whitmer et al. (2019) and Zhang et al. (2021). Whitmer et al. (2019) indicated that self-reported grit in a learning management system was strongly correlated with the number of attempts made during an assessment. Zhang et al. (2021) reported that persistence was correlated with a decision to reattempt an item after providing an incorrect response. They further identified a correlation between persistence and the percentage of eventual correct responses. This indicates that wheel-spinning propensity is correlated with item-level persistence; thus, students who show a tendency to reattempt an item until responding correctly are less likely to exhibit wheel-spinning behavior on the topic. However, this correlation requires further validation. The persistence and wheel-spinning propensities were strongly correlated with the number of topics the students persisted on and that the students exhibited wheel-spinning behavior on, respectively; thus, the proposed approach reflects the likelihood of the corresponding behaviors. The present study provides an unobtrusive and practical approach for modeling individuals' persistence and wheel-spinning propensities, which are key psychometric traits associated with long-term educational and work-related outcomes (Datu et al., 2017; Duckworth et al., 2007).

We found no correlation between persistence propensity and wheel-spinning propensity, which indicated that high persistence propensity may not necessarily lead to high wheel-spinning propensity. This finding suggests that non-persistence and wheel-spinning propensities may result from various factors and should be differentiated. The low correlation between the two propensities represents that both latent traits should be included when creating learner profiles. Modeling latent traits may have several advantages. First, these traits may be used in various prediction studies. Although the current effective models facilitate the early prediction of wheel-spinning and non-persistence propensities on a given task (Mu et al., 2020; Wang et al., 2020; Zhang et al., 2019), the level of predictive performance decrease substantially during the next assignment (Botelho et al., 2019). The modeling approach used in the present study may be used to predict the non-persistence and wheel-spinning behaviors for the next assignment. Thus, a researcher can estimate students' current latent traits using only attempt data in previous tasks to predict the students' non-persistence and wheel-spinning behaviors for the next task. Second, the estimated traits can be used to create learner profiles for personalized learning, such as learning paths or quiz recommendations. In addition, the topic properties obtained using the IRT model can be used to provide personalize interventions. For example, recommending topics with low persistence difficulty to students with low persistence propensity may prevent them from quitting.

We calculated the posttest scores of the students and found that the two latent traits influenced their academic performance. As expected, the students with high persistence and low wheel-spinning propensities exhibited the best performance. Although some studies have indicated a correlation between persistence and academic performance (Borghans et al., 2008; Poropat, 2009), others have reported no apparent correlation between the two (Duckworth et al., 2007; Zhang et al., 2021). The inclusion of wheel-spinning propensity may affect the correlation results because persistence may not always be productive. As mentioned, students may guit after their first attempt for various reasons; they may find the topic extremely difficult or boring or believe that they understand the topic well and thus stop after achieving success on their first attempt. In the present study, some of the students might have preferred attempting the exercises in their paper-based textbook or might not have been assessment-oriented because attempting the exercises and submitting the responses on BookRoll were not mandatory. Furthermore, some students exhibited wheel-spinning behavior after persistently attempting exercises and submitting their responses on BookRoll; this suggests that although these students made effort, they could not be successful. Students who have been struggling for a long time should be stopped from persisting and restudy the knowledge instead. Previous studies have shown that wheel-spinning behavior correlates with gaming behaviors, such as random guessing (Beck et al., 2014). Thus, further investigation of the factors contributing to persistence and wheel-spinning propensities in the current context is essential and will be our next step.

Implications

Our findings have several implications. First, the study proposed an unobtrusive and practical approach for modeling crucial traits that are otherwise difficult to evaluate. Although previous studies have attempted to predict and identify non-persistence and wheel-spinning behaviors, these approaches were unable to quantify these latent traits. Thus, our study may serve as a reference for future studies aimed at modeling persistence and wheel-spinning propensities using other types (e.g., temporal data) and granularity (e.g., item-level and topic-level) of data. The latent traits estimated in our study may be used to predict students' quitting and wheel-spinning behaviors for future tasks. Because non-persistence and wheel-spinning propensities negatively influenced academic performance, future studies should consider using latent traits for the early identification of at-risk students. Second, our study may facilitate the modeling of persistence propensity in other contexts (e.g., persistence in reading or collaborative learning). The correlation between the persistence propensity in various contexts can be explored. Furthermore, various models may be used to assess other psychometric traits and behaviors, such as the propensity to copy answers, cheat, or procrastinate, which can be combined to create learner profiles for personalized learning. For practice, teachers may identify students who tend to quit or exhibit wheel-spinning behavior during an assessment and design appropriate interventions to help the students, such as sending warning messages. The topic properties obtained using the IRT model may also enable teachers to improve their students' latent traits. For example, recommending topics with low persistence difficulty to students who tend to quit or exhibit wheel-spinning behavior during an a given topic. Finally, these topic properties may help teachers to improve their teaching approach. The topics on which many students quit or exhibit wheel-spinning behavior may be perceived as difficult or complex. Thus, teachers must invest extensive effort in teaching these topics and ensure the required levels of baseline knowledge in their students before permitting them to attempt exercises.

