Research and Practice in Technology Enhanced Learning

RESEARCH Free and Open Access

Representing learning progression through activity monitoring with a focus on "learning from failure"

Taisei Yamauchi [1*], Heinz Ulrich Hoppe [2], Brendan Flanagan [3,5], Yiling Dai [4], Hiroaki Ogata [5]

*Correspondence: yamauchi.taisei.28w@st.kyoto-u.ac.jp
Graduate School of Social Informatics, Kyoto University, Kyoto, Japan.
Full list of author information is available at the end of the article

Abstract

This study addresses the challenge of characterizing learning progression in self-regulated learning environments only by activity monitoring. Existing studies often overlook self-regulated learning environments without predefined mastery criteria. The study fills a gap there by emphasizing learning from failure and offers a novel perspective on understanding learning dynamics in less structured settings. It empirically analyzes whether graded self-reporting can adequately capture students' learning progression using available student logs, and shows that graded self-reporting better reflects grades and the learning process quantitatively. It also introduces learning progress graphs generated from action logs to quantify and visualize learning progression, aiming to capture individual learning trajectories and to identify wheel-spinning in unguided contexts. Through two evaluations, involving educational technology researchers and experts, the study assesses the adequacy and interpretability of learning progression graphs for detecting wheel-spinning. Findings suggest that learning progression graphs effectively identify wheel-spinning on the basis of learning progression graphs, highlighting the importance of understanding student active steps without predefined target criteria. However, limitations include reliance on self-assessment, limited system recommendations, and the absence of teacher feedback. The study also notes challenges in interpreting learning progression without predefined guidance.

Keywords: learning progression, self-regulated learning, learning from failure, activity monitoring, wheel-spinning

Introduction

The term "wheel-spinning" has been introduced to characterize a situation in which a learner spends excessive time and effort in a learning environment without mastering the intended learning goals (Beck & Gong, 2013). Wheel-spinning, as a form of unproductive persistence, has been discussed and studied in the context of intelligent tutoring systems (Wang et al., 2020; Zhang et al., 2019) as well as in game-based learning (Owen et al., 2019). Most of these studies rely on specific mastery learning criteria (Pelánek, 2018). In contrast, we plead for characterizing learning progression on the basis of behavioral data in the form of action logs. This a both a more general approach (not only targeting wheel-spinning) and it can be more flexibly applied to many real learning settings. In our study, we conceive wheel-spinning as ongoing engagement and activity without positive progression, not necessarily only related to repeated failure. As a first step, we have tackled the issue of capturing learning progress in a way that allows for comparing and interpreting individual learning trajectories. This brought us to generating “learning progress graphs” (LPGs) from the action logs (Yamauchi et al., 2024). Although useful for many different purposes, the graphs can also be used specifically to identify wheel-spinning.

In the investigated scenario, junior high school students use an online platform for self-regulated learning (SRL) with mathematics exercises during their summer vacation. The data are from students in three different grades from the same school with an overall of 354 participants. The studied learning setting can be characterized as an extreme form of SRL (Owen et al., 2019) in which the students’ engagement is based on free choice without any kind of supervision or instruction during the learning period. Analyzing and measuring SRL has been identified as an important research challenge (Sacks & Leijen, 2014; Winne, 2010). The author (Yamauchi et al., 2024) has defined a formula for calculating learning progression based on the results of the exercise, depending on whether the result is success or failure. The developed wheel-spinning graph was designed to represent students’ learning progression. However, the study suggests that to more clearly convey students’ learning progression, it is important to distinguish between meaningful and meaningless failures.

In this study, we applied the concept of learning from failure (Jackson et al., 2022) and attempted to formulate the impact of the degree of failure on learning progression. Productive failure (Kapur, 2008; Kapur & Bielaczyc, 2012; Kapur, 2014) is the subject to be compared to our approach, but the purpose of using learning from failure in our study is to provide a unique perspective on how students learn from their experiences in unstructured environments, which is different from the learning design or instructional approach aimed at in productive failure. In addition, there is little empirical research on learning from failure (Darabi et al., 2018) while this study will provide new insights into this through the school practice.

We have targeted two aspects to evaluate the quality of LPGs for characterizing learning progression and specifically for identifying wheel-spinning. The first one is “expressiveness” as the information-richness of the model representation. I.e., does the LPG representation transmit the information that is relevant for characterizing the corresponding learning sequence? In this regard, we first explained that information is conveyed through case studies. Furthermore, we quantitatively discussed the effects of learning progression expression, particularly through gradation of failure, using progression, wheel-spinning, and examination grade as indicators, and using their correlations. Secondly, “adequacy” addresses the question in how far the model-based interpretation corresponds to human expert judgement of the cases. To evaluate the properties of LPG in terms of expressiveness and adequacy, and to check the appropriate way to express LPG, we have conducted an expert study with example LPGs based on real student data.

We will tackle these exercises by answering the two RQs:

RQ1: Can learning progression graphs express students’ unguided learning behavior, considering well the learning effect of failure?

RQ2: Between different variants of LPGs – which one is more adequate, also considering the instructional monitoring support?

Background and Related Work

Our aim in this research is to find a way of interpreting “wheel-spinning” in the context of unguided self-regulated learning (SRL).

Assessment and monitoring of SRL

SRL is typically defined as self-generated thoughts, feelings, and behaviors that are planned and cyclically adapted to achieve personal learning goals (Zimmerman, 2000). Self-regulation involves the capacity to cultivate knowledge, skills, and attitudes that are applicable across various learning environments and can be transferred from the contexts in which they were acquired to both leisure and professional settings (Boekaerts, 1999). The field of learning analytics, which focuses on interventions and learning processes, provides valuable opportunities to analyze SRL and yield meaningful insights (Roll & Winne, 2015). SRL activity consists of a sequence that includes (1) activation of perceptions through goal setting, (2) metacognitive awareness through monitoring SRL processes, (3) control of the various aspects of the self, and (4) reflections on the self (Pintrich, 2000).

The establishment of a computerized learning environment will significantly advance the science of learning about self-regulated learning and is considered essential for the application of this knowledge to educational settings (Winne, 2010). Intervention has also been shown to be effective for SRL, and the general SRL model with intervention characteristics is of incremental and cyclical in nature, including phases of preparation, performance and appraisal. The present study also fits into this context in that the students are given a particular mathematical problem to solve (preparation), they solve it (performance), and they are free to reflect on the results using the intervention (appraisal) (Edisherashvili et al., 2022). However, our scenario does not allow for the clear identification of SRL phases due to absence of an externally imposed process structure. Hatala et al. (2023) describe a pattern of self-regulated learning for students in the domain of programming, showing what processes produce good results for self-regulated learning and how stable they are.

Interdisciplinary cooperation between the fields of learning science, learning analytics, and AI is essential for integrating data-driven approaches with theory-driven verification metrics (Molenaar et al., 2023). The context that we focus on in this study is the most autonomous form of self-regulated learning, in contrast to self-directed learning, in that students work on the teacher-prescribed tasks and additional problems given to them during their school's summer holidays in a completely free environment (Sacks & Leijen, 2014). In this study, we used a self-regulated learning context assessed by whether the goals set were too modest, based on activity monitoring that evaluates students' self-reports of the problems they solved. We do not distinguish internal elements of the SRL process but rely on analyzing the dynamics of student activity.

Wheel-spinning and learning progression

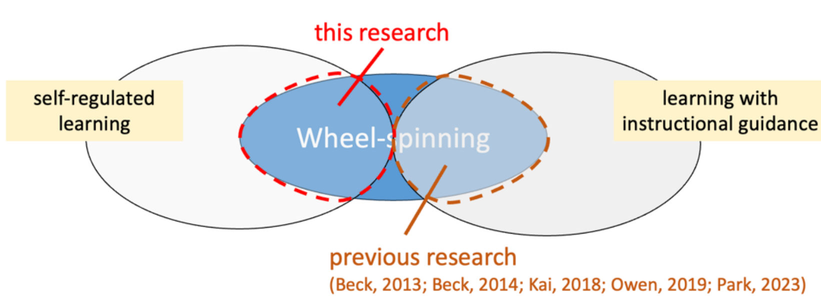
The notion of wheel-spinning refers to a situation in which learners spend considerable time learning a topic without actually achieving mastery (Beck & Gong, 2013) or substantial progression. This concept has been applied to the studying-learning interaction with intelligent tutoring systems where students may be stuck in the mastery learning cycle without achieving mastery criteria (Beck & Rodrigo, 2014). This notion of wheel-spinning is intimately related to the idea of mastery. Note that all notion of mastery level or wheel-spinning is different from the context of the research or the situation in which the subject is placed. Wheel-spinning, often regarded as unproductive persistence, is also compared with productive persistence.

In the educational context, we have a number of prior definitions and thresholds of mastery and wheel-spinning. Previous studies define mastery prior to wheel-spinning (Beck & Gong, 2013; Beck & Rodrigo, 2014; Kai et al., 2018; Owen et al, 2019; Palalog et al., 2016; Park, 2023). Beck & Gong (2013) and Park (2023) set the wheel-spinning threshold as the status of that student who practiced the same skill set over 10 times but failed to achieve mastery. Beck & Rodrigo defined wheel-spinning as a situation in which students fail to master a skill after having used a computer tutor with a certain effort and during a certain time. Kai et al. (2018) defined wheel-spinning as a type of unproductive persistence where students spend too much time struggling without achieving mastery of skills. Specifically, it refers to students who either do not pass a retention test after initially demonstrating mastery (by answering three problems correctly in a row on or after the 10th problem in a problem set) or who complete 10 problems without achieving mastery and thus never receive a retention test. In a game-based learning environment, Palalog et al. (2016) set a different definition of wheel-spinning as a non-learning behavior characterized by continuous but futile effort, where students fail to achieve mastery in a timely manner. Specifically, in the context of the Physics Playground game, wheel-spinning is identified when students are unable to successfully complete a level after multiple attempts (8 or more) or after spending a significant amount of time (more than 160 seconds) on it. In Owen et al. (2019) paper, wheel-spinning is defined as a form of unproductive effort where students spend a considerable amount of time struggling to learn a topic without achieving mastery. It is associated with reduced motivation and avoiding asking for help when needed. The paper differentiates wheel-spinning from productive persistence, which involves steady progress and eventual success despite challenges. Most studies focused on wheel-spinning as the attempts/effort without mastery, while no other papers research wheel-spinning in the environment of no instructional guidance that does not have a specific mastery.

In this research, first, we defined progression based on each exercise's difficulty and wheel-spinning as no positive change of progression. That means wheel-spinning is not necessarily related to repeated failure but would also be identified based on missing increments in difficulty and variation of learning tasks (exercises). Our approach aims to putting wheel-spinning in the broader context of identifying and quantifying learning progression through monitoring behavioral dynamics (see Fig. 1). Combined with a specific visualizing method, this provides a new way to characterize wheel-spinning.

Figure 1

Comparison of research on wheel-spinning



Learning from failure

The concept of "learning from failure" has gained significant attention in educational research, with various studies exploring its potential benefits and challenges (Jackson et al., 2022). A common theme among these studies is the examination of emotional and cognitive barriers that hinder learning from failure. Previous research has shown how failure can threaten self-esteem and motivation, making it difficult for individuals to engage with and learn from their mistakes (Carlson & Fishbach, 2024; Eskreis-Winkler & Fishbach, 2019; Eskreis-Winkler & Fishbach, 2022). These papers emphasize the need for interdisciplinary approaches, integrating insights from psychology, education, and organizational behavior to develop strategies for overcoming these barriers.

The notion productive failure (PF) underlines the role of failure as a powerful source of learning. Kapur (2008) highlighted PF as a pedagogical design or an instructional approach, where students are encouraged to generate solutions to complex problems before receiving formal instruction. Studies (Kapur, 2008; Kapur & Bielaczyc, 2012; Kapur, 2014) have focused on how engaging students in problem-solving without initial guidance can lead to deeper understanding and improved problem-solving skills. For example, Kapur (2014) argues that PF enhances conceptual understanding and knowledge transfer more effectively than merely evaluating others' solutions, while Kapur & Bielaczyc (2012) demonstrate that PF students outperform their direct instruction counterparts in posttests involving well-structured and complex problems.

Furthermore, several studies acknowledge the limitations of their research, such as specific sample settings, potential biases, and the need for further investigation. Studies (Kapur, 2008; Kapur, 2014; Kapur & Bielaczyc, 2012; Eskreis-Winkler & Fishbach, 2019; Darabi et al., 2018; Caliskan, 2021) call for more comprehensive studies to validate findings and explore broader applications. For instance, Darabi et al. (2018) highlight the scarcity of experimental research on learning from failure and advocates for more robust empirical foundations to better understand its effectiveness as an instructional strategy. Moreover, the observation that the ability to identify significant failure differs by school level, while many school policies generally regard failure as an opportunity for learning (Caliskan, 2021), further supports the need of empirical studies.

Unlike previous research that has primarily focused on structured environments with mastery criteria, this study explores self-regulated learning (SRL) contexts without predefined guidance. "Learning from failure" in the context of this paper is treated as an observational and analytic category, providing a unique perspective on how students navigate and learn from their experiences in less structured settings. This empirical approach offers valuable insights into optimizing instructional strategies and fostering a culture that embraces failure as a learning opportunity.

Method

System environment and application context

The experimental setting for our empirical research makes use of the digital learning platform LEAF (Flanagan & Ogata, 2018). System users first enter the environment by logging in to the Moodle LMS, and then proceed to the BookRoll platform which includes an analysis tool called LogPalette. All logs both at BookRoll and at LogPalette are collected in a learning record store (LRS), and the systems can also use these logs for analysis. The BookRoll system gives access to pooled exercises. Regarding the exercise materials, there are three types of pooled materials. One type consists of the exercises prescribed by teachers, another type comprises the exercises that students take on as additional exercises, and the other consists of the textbook pages that help to solve these exercises. Each page of the learning material was labeled as one unit in advance by publishers or mathematics teachers, based on the Japanese mathematical curriculum prescribed by MEXT (Ministry of Education, Culture, Sports, Science and Technology).

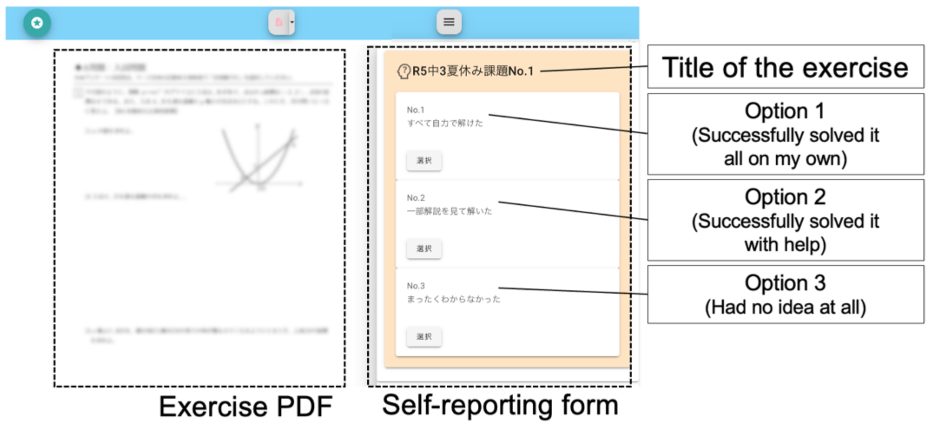
Learning scenario and data capture

This research was conducted in a Japanese junior high school with 354 participants from grades 7-9 (12-15 years old). They were all on summer vacation during the research (from July 12 to August 25, 2024). They were provided with teacher-customized exercises through BookRoll as assignments, and their results were recorded in the LRS. Teachers provided a single e-book containing multiple exercises. Although the exercises were presented in a set order, students were free to complete them in any sequence they preferred. This was not guided by a tutoring system. We observed that some students still chose to work on the exercises in the order they were presented. This activity can be categorized as self-regulated learning, since the students' activity was based on free initiative, and they did not receive any additional instructions during the respective period (Sacks & Leijen, 2014).

For the general exercise activity, students reported the result of solved exercises by entering whether their answers were correct or incorrect on the selective form, referred to Fig. 2. Students were not forced to work on the prescribed exercises, i.e., whether they were taken up and completed or not depended on the free choice of the students. Students could also use recommendation system, either drawing from the pool of additional exercises or as new attempts with exercises that were not completed correctly and without using help information previously. In addition, the system also recommended textbook pages to the students in case of repeated failure. They could solve other exercises by themselves as self-regulated learning on the system, or even without the system.

Figure 2

User interface for viewing and answering exercises on BookRoll. The option "Successfully solved it with help" in the figure was selected when the student reach correct answer with referring to textbook, other exercises or other supports.



There is a total of 333,466 logs by 238 students in the subjective course during the experiment, 93,539, 121,001, 118,745 logs of which were created by grade 7, 8, 9 students, respectively. Each log was associated with each action taken by the student, and the breakdown of these logs is as shown in Table 1. When the student opened or closed a content in the BookRoll, the action "start" or "end" was recorded. When the student turned the page back and force, then the action "move" was recorded. When students answered an exercise, created a memo, highlighted the content, added a bookmark or drew an image, then the action "answerd", "created memo", "highlighted content", "added bookmark" or "drew image" was recorded, respectively.

For the analysis, we excluded the logs between a page movement and next page movement if the student took less than 10 seconds and no other actions like "answerd", "created memo", "highlighted content", "added bookmark" or "drew image". From the remaining logs, we extracted 215,980 instances of consecutive attempts by the same student on the same page, each starting with "start" and ending with "end". Among these, there are a total of 7,097 exercise self-reporting logs by all students among the three grades. Of all self-reporting logs, 5,317 reported "successfully solved it all on my own", 1,532 did "successfully solved it with help" and 248 did "had no idea at all". This process was performed to handle multiple problem-answer logs within a single sequence collectively.

They participated in an examination after the summer vacation on August 26th, 2024. The content of the examination was different by student grades, consisting of the assignment that teachers provided with. In these tests, some of the exercises from the summer assignment are chosen by the teacher and set as test questions.

Table 1

Statistical information of the log data

Student action	Number of the logs
Regarding the start and end of content	28,415
Regarding the page movement of the content	166,464
Regarding exercise answer	25,615
Regarding memo creation	108,072
Regarding highlighting the content	2,100
Regarding bookmark attachment	2,623
Regarding image drawing	177

Operationalizing and quantifying learning progression

Following our aim to describe learning progression conforming to the free SRL context, we have defined "active" as a basic term and specified two formulae for the "difficulty" of exercises and for quantifying "learning progression".

To trace the student behavior through the activity, we took the activity logs of each student. In this research, the attribute "active" means the corresponding student has delivered (i.e., viewed and reported on) at least one exercise.