Limitations and future research

Our study has some limitations. First, attempting the exercises on all topics was not mandatory, and the number of items per topic varied, which resulted in data sparsity. Because the students selected the topics for practice, the attempted topics markedly varied across the students. Some students attempted even those topics that their teachers did not teach. The number of topics attempted by each student varied despite data preprocessing, which might have influenced the evaluation outcomes of the latent traits. For example, of two students who quit on all of their attempted topics, the student who attempted fewer topics might have had higher persistence propensity than the one who attempted more topics. Data sparsity might have also affected the evaluation outcomes of topic properties. Some topics exhibited the maximum value of wheel-spinning difficulty. This might be because only a few students attempted the aforementioned topics. Future studies are recommended to devise stricter strategies for evaluating the latent traits, such as by selecting a specific number of items and topics as the study exercise and making it mandatory for all students to attempt the exercise. Second, the definitions of persistence and wheel-spinning behaviors in the current context might have differed from those in the relevant literature. Because definitions may considerably influence evaluation outcomes, other definitions can be explored to evaluate these behaviors. In addition, we modeled persistence and wheel-spinning propensities in formative assessments. Future studies can compare the latent traits evaluated using the IRT model with the existing measures of persistence that consider both assessment and other activities through self-report or controlled experiments. It should also be noted that the findings in this study may not generalize to different classroom settings, grade levels, subjects, and geographic locations due to heterogeneity and variability in introductory high school mathematics curricula and the aptitude levels of the students enrolled in this particular course. Finally, we analyzed only the quantitative and aggregate data (i.e., the number of items attempted and the final responses of the students) on each topic. Future studies can include qualitative and other types of data, such as interview, self-reported,

and temporal data to investigate key factors that influence students' tendencies to persist or exhibit wheel-spinning behavior in learning tasks.

Acknowledgements

Not applicable.

Author contributions

ACMY and HO contributed to the research conceptualization and methodology. Data collection was performed by HO. ACMY analyzed the data and wrote the manuscript. HO provided substantial comments to improve all manuscript versions. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The data of this study is not open to the public due to participant privacy.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 26 September 2023 Accepted: 27 November 2023

Published online: 04 December 2023

References

Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-efficacy: The exercise of control.

- Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. In International conference on artificial intelligence in education (pp. 431–440). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_44
- Beck, J., Rodrigo, M., & Mercedes, T. (2014). Understanding wheel spinning in the context of affective factors. In International conference on intelligent tutoring systems (pp. 162–167). Springer, Cham. https://doi.org/10.1007/978-3-319-07221-0 20
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74.
- Borghans, L., Meijers, H., & Ter Weel, B. (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, 46(1), 2–12. https://doi.org/10.1111/j.1465-7295.2007.00073.x
- Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A., & Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. *IEEE Transactions on Learning Technologies*, 12(2), 158–170. https:// doi.org/10.1109/TLT.2019.2912162
- Botelho, A., Wan, H., & Heffernan, N. (2015). The prediction of student first response using prerequisite skills. In Proceedings of the Second (2015) ACM Conference on Learning@ Scale (pp. 39–45). https://doi.org/10.1145/2724660.2724675
- Choi, Y., & McClenen, C. (2020). Development of adaptive formative assessment system using computerized adaptive testing and dynamic Bayesian networks. *Applied Sciences*, *10*(22), 8196. https://doi.org/10.3390/app10228196
- Cloninger, C. R., Svrakic, D. M., & Przybeck, T. R. (1993). A psychobiological model of temperament and character. Archives of General Psychiatry, 50(12), 975–990. https://doi.org/10.1001/archpsyc.1993.01820240059008
- Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the motivated strategies for learning questionnaire. *Learning and Individual Differences*, 21(4), 337–346. https://doi.org/10.1016/j.lindif.2011.03.002
- Datu, J. A. D., Yuen, M., & Chen, G. (2017). Development and validation of the Triarchic Model of Grit Scale (TMGS): Evidence from Filipino undergraduate students. *Personality and Individual Differences*, 114, 198–205. https://doi.org/10. 1016/j.paid.2017.04.012
- Desmarais, M. C., & Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. User Modeling and User-Adapted Interaction, 22(1), 9–38. https://doi.org/10.1007/s11257-011-9106-8
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92(6), 1087. https://doi.org/10.1037/0022-3514.92.6.1087