For each exercise Q_i in the exercises, we defined the *difficulty* D_{Q_i} as follows:

$$D_{Q_i} = \frac{N_{\bar{C}} + 1}{(N_C + 1) + (N_{\bar{C}} + 1)} \tag{1}$$

where N_C represents the number of people whose initial answer of the Q_i is "successfully solved it all on my own" and $N_{\bar{C}}$ represents the number of people whose initial answer of the Q_i is "successfully solved it with help" or "had no idea at all". As the more students could not solve the exercise on their own, the more difficult the exercise is. This definition derives from the number of correct or incorrect students plus 1; this is the smoothing whose value will be set to 0.5 when nobody is trying to answer the exercise.

We converted n-th self-reported results as the value R_n, which is between 0 and 1. The results of the students' self-reports are important for showing the situation of each student at each step. In the learning process of the students, if they were unable to complete an exercise, the situation of the students will differ depending on whether they "had no idea at all" or "successfully solved it with help" during the process of solving it, and these differences may affect the progression. Therefore, we decided to analyze the learning progression using two patterns: one where the results are divided into binary categories, and the other where the option of "successfully solved it with help" is taken into account. For the comparison of the effect on the student "mid-status" consideration for formulating learning progression, we introduced two converting ways to decide R_n, R_n^{binary} and $R_n^{ternary}$, each of which is defined as follows:

$$R_n^{binary} = \begin{cases} 1 \text{ (if “successfully solved it all on my own”)} \\ 0 \text{ (if “successfully solved it with help” or “had no idea at all”)} \end{cases} \quad (2)$$

$$R_n^{ternary} = \begin{cases} 1 \text{ (if “successfully solved it all on my own”)} \\ 0.5 \text{ (if “successfully solved it with help”)} \\ 0 \text{ (if “had no idea at all”)} \end{cases} \quad (3)$$

Finally, we calculated the progression ΔP_n with the current progression P_n and the previous progression status P_{n-1} as follows:

$$\Delta P_n = P_n - P_{n-1} = \begin{cases} 0 \left(\text{if } Q_n \text{ is repeated with most recent occurence } Q_k \text{ and } R_n \leq R_k\right) \\ \dfrac{(1-P_{n-1})D_{Q_n}R_n - P_{n-1}\left(1-D_{Q_n}\right)(1-R_n)}{2} \text{ (otherwise)} \end{cases} \quad (4)$$

Based on this stepwise calculation of differences, an LPG is constructed as a graph (or plot) with linear segments that accumulate the ΔP_n successively. Note that $P_0 = 0$.

The formula meets two important basic requirements: All P_n would be between 0 and 1, and failure on a more difficult exercise should be less negative than failure on easier exercise. In addition, we attempted to formulate a learning progression based on the assumption that simply repeating the same exercises, even if done successfully, is not desirable for learning advancement. This is why the first case of the formula (4) is separated. LPGs can be easily interpreted as indicators of learning progression based on the line gradients increasing, decreasing, or stable. As to be expected, the gradient is affected more by the result of a newer exercise, and a more difficult exercise generates less decrease (proportional to $(1-D)$) in the negative case but higher increase (proportional to D) in the positive case than an easier one.

Visual representation

Visual methods in education are gaining popularity, yet additional research is required to evaluate their enhanced contributions in terms of effectiveness, efficiency, and other learning-related criteria (Klerkx et al., 2014). Although we are not interested in content visualization, one study suggests the visualization of a timeline with learning events, which can organize the complicated multiple classroom events more easily (Dillenbourg et al., 2018). Applying visualization in the SRL context is also conducted by previous research, which makes the process clearer with the graph path (Zheng et al., 2021).

Our approach introduces a new graph visualization for grasping students' learning progression and possibly wheel-spinning behavior based on active sequences with the difficulty of the exercises using line diagrams with additional markers. Figure 3 shows one example of a student's active sequence graph. The x-axis of the graph represents incremental discrete steps of activity (such as self-reporting the results by working on learning materials and clicking the link to the recommended materials or the exercises on the assignment list) that we call “active steps.” The y-axis of the graph represents the difficulty of each exercise as explained above. The markers represent the active work on the exercise. The number attached to the marker represents the ID of the learning material. The color of the marker represents whether the student reported success (green), mid-status (yellow) or failure (red). If the same problem had been solved in the past, a black border was placed on the marker. This black border was represented thicker when a self-report with improved understanding was given, for example, in question 128 in Fig. 3.

In analyzing students’ learning progression, we identified a limitation that we refer to as the saturation effect. This effect occurs when students successively solve very easy problems correctly on the first attempt: the change in learning progression ΔP_n becomes so small that the curve appears nearly flat, making it difficult to visually detect meaningful progress (Yamauchi et al., 2024). Fig. 5(a) illustrates this problem. Although the student continues to make progress, the curve gives the misleading impression of stagnation because the increments are too subtle to be observed.

To address this issue, we applied a transformation of the curve based on exponentiation of ΔP_n. Specifically, we defined y-axis value on the graph as $(\Delta P_n)^\delta$ if $\Delta P_n > 0$, and as 0 otherwise, where δ is a constant satisfying $0 < \delta < 1$. This transformation amplifies small positive changes in ΔP_n, while treating negative values as zero. As shown in Fig. 5(b), this adjustment makes small improvements more visible, thereby mitigating the saturation effect. Furthermore, “no progression” can be defined as the state where $y = 0$ persists on the graph. Plotting the progress of the student featured in Fig. 3 according to this approach yields the graph shown in Fig. 4. A previous paper follows this effectiveness, which states that graph interpretation accuracy improves when the slope-mapping constraint is honored, meaning the queried variable is on the vertical axis, allowing steeper lines to represent faster changes (Gattis & Holyoak 1996).

For experts or teachers, this transformation means that the skewed curves can be interpreted more reliably. Instead of overlooking minor but genuine progress, they can now recognize and evaluate subtle improvements in students’ learning. At the same time, the criterion for identifying potential wheel-spinning behavior becomes clearer, providing a more consistent basis for monitoring student performance and offering feedback.

We tried to see where the students weren’t progressing in their learning behaviors, so we set red areas in the graph if there is three-successive 0 or less gain in P_n, which can work as one of the wheel-spinning criteria.

For the holiday scenario from which our data were harvested, proportional scaling of the events over a timeline is not adequate since there would be long phases of inactivity and very short and dense bursts of active work with the exercises. Our choice focuses on the sequences of actions and their evolution for individual students. For classroom scenarios within the time limitations of a lesson, a proportional timeline could be used.

Figure 3

Example of a graph with y-axis of learning progression that shows the sequence of a student's entire active steps

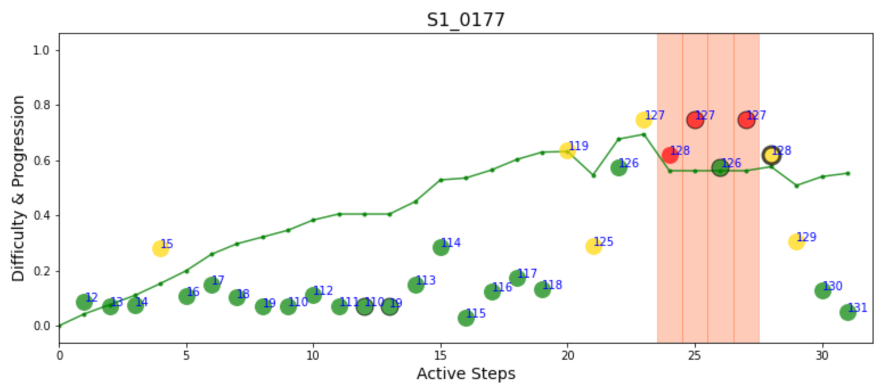


Figure 4

Example of a graph with y-axis of ΔP_n that shows the sequence of a student's entire active steps

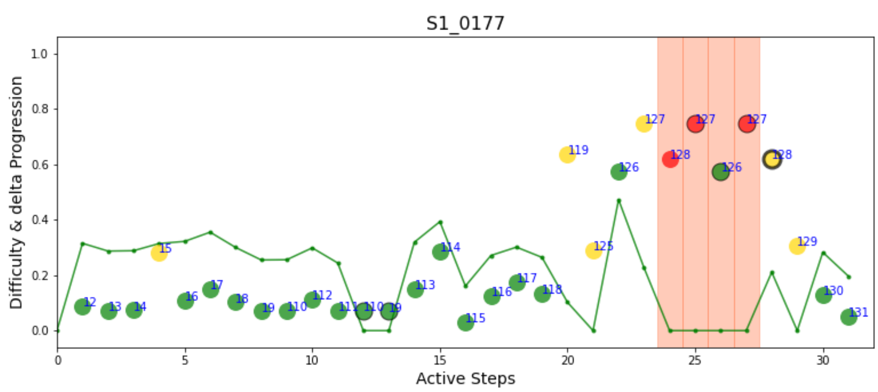
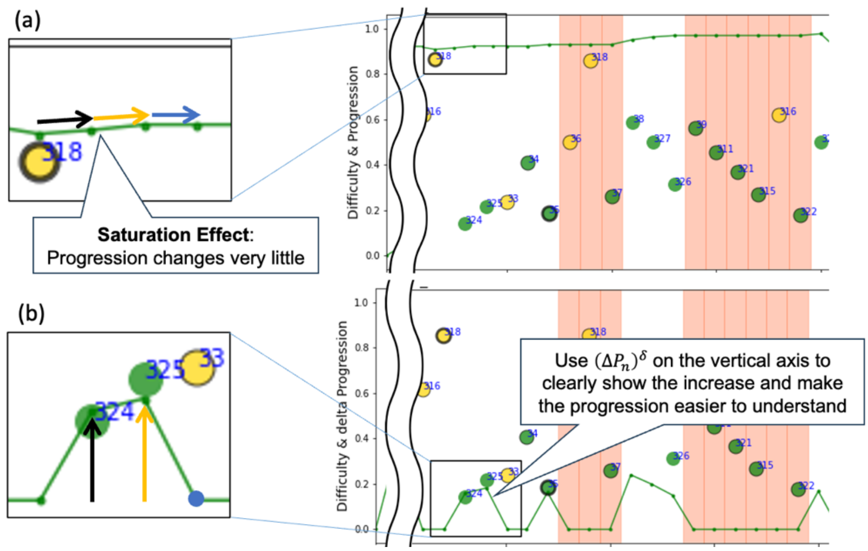


Figure 5

Two representation of a student's learning progression, with saturation effect on a graph ((a), drawn using the same method as Figure 3) and proposed solution ((b), drawn using the same method as Figure 4)



Introduction of indicators for evaluation of formula

For the evaluation of the formula for learning progression, it would be important factor to have ability to describe a student current status better. For the comparison of the calculation using R_n^{binary} or $R_n^{ternary}$, we created five kinds of indicators to quantitatively examine the relationship between learning progression, wheel-spinning, and grades.

Overall progression is inevitably important for student learning status. After calculating the learning progression for each active step, the sum of the progressions was calculated as "amount_progression". For the comparison to "amount_progression" from the perspective of indicators which better indicates students' learning outcomes, we introduced the variable "amount_active" to simply capture the total number of active steps taken by students. Additionally, the deviation values of the grades were calculated for each grade level and were defined as "deviation". Why we didn't use the test scores was the fact that different grades worked on different problems and took different final tests. The following formula represents how "amount_progression" and "deviation" were calculated, where $score_{student}, mean_{score}, std_{score}$ represents the student's score, mean of scores of all students, standard deviation of scores of all students, respectively:

$$amount\_progression = \sum P_n \tag{5}$$

$$deviation(student) = 50 + \frac{10 \times (score_{student} - mean_{score})}{std_{score}} \tag{6}$$

Evaluation of the utilization of learning progression graphs for wheel-spinning detection

To test the adequacy and interpretability of the wheel-spinning visualization, we took two evaluations.

The first evaluation was conducted for evaluation of learning progression graph from the perspective of adequacy. It was conducted by ten researchers specializing in educational technology, who were randomly selected from among experienced researchers familiar with online learning, educational technology, and data analysis, and who were willing to complete the questionnaire. We presented participants with six randomly selected examples, some of which included a wheel-spinning area, while others did not. We first provided a definition of wheel-spinning and explained how to interpret the graph, using six randomly selected examples based on the wheel-spinning criteria described above (treated as a “working hypothesis”). Participants were then asked to complete six questions about the wheel-spinning areas of the graphs, along with indicating their confidence in each answer. The confidence is assessed with 5-point Likert scale. We also asked to describe adequacy and interpretability of the graph.

The second evaluation was based on the first evaluation and aimed at assessing the improvement of the saturation effect. First, four experts, who were randomly selected and were willing to answer the questionnaire among skillful researchers familiar with online learning educational technology and data analysis, were asked to rate the wheel-spinning and confidence level based on the graph created from the values of “active step” and “progression (P_n),” as shown in Fig. 3, with 6 examples. Next, the same four experts were asked to rate their confidence in wheel-spinning on a graph created from the values of “active step” and “delta progression (ΔP_n),” as shown in Fig. 4, with 6 examples. Finally, they were asked to freely write their opinion about which one improves the saturation effect more when comparing the two charts. These four experts’ opinions were additionally evaluated by another expert, if the judge of the four are even.

For details on the contents of the questionnaires, please refer to the supplementary file.

Results

Case-based analysis

Fig. 6 shows one example of learning progression graphs by one student logs with $R_n^{ternary}$. In the figure, the student reported “having no idea at all” on the exercise 114 with the difficulty of 0.3, and also reported on exercise 126 with the difficulty of 0.6. The gradient of the learning progression line at exercise 114 is steeper than that of exercise 126. This means, learning progression graph can represent the difficulty of the exercises, which is that the more difficult question the student made failure, the less steep the learning progression graph is.

The student reported "successfully solving it with help" on the exercise 119 with higher difficulty than the learner's current learning progression, and also reported on exercise 124 with lower difficulty than the learner's current learning progression. The gradient of the learning progression line at exercise 119 is positive, while that of exercise 124 is negative. This means, learning progression graph can be interpreted according to the difficulty of the exercises which is reported as "successfully solved it with help". If the student has solved more difficult exercise successfully with help, learning progression represents that the student made deeper understanding of such a difficult exercise. On the other hand, if the student has solved less difficult exercise successfully with help, learning progression represents the student asked for help on the exercise which should be solved successfully on the student's own.

Figure 6

Example of a learning progression graph with $R_n^{ternary}$ including different gradient of learning progression at each time of exercise solving

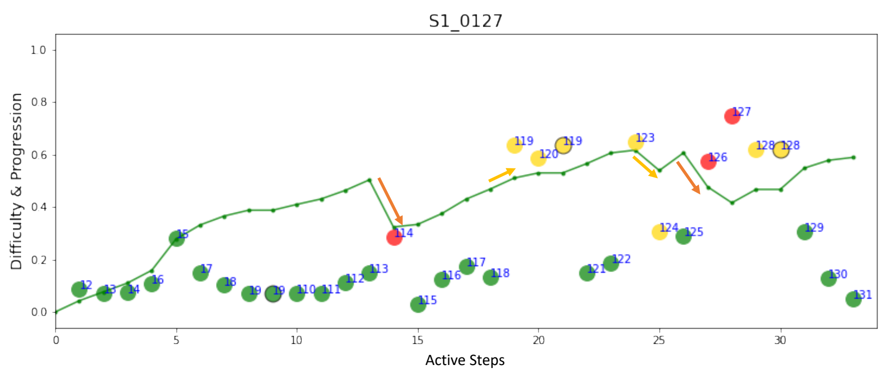
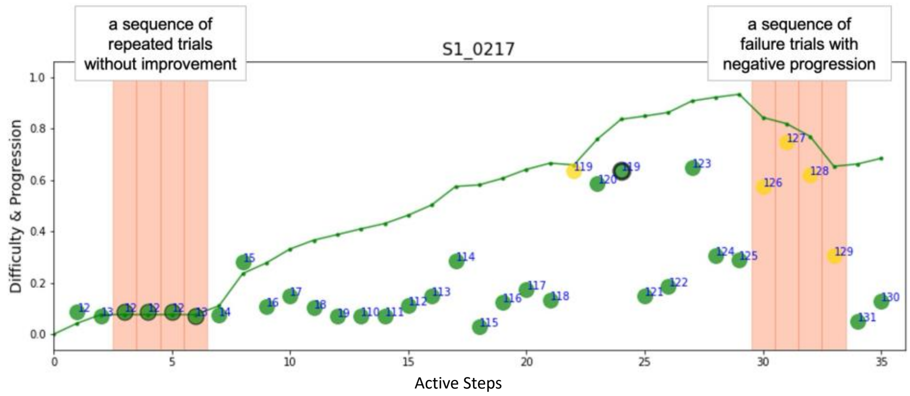


Figure 7 shows another example of learning progression graphs by one student logs with $R_n^{ternary}$. In the figure, there are 2 areas that indicates wheel-spinning. The left of wheel-spinning area represent that the student was trying to solve the same exercises successively many times. The right area represents the three or more successive failure. This representation indicates that we could treate both cases as wheel-spinning phenomenon, which the graph shows as red-shaded area.

Figure 7

Example of a learning progression graph with $R_n^{ternary}$ including two types of wheel-spinning area



The following two figures, Fig. 8 and 9 show graphs of trials by the same student in which self-reported results are determined for two cases, R_n^{binary} and $R_n^{ternary}$, respectively. In Fig. 8, wheel-spinning occurs because "solved successfully with help" and "had no idea" are treated as failures. On the other hand, by gradating "solved it successfully with help" and "had no idea," it was found that progress in understanding could be seen even in the midst of wrong approaches. For example, for question 126 in the graph, the first effort was "had no idea", but the next effort was "solved it successfully with help". In addition, the graph in Fig. 9 indicates that the situation was not considered wheel-spinning. Based on these results, it can be concluded that judgment aligns with an increase in students' understanding.