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. Knowledge Management & E-Learning: An International Journal, 10(4), 469–484.

- Jia, J., & Le, H. (2020). The design and implementation of a computerized adaptive testing system for school mathematics based on item response theory. In *International Conference on Technology in Education* (pp. 100–111). Springer, Singapore. https://doi.org/10.1007/978-981-33-4594-2_9.
- Kai, S., Almeda, M. V., Baker, R. S., Heffernan, C., & Heffernan, N. (2018). Decision tree modeling of wheel-spinning and productive persistence in skill builders. *Journal of Educational Data Mining*, 10(1), 36–71. https://doi.org/10.5281/ zenodo.3344810
- Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. In *Proceedings of the second (2015)* ACM conference on learning@ scale (pp. 57–66). https://doi.org/10.1145/2724660.2724680.
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 1–11.

Melesko, J., & Novickij, V. (2019). Computer adaptive testing using upper-confidence bound algorithm for formative assessment. *Applied Sciences*, 9(20), 4303. https://doi.org/10.3390/app9204303

Mu, T., Jetten, A., & Brunskill, E. (2020). Towards Suggesting Actionable Interventions for Wheel-Spinning Students. International Educational Data Mining Society.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95(4), 667. https://doi.org/10.1037/0022-0663.95.4.667

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322. https://doi.org/10.1037/a0014996

Prabhu, V., Sutton, C., & Sauser, W. (2008). Creativity and certain personality traits: Understanding the mediating effect of intrinsic motivation. *Creativity Research Journal*, *20*(1), 53–66. https://doi.org/10.1080/10400410701841955

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21(1), 25–36. https://doi.org/10.1177/0146621697211002

Scoular, C., & Care, E. (2020). Monitoring patterns of social and cognitive student behaviors in online collaborative problem solving assessments. *Computers in Human Behavior, 104,* 105874. https://doi.org/10.1016/j.chb.2019.01.007

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior, 29*(6), 2568–2572. https://doi.org/10.1016/j.chb.2013.06.033

Ventura, M., Shute, V., & Zhao, W. (2013). The relationship between video game use and a performance-based measure of persistence. *Computers & Education, 60*(1), 52–58. https://doi.org/10.1016/j.compedu.2012.07.003

Wang, Y., Kai, S., & Baker, R. S. (2020). Early detection of wheel-spinning in ASSISTments. In *International conference on artificial intelligence in education* (pp. 574–585). Springer, Cham. https://doi.org/10.1007/978-3-030-52237-7_46.

Whitmer, J., Pedro, S. S., Liu, R., Walton, K. E., Moore, J. L., & Lotero, A. A. (2019). The constructs behind the clicks. ACT Research Report, 26.

Wolters, C. A. (2003). Understanding procrastination from a self-regulated learning perspective. *Journal of Educational Psychology*, 95(1), 179. https://doi.org/10.1037/0022-0663.95.1.179

Zhang, C., Huang, Y., Wang, J., Lu, D., Fang, W., Stamper, J., Fancsali, S., Holstein, K., & Aleven, V. (2019). Early detection of wheel spinning: comparison across tutors, models, features, and operationalizations. *International Educational Data Mining Society*.

Zhang, S., Bergner, Y., DiTrapani, J., & Jeon, M. (2021). Modeling the interaction between resilience and ability in assessments with allowances for multiple attempts. *Computers in Human Behavior, 122*, 106847. https://doi.org/10.1016/j. chb.2021.106847

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory into Practice, 41(2), 64–70.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at > springeropen.com