Figure 8

Example of a learning progression graph with two-option self-reported results R_n^{binary}

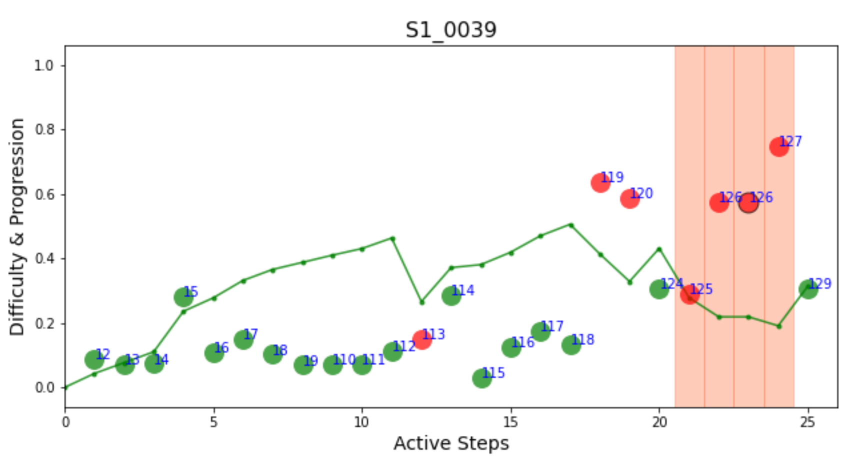
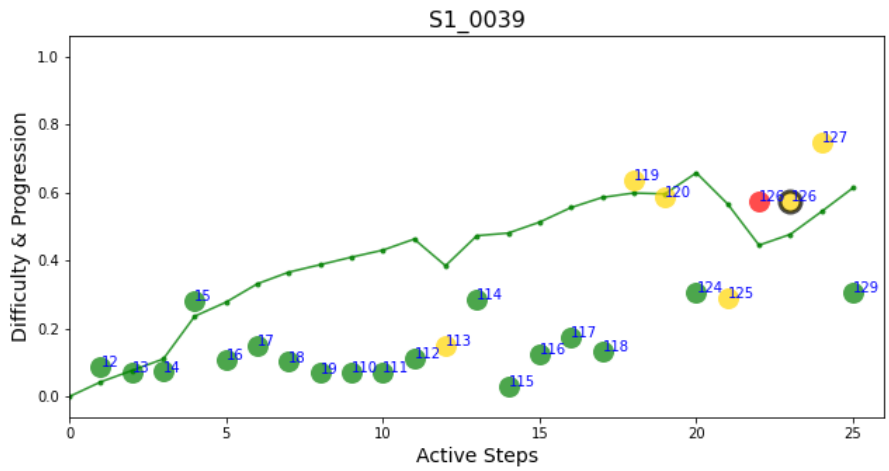


Figure 9

Example of a learning progression graph with three-option self-reported results $R_n^{ternary}$



Correlation analysis

To compare whether failure from learning is better represented, we created three specific indicators: "amount_progression," "amount_active," and "deviation." We first search the correlation between "amount_active" and "amount_progression" to grasp if the progression represents the features of the datasets we used. We employed "deviation," a metric based on exam scores outside of summer break activities, as a quantitative evaluation of learning progression. This enabled us to compare how well the different methods of calculating learning progression reflect subsequent student performance. By examining correlations between "amount_active" and "deviation," or "amount_progression" and "deviation," we aimed to gain insights into the relationship between learning progression indicators and overall student outcomes.

For the two indicators "amount_active" and "amount_progression," we found that both using R_n^{binary} and using $R_n^{ternary}$ had strong correlations, but the correlation was stronger when using $R_n^{ternary}$ (Fig. 10 (a), (b)). The reason for this is thought to be due to the characteristics of the data set used, rather than a change in the learning progression calculation formula. Looking at Fig. 10, if we focus on the points around 10 ≤"amount_active"≤ 60 and 0 ≤"amount_progression"≤ 0.3, there are many points in Fig. 10 (a), but almost none in (b). In this data set, only 3.5% of all self-reportings were "had no idea at all", and 86.1% of the self-reports that were considered "failure" in R_n^{binary} were treated as "successfully solved it with help" in $R_n^{ternary}$, so this seems to be why there have been few students with extremely low learning progression calculated by $R_n^{ternary}$.

Furthermore, when we looked at the correlation between "amount_active" and "deviation" for students who had wheel-spinning periods, we found a weak correlation of around 0.2

for both R_n^{binary} and $R_n^{ternary}$ (Fig. 11 (a), (b)). On the other hand, when looking at the correlation between "amount_progression" and "deviation", a correlation of 0.34 and 0.43 was seen when using R_n^{binary} and $R_n^{ternary}$ respectively (Fig. 11 (c), (d)). From these results, we can see that progression based on difficulty and repetition is a better reflection of students' comprehension than simply working through a large amount of material. Furthermore, the p-value for the correlation between "amount_progression" and "deviation" is less than 0.001, which is extremely low compared to the p-value for the correlation between "amount_active" and "deviation", showing that the results are more reliable as a reflection of students' comprehension. In addition, the correlation between "amount_progression" and "deviation" differs depending on whether the calculation method used is R_n^{binary} or $R_n^{ternary}$. In Fig. 11 (c), there are many students with high grades even when $0 \leq$"amount_progression"≤ 0.2, but in Fig. 11 (d), this is not the case, and rather, students with high grades are concentrated to some extent in the $0.8 \leq$"amount_progression"≤ 1.0 range. From this, we can see that using $R_n^{ternary}$ is more likely to reflect the students' understanding than using R_n^{binary}.

Figure 10

Correlation analysis between the amount of active steps and the amount of progression calculated by (a) R_n^{binary} (left) or (b) $R_n^{ternary}$ (right)

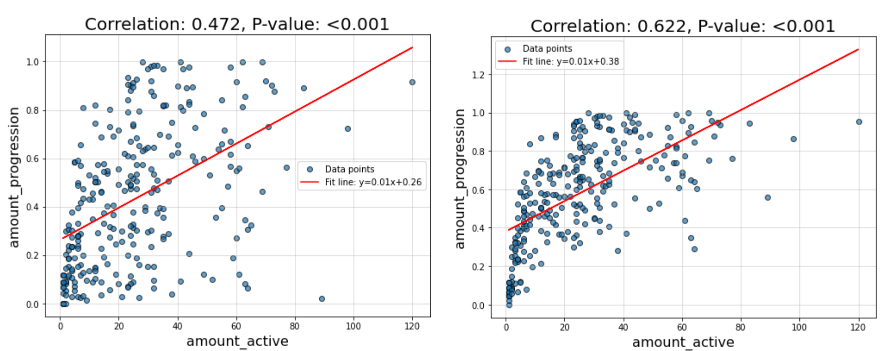
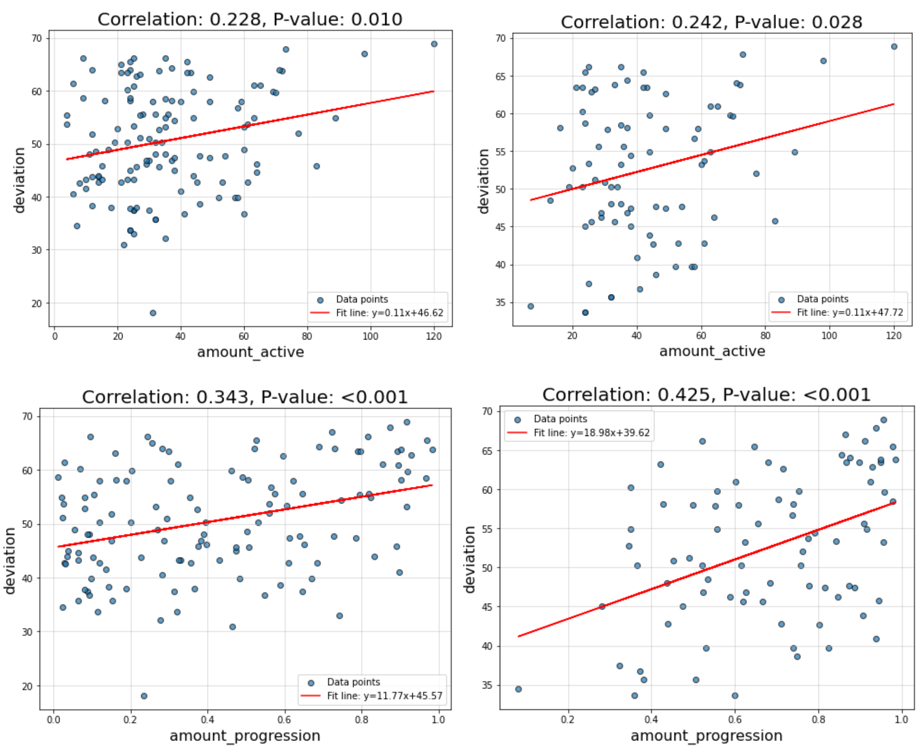


Figure 11

Correlation analysis of deviation of the exam score of the students who have had wheel-spinning periods. Correlation was calculated between deviation and (a) amount of active steps by R_n^{binary} (upper left), (b) amount of active steps by $R_n^{ternary}$ (upper right), (c) amount of progression by R_n^{binary} (lower left), (d) amount of progression by $R_n^{ternary}$ (lower right)



Expert assessment

First evaluation

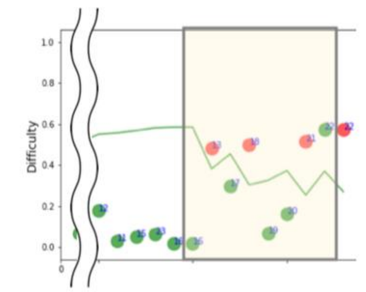
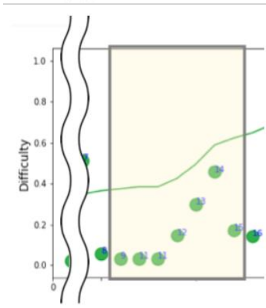
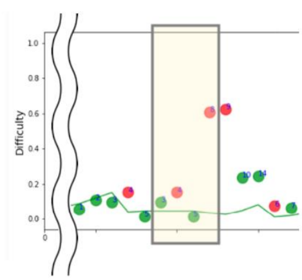
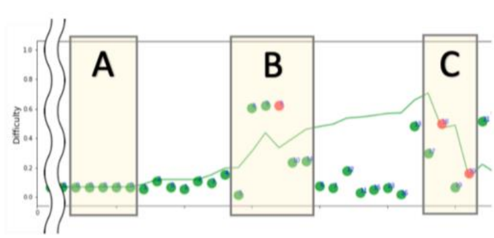
Table 2 shows all the questions from the first evaluation along with their corresponding answers. In the table, we show the quiz number, target area of example questions, the number of those who recognized the wheel-spinning area or non-wheel-spinning area with confidence, and the model's answer (working hypothesis referred in section 3.5). The overall agreement rate between expert responses and expected answers was 65%.

According to the survey results, adequacy is almost enough for participants because the correct answer is more than wrong answers for four questions, where Q3 and Q4-C is not. For Q3, there is no progression in that area, but half of the participants recognized it as non-wheel-spinning area. On the other hand, for Q4-C, there is positive ΔP_n and thus there is not wheel-spinning, but over the half of the participant recognized it differently. Although the graph is numerically correct, the visual interpretation may suffer from a saturation effect as a consequence of which is the ΔP_n is too small to be no longer recognizable. Therefore, we should reconsider this issue to improve the graph's adequacy.

We can also discuss the interpretability of the graph through the confidence and description of the questionnaire. In the survey, low confidence indicates a "lack of interpretability", which may be related to problems with the LPG but may also come from ambiguous situations. One of the participants stated that "*I agree, to assess repeating the same exercises again and again (whether green or red) can be categorized as wheel spinning. but failing exercises and needing several attempts, might be very helpful for learning - do you make such a distinction, or is all wheel spinning and negative? Might the duration of missing progress be important to judge something as problematic wheel spinning?*" and another stated "*In the annotated wheel-spinning domain, except for the challenge of extremely difficult problems, the problem seems to be solved in some stepwise manner, and in this respect, we felt that identification as wheel-spinning was not appropriate.*" Although the graph conveys wheel-spinning area well, we should reconsider the wheel-spinning phenomenon in the context of successive failure.

Table 2

Results of the first expert evaluation about wheel-spinning (WS)

#	Target area	Num of people (average confidence) WS	Num of people (average confidence) Not WS	Answer based on working hypothesis (agreement with experts)
Q1		3 (3.67)	7 (3.43)	Not WS (70%)
Q2		3 (3.67)	7 (4.14)	Not WS (70%)
Q3		5 (3.00)	5 (3.33)	WS (50%)
Q4-A		6 (3.67)	4 (4.25)	WS (60%)
Q4-B		0	10 (3.90)	Not WS (100%)
Q4-C		6 (3.80)	4 (3.50)	Not WS (40%)

Second evaluation

Table 3 shows all questions of the second evaluation with the answer of them. In the table, as with table 1, we show the quiz number, target area of example questions, the number of those who recognized the wheel-spinning area or non-wheel-spinning area with confidence, and the model's answer. The overall agreement rate between expert responses and expected answers was 39% for Q1 and 50% for Q2.

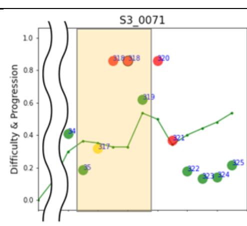
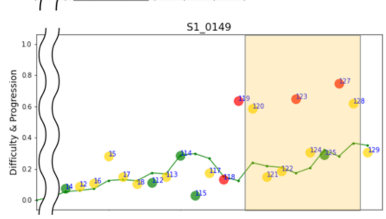
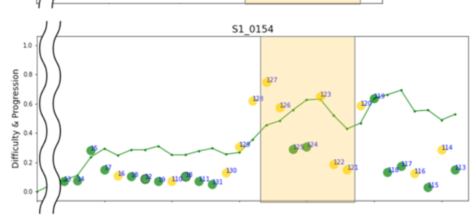
According to the survey results, there is room for debate about adequacy. In Q1-3, all four evaluators determined (with high confidence) that it was wheel-spinning, but in our opinion, this was not wheel-spinning because there was no decrease in learning progression at three or more consecutive points. This misunderstanding may have occurred because the gradient of the two consecutive decreases in learning progression was very steep, but in reality, the gradient was not used to determine wheel-spinning, so it is thought that the gradient was misleading. In Q1-4-A and Q1-4-B, the majority of the choices were the opposite of what was expected. One point worth noting is that there are some places in these graphs where it is not clear whether the gradient is positive or negative. These are the "saturation effects" mentioned in the previous chapter, and it is thought that these make it look as if there is no progression even when there is actually progression.

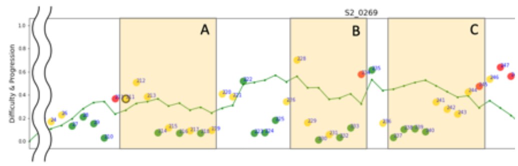
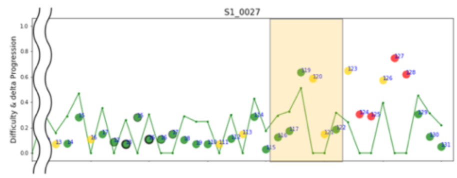
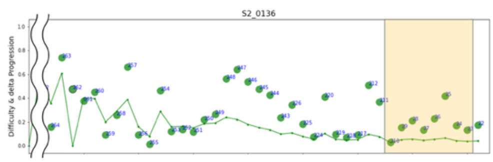
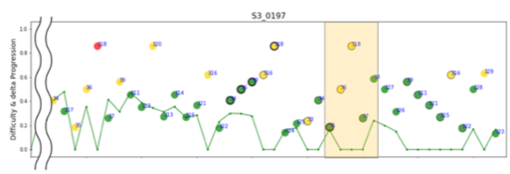
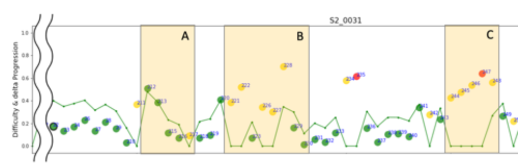
On the other hand, there were also some questions about Q2 that gave results that differed from our expectations. In Q2-4-A, many evaluators responded that wheel-spinning was occurring, contrary to expectations. This is thought to be because the gradient is continuously decreasing. On the other hand, in Q2-3, which has three successive zero delta-progression sequence, there were many evaluators who answered that wheel-spinning had not occurred. Looking at the results of Q2-4-C, which has four successive zero delta-progression sequence, even though the graphs are similar, the results are different, so it seems that the number of consecutive learning progression changes that are not considered to be wheel-spinning differs from ours. In other words, in our assumption, three consecutive non-positive delta-progression is wheel-spinning, but four may be a better indication of wheel-spinning. In addition, in Q2-2, there were more evaluators who answered that it was wheel-spinning. This is thought to be because the delta progression of the latter half of the exercises became very small, as the students were able to solve all the exercises correctly within the range. This is the only saturation effect in this graph representation that can be considered. This is because if the students report even one mistake, there will be room for progression.

We also asked the evaluators to rate the graphs that were more suitable for wheel-spinning and asked them to give their reasons. As a result, three of the evaluators said that the graph expression used in Q1 was more suitable, while only one person said the same for Q2, which was different from our expectations. One of the evaluators who said Q1 was better explained their choice by saying, "*The green line graph shows a certain trend, and it seems to follow the trend of most of the marker plots*," while for the graph representation of Q2, they answered "*The green line graph fluctuates wildly, so it is difficult to grasp the trend of learning progress. In addition, the wild fluctuations in the graph do not seem to* *match the marker plots.*” From this, it can be assumed that this evaluator was not looking at just one element of the graph but was focusing on an intuitive understanding of the graph as a whole. The evaluator who answered that Q2 was better responded to the graph expression in Q2 as follows: “*Since it shows changes in learning progression, if it is at 0, it can be judged as ‘wheel-spinning’, and it is easy to understand, like a progress chart when downloading from a computer.*” On the other hand, the evaluator who answered that Q1 was better responded to the graph expression in Q1 as follows: “*Even though learning progression has dropped, there are parts that are not judged as ‘wheel-spinning’, so it is difficult to judge.*” This response is in line with our intentions for the graph representation, and it shows that some evaluators realized that it is enough to focus on specific parts of the graph. From the responses to each reason, it can be seen that there are still advantages and disadvantages to both graph representations. In other words, for the graph representation of Q1, the point that a decline in learning progression is not necessarily considered to be wheel-spinning, and for the graph representation of Q2, the point that the overall learning trend is difficult to understand intuitively, should be improved.

Table 3

Results of the second evaluation about wheel-spinning (WS)

#	Target area	Num of people (average confidence)		Expected Answer (agreement with experts)
		WS	Not WS	
Q1-1		3 (3.00)	2 (3.00)	WS (60%)
Q1-2		3 (2.67)	1 (4.00)	WS (75%)
Q1-3		4 (3.75)	0	Not WS (0%)
Q1-4-A		3 (2.67)	1 (3.00)	Not WS (25%)

Q1-4-B		4 (4.25)	0	Not WS (0%)
Q1-4-C		3 (4.33)	1 (2.00)	WS (75%)
Q2-1		1 (5.00)	3 (4.00)	Not WS (75%)
Q2-2		3 (4.00)	2 (4.50)	Not WS (40%)
Q2-3		1 (5.00)	3 (3.67)	WS (25%)
Q2-4-A		3 (4.33)	1 (5.00)	Not WS (25%)
Q2-4-B		2 (4.00)	3 (3.33)	Not WS (60%)
Q2-4-C		3 (3.67)	1 (4.00)	WS (75%)

Discussion

RQ1: Can learning progression graphs express students' unguided learning behavior?

Through this study and approach, we have connected the area of SRL research with behavioral data analysis of students' activities as suggested by Molenaar et al. (2023). We observed learning progression in SRL by focusing on behavioral dynamics, that is, students' activities over time. The long-term learning analysis yielded different outcomes depending on students' time-use patterns in SRL (Rienties et al., 2019). It may also represent a new discovery in the field of learning from failure, an area that has not been fully explored in empirical research (Darabi et al., 2018), particularly in terms of analysis based solely on observed student activity.

This study incorporated the perspective of learning from mistakes, and by establishing gradations in the trends of students' errors, it sought to more accurately grasp students' understanding. First, we found a weak correlation between the number of active steps during the summer vacation period and grades. On the other hand, the learning progression, which was formulated by taking into account the failure of exercises and the operation of repeatedly solving the same exercises, was found to correlate more strongly with students' grades than the sheer number of active steps. This suggests that while there is a certain learning effect from the number of exercises solved, there is a limit to that effect, and that it is important to have a method that helps the student overcome their mistakes when reviewing. It goes without saying that studying a lot has a causal effect on test results or grades (Stinebrickner & Stinebrickner, 2008), but this conclusion is consistent with the results of previous research (Schuman et al., 1985) that the correlation between study time and grades is at best minimal.

This research contributes a new approach to analyzing SRL activities in the long span. According to the idea of progression curves of the visualization we proposed, there is no instruction or intervention during the monitoring period. Another perspective of this research is that students' performance can be enhanced if they engage in self-reporting even in the SRL context (Oudman et al., 2022). However, the literature on SRL introduces a "richer" context, with different phases, scaffolding and feedback (Edisherashvili et al., 2022; Pintrich, 2000), while this research based on observational data was conducted under specific constraints, characterized by an extremely autonomous form of SRL in the absence of teacher intervention and feedback during the activity period. It might be argued that this was different for the system-generated additional recommendation however these were used only very rarely.

RQ2: Between different variants of LPGs – which one is more adequate, also considering the instructional monitoring support?

The first survey results referred to section 4-3-1 indicate that the current LPG graph layout is affected by a "saturation effect". The graphical evaluations were conducted by experts. In cases like Q1-4A (expert agreement: 25%) and Q1-4B (expert agreement: 0%), the low agreement levels are believed to be due to the saturation effect. For Q1-3 (expert agreement: 0%), although there were downward trends in the graph, they were not sufficient to clearly support an interpretation of wheel-spinning across three cognitive regions. To address this issue, a revised graph layout like the one shown in Fig. 4 was introduced, which provides richer information and enables more differentiation in the analysis of progressive learning sequences. However, signs of the saturation effect still appear in the Q2-2 graph (expert agreement: 40%). In contrast, Q2-3 (expert agreement: 25%) and Q2-4A (expert agreement: 25%) showed alignment with participants who correctly understood the graph characteristics. This suggests limited but noteworthy potential rather than strong evidence of effectiveness.

These findings imply that the inconsistent judgments are not simply due to a lack of interpretability of the LPGs themselves, but rather to differences among experts in how wheel-spinning is understood and defined. No operational definition of wheel-spinning had been provided to the experts in the study because we wanted to evoke subjective judgements and not a replication of a reasoning that could be automated. Accordingly, the experts applied their own interpretations, and some divergence in their judgments had to be expected. Thus, while the LPGs can serve as a valuable basis for discussion, the contested nature of wheel-spinning constrains the consistency of interpretations.

Regarding graph design, both versions of the LPGs show advantages and disadvantages. The original version (learning progression on the y-axis) allows overall trends to be seen more intuitively, while the revised version (change in learning progression on the y-axis) provides clearer analytical criteria for identifying potential wheel-spinning and offers richer detail. Expert preferences were divided, and no statistical evidence supports a definitive claim of superiority. Nevertheless, the revised version appears promising as a more information-rich representation of learning progression, particularly if future studies include training participants in how to interpret the visualizations.

This interpretation is consistent with previous research on graph perception. Gattis and Holyoak (1996) have shown that accuracy in graph interpretation is enhanced when the slope-mapping constraint is followed, namely when the queried variable is placed on the vertical axis so that steeper lines represent faster changes. However, as our results show, even when such constraints are applied, discrepancies between intended interpretation and expert perception may remain, particularly when steep gradients or saturation effects occur. Moreover, Glazer (2011) highlights that the ability to interpret graphs is influenced by multiple factors, including the characteristics of the graph, its content, and the prior knowledge of the viewer. Berg and Phillips (1994) further demonstrated that interpreting changes in graphs requires a higher level of understanding than interpreting values, which may explain why some experts favored the revised graphs for their ability to highlight wheel-spinning, while others preferred the original for its more intuitive depiction of overall trends.

Future work should therefore explore ways to improve the handling of the saturation effect and investigate whether graph training can support more consistent interpretation. It would also be valuable to examine whether LPGs can assist teachers and other human experts in predicting learning success or failure, as suggested by the correlations reported in the lower row of Figure 11. While conclusive evidence of effectiveness has not yet been obtained, the revised LPGs demonstrate potential as a promising tool for representing learning progression, providing a richer basis for analysis, and guiding further refinement and educational application.

Conclusion

This research discusses the concept of "learning progression" in the context of instruction-free learning among junior high school students using an online math exercise platform during their summer holidays. To address the challenge of capturing learning progression, the study introduces "learning progression graphs" (LPGs) generated from action logs. Based on the observational data of the students' responses, the difficulty of the exercises derived from students' self-reporting and the learning progression defined by the type of students' self-reporting were found to be good indicators of the students' understanding of the exercises (i.e. their grades). It was also concluded that gradation of the students' failures serves as a better indicator of their progression.

Regarding the learning outcomes, this experiment relied solely on students' self-assessments in terms of self-reported success or failure. The results of solving exercises had to be analyzed based on the topics of study materials. By defining knowledge and attaching it as meta-information to the exercises that students solve, it should be possible to formulate learning progression based on knowledge states. The advantage of this formulation is that it allows the application of models that measure understanding through knowledge states, such as knowledge tracing (Corbett & Anderson, 1994) and ZPD (Vygotsky, 1978), to handle more detailed information, such as which knowledge is causing wheel-spinning. In fact, an approach has been proposed to predict student performance using the ZPD model in order to adapt teaching strategies (Chounta et al., 2017), but our research has the potential to expand on existing research by approaching the issue from a more observational perspective.

For students, LPGs can be used as feedback and reflection support on individual learning activities, allowing them to review their own learning activities. That is, if there are a series of errors in their answers, they may shift to easier questions or revisit supplementary materials such as textbooks. Teachers can monitor their students' efforts with LPG; they can refer to exercise difficulty to see which questions students are struggling with. They also can use this to plan what to teach and how to teach questions. In addition to educators, this visualization algorithm can also be integrated into educational recommendation systems and open learner models (Bull & Kay, 2010). By checking the difficulty of an exercise and the learning progression of a student, it will be possible to recommend easier or new exercises. In addition, as this visualization is created based on students' self-reporting, it can reveal patterns of wheel-spinning when combined with self-reported data. At the same time, prior work has shown that students may also engage in unproductive behaviors such as "gaming the system," progressing without genuine learning (Baker, Corbett, & Koedinger, 2004), which suggests that LPGs should be designed with attention to how learners might strategically respond to feedback.

Abbreviations

SRL: Self-Regulated Learning; LPG: Learning Progression Graph; AI: Artificial Intelligence; PF: Productive Failure; LRS: Learning Record Store; MEXT: Ministry of Education, Culture, Sports, Science and Technology; ZPD: Zone of Proximal Development; WS: Wheel-Spinning.

Acknowledgements

We would like to express our deepest gratitude to the participating teachers, whose practical insights were instrumental in bridging the gap between academic theory and classroom reality. Our sincere thanks also go to our technical staff; their expertise and meticulous attention to detail were fundamental to the success of this project.

Author's contributions

TY designed the experiment, performed data analysis and drafted the initial manuscript. BF, YD provided the support to implement the system and edited the manuscript. UH supervised the research overview and edited the manuscript. HO was responsible for funding acquisition and supervision. All authors read and approved the final manuscript.

Author's information

Taisei Yamauchi is a doctoral student at the Graduate School of Informatics, Kyoto University. Since 2025, he has served as a Research Fellow of the Japan Society for the Promotion of Science (JSPS). He earned his Bachelor of Engineering in Applied Chemistry from Kyoto University, followed by a Master of Informatics from the Graduate School of Informatics. His research interests lie in the fields of metadata labeling, exercise generation, personalized learning optimization, and learning analytics.

Heinz Ulrich Hoppe is an emeritus professor of "Collaborative and Learning Support Systems" of the University of Duisburg-Essen (Germany) and a fellow of the Asia-Pacific Society for Computers in Education (APSCE). His research is focused on computational techniques for analyzing and supporting collaboration, learning and knowledge building in various contexts. Currently, he is engaged as a senior researcher and head of the board of the Rhine-Ruhr Institute for Applied System Innovation (RIAS).

Yiling Dai is an Assistant Professor at the Graduate School of Advanced Science and Engineering, Hiroshima University. She received a bachelor's degree from Zhejiang University, a master's degree from the Graduate School of Business, Rikkyo University, and a PhD degree from the Graduate School of Informatics, Kyoto University. Her research interests include: Information Retrieval, Knowledge Discovery, Educational Data Mining and Learning Analytics.

Brendan Flanagan is an Associate Professor at the Center for Innovative Research and Education in Data Science, Institute for Liberal Arts and Sciences, and the Graduate School of Informatics at Kyoto University. He received a bachelor's degree from RMIT University and master's and Ph.D. degrees from the Graduate School of Information Science and Electrical Engineering, Kyushu University. His research interests include learning analytics, educational data science, educational data mining, NLP/text mining, machine learning, computer assisted language learning, and the application of blockchain in education.

Hiroaki Ogata is a full Professor at the Academic Center for Computing and Media Studies, Kyoto University. His research interests include learning analytics, evidence-based education, educational data mining, educational data science, computer supported ubiquitous and mobile learning, and CSCL.

Funding

This work was partly supported by JST BOOST JPMJBS2407, JSPS JP25KJ1627, JP24K20902, JP20H01722, JP23K25698 and JP23H01001, JP21K19824, JP23H00505, NEDO JPNP20006, and CSTI SIP Program JPJ012347.

Availability of data and materials

Not applicable.

Declarations

Competing interests

H. Ulrich Hoppe and Hiroaki Ogata are members of the Editorial Board of RPTEL. Other than that, the authors declare that there are no competing interests.

Author details

[1] Graduate School of Informatics, Kyoto University, Japan. [2] RIAS Institute, Bürgerstr. 15, 47057 Duisburg, Germany. [3] Graduate School of Advanced Science and Engineering, Hiroshima University, Japan. [4] Center for Innovative Research and Education in Data Science, Institute for Liberal Arts and Sciences, Kyoto University, Japan. [5] Academic Center for Computing and Media Studies, Kyoto University, Japan.

Received: 31 January 2025 Accepted: 17 December 2025
Published online: 1 January 2027 (Online First: 15 April 2026)

References

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. In *Intelligent tutoring systems: Proceedings of the 7th international conference.* (pp. 531–540).

Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. In *Artificial Intelligence in Education: 16th International Conference, 16*, 431–440.

Beck, J. E., & Rodrigo, M. M. T. (2014). Understanding wheel spinning in the context of affective factors. In *Intelligent Tutoring Systems: 12th International Conference, 12*, 162–167.

Berg, C. A., & Phillips, D. G. (1994). An investigation of the relationship between logical thinking structures and the ability to construct and interpret line graphs. *Journal of research in science teaching, 31*(4), 323–344. https://doi.org/10.1002/tea.3660310404

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31*, 6, 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Bull, S., & Kay, J. (2010). Open learner models. In *Advances in Intelligent Tutoring Systems*. 301–322.

Caliskan, O. (2021). Learning from failure or failing to learn: Perspectives of school principals. *Journal of Educational Administration, 59*(4), 422–436. https://doi.org/10.1108/JEA-07-2020-0158

Carlson, R. W., & Fishbach, A. (2024). Learning from failure. *Motivation Science, 10*(3), 160–170. https://doi.org/10.1037/mot0000338

Chounta, I. A., Albacete, P., Jordan, P., Katz, S., & McLaren, B. M. (2017). The "Grey Area": A computational approach to model the zone of proximal development. In *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, 12* (pp. 3–16).

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction, 4*, 253–278. https://doi.org/10.1007/BF01099821

Darabi, A., Arrington, T. L., & Sayilir, E. (2018). Learning from failure: A meta-analysis of the empirical studies. *Educational Technology Research and Development, 66*, 1101–1118. https://doi.org/10.1007/s11423-018-9579-9

Dillenbourg, P., Prieto, L. P., & Olsen, J. K. (2018). *Classroom orchestration*. In *International Handbook of the Learning Sciences* (pp. 180-190). Routledge.

Edisherashvili, N., Saks, K., Pedaste, M., and Leijen, Ä. (2022). Supporting self-regulated learning in distance learning contexts at higher education level: Systematic literature review. *Frontiers in Psychology, 12*, 792422. https://doi.org/10.3389/fpsyg.2021.792422

Eskreis-Winkler, L., & Fishbach, A. (2019). Not learning from failure—The greatest failure of all. *Psychological Science, 30*(12), 1733–1744. https://doi.org/10.1177/0956797619881133

Eskreis-Winkler, L., & Fishbach, A. (2022). You think failure is hard? So is learning from it. *Perspectives on Psychological Science, 17*(6), 1511–1524. https://doi.org/10.1177/17456916211059817

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: An International Journal, 10*, 4, 469–484. https://doi.org/10.34105/j.kmel.2018.10.029

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 231. https://doi.org/10.1037/0278-7393.22.1.231

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education, 47*(2), 183–210. https://doi.org/10.1080/03057267.2011.605307

Hatala, M., Nazeri, S., & Kia, F. S. (2023). Progression of students' SRL processes in subsequent programming problem-solving tasks and its association with tasks outcomes. *The Internet and Higher Education, 56*, 100881. https://doi.org/10.1016/j.iheduc.2022.100881

Jackson, A., Godwin, A., Bartholomew, S., & Mentzer, N. (2022). Learning from failure: A systematized review. *International Journal of Technology and Design Education, 32*(3), 1853–1873. https://doi.org/10.1007/s10798-021-09661-x

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Kapur, M. (2014). Comparing learning from productive failure and vicarious failure. *Journal of the Learning Sciences, 23*(4), 651–677. https://doi.org/10.1080/10508406.2013.819000

Kai, S., Almeda, M. V., Baker, R. S., Heffernan, C., and 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

Klerkx, J., Verbert, K., & Duval, E. (2014). *Enhancing learning with visualization techniques*. Handbook of Research on Educational Communications and Technology. 791–807.

Molenaar, I., de Mooij, S., Azevedo, R., Bannert, M., Järvelä, S., and Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior, 139*, 107540. https://doi.org/10.1016/j.chb.2022.107540

Oudman, S., van de Pol, J., & van Gog, T. (2022). Effects of self-scoring their math problem solutions on primary school students' monitoring and regulation. *Metacognition and Learning*. 1–27. https://doi.org/10.1007/s11409-021-09281-9

Owen, V. E., Roy, M., & Thai, K.P. (2019). Detecting wheel-spinning and productive persistence in educational games. *International educational data mining society*.

Palaoag, T. D., Rodrigo, M. M. T., Andres, J. M. L., Andres, J. M. A. L., & Beck, J. E. (2016). Wheel-spinning in a game-based learning environment for physics. In *Intelligent Tutoring Systems: 13th International Conference, 13*, 234–239.

Park, S. (2023). Discovering unproductive learning patterns of wheel-spinning students in intelligent tutors using cluster analysis. *TechTrends*, 1–9. https://doi.org/10.1007/s11528-023-00847-9

Pelánek, R. (2018). Conceptual issues in mastery criteria: Differentiating uncertainty and degrees of knowledge. *Articial Intelligence in Education*, 450–461. https://doi.org/10.1007/978-3-319-93843-1_33

Pintrich, P. R. (2000). *The role of goal orientation in self-regulated learning*. In M. Boekaerts, P. R. Pintrich, and M. Zeidner (Eds.), Handbook of Self-regulation. 451–502. https://doi.org/10.1016/B978-012109890-2/50043-3

Rienties, B., Tempelaar, D., Nguyen, Q., & Littlejohn, A. (2019). Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Computers in Human Behavior, 100*, 345–357. https://doi.org/10.1016/j.chb.2019.07.007

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics, 2*(1), 7–12. https://doi.org/10.18608/jla.2015.21.2

Sacks, K., & Leijen Äli. (2014). Distinguishing self-directed and self-regulated learning and measuring them in the e-learning context. *Procedia-Social and Behavioral Sciences, 112*, 190–198. https://doi.org/10.1016/j.sbspro.2014.01.1155

Schuman, H., Walsh, E., Olson, C., & Etheridge, B. (1985). Effort and reward: The assumption that college grades are affected by quantity of study. *Social Forces, 63*(4), 945–966. https://doi.org/10.1093/sf/63.4.945

Stinebrickner, T. R., & Stinebrickner, R. (2008). The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy, 8*(1), Article 28. https://doi.org/10.2202/1935-1682.1868

Vygotsky, L. (1978). *Interaction between Learning and Development* (pp. 79–91). Linköpings universitet.

Wang, Y., Kai, S., & Baker, R. S. (2020). Early detection of wheel-spinning in ASSISTments. In *Artificial Intelligence in Education: 21st International Conference, 1*(21), 574–585.

Winne, P. H. (2010). Improving measurement of self-regulated learning. *Educational Psychologist, 45*(4), 267–276. https://doi.org/10.1080/00461520.2010.517150

Yamauchi, T., Hoppe, U.H., Dai, Y., Flanagan, B., & Ogata, H. (2024). Representing learning progression of unguided exercise solving: A generalization of wheel-spinning detection. In *32nd International Conference on Computers in Education, 1*, 686–695.

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*.

Zheng, J., Huang, L., Li, S., Lajoie, S. P., Chen, Y., & Hmelo-Silver, C. E. (2021). Self-regulation and emotion matter: A case study of instructor interactions with a learning analytics dashboard. *Computers & Education, 161*, 104061. https://doi.org/10.1016/j.compedu.2020.104061

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of Self-regulation*. 13–39.

Publisher's Note