Research and Practice in
Technology Enhanced Learning

RESEARCH Free and Open Access

# Integrating self-explanation and operational data for impasse detection in mathematical learning

Ryosuke Nakamoto [1*], Brendan Flanagan [2], Yiling Dai [3], Taisei Yamauchi [1], Kyosuke Takami [4] and Hiroaki Ogata [3]

*Correspondence:
s0527225@gmail.com
Graduate School of Social Informatics,
Kyoto University,
Kyoto, Japan.
Full list of author information is available at the end of the article

**Abstract**

Self-explanation is increasingly recognized as a key factor in learning. Identifying learning impasses, which are significant educational challenges, is also crucial as they can lead to deeper learning experiences. This paper argues that integrating self-explanation with relevant datasets is essential for detecting learning impasses in online mathematics education. To test this idea, we created an evaluative framework using a rubric-based approach tailored for mathematical problem-solving. Our analysis combines various data types, including handwritten responses and digital self-explanations from 93 middle school students. Using hierarchical logistic regression, we examined feature groups such as Self-Explanation Quality, Handwriting Features, and Overall Level of Action. Models based solely on self-explanation achieved a 74.0% accuracy rate, while adding more features increased the final model's accuracy to 80.06%. This improvement highlights the effectiveness of an integrated approach. The combined model, which merges generated handwriting features counts with self-explanation features, shows the importance of both qualitative and quantitative measures in identifying learning impasses. Our findings suggest that a comprehensive approach, leveraging detailed operational data and rich self-explanation content, can enhance the detection of learning challenges, providing insights for personalized education in online learning environments.

**Keywords:** Self-explanation, Impasse detection, Online mathematics education, Educational data analysis

## Introduction

The integration of digital learning platforms has revolutionized educational practices, particularly within the domain of mathematics where self-explanation is recognized as a critical learning activity (Chi et al., 1994; Rittle-Johnson et al., 2017). These platforms offer interfaces that encourage students to engage in deep self-explanation, a practice that has been shown to enhance understanding and problem-solving capabilities (Conati & VanLehn, 2000). Structured self-explanation templates aid in articulating thought processes, extending beyond traditional educational tools to encompass feedback mechanisms and automated analysis systems (Berthold & Renkl, 2009; Nakamoto et al., 2024). Such self-explanatory data is invaluable for both educators and automated systems to gain insights into cognitive processes, shaping personalized educational strategies (Crippen & Earl, 2007). With the surge of online learning platforms, the need for effective, autonomous learning strategies is more pronounced than ever. In this context, self-explanation emerges not only as a facilitator of individual learning processes but also as a critical element in monitoring and enhancing the quality of learning.

In parallel, educational technology has concentrated efforts on identifying and addressing cognitive barriers or “impasses” that impede mathematical comprehension (Brown & VanLehn, 1980; Carroll & Kay, 1988; Siegler & Jenkins, 1989; VanLehn et al., 1992). Modern pedagogical approaches aim to foster not just content delivery but also deep cognitive engagement and sustained curiosity (Hattie et al., 2009). Recognizing impasses is crucial, as their presence signifies significant learning hurdles that can prompt introspective and explorative learning, enhancing the educational experience. Addressing these impasses directly has been proven to significantly elevate learning outcomes.

Detecting learning impasses, especially in unstructured educational tasks, poses considerable challenges, often requiring significant time and resources. This complexity presents a notable burden for educators and limits the widespread implementation of impasse detection strategies. To address these challenges, our study proposes a novel approach: the creation of an automated impasse detection system that leverages the advantages of self-explanations. We argue that when learners engage in the process of self-explanation, they generate rich conceptual data. This data, when analyzed in conjunction with relevant datasets, can be instrumental in detecting learning impasses—points where learners struggle or fail to grasp key mathematical concepts. Identifying these impasses is crucial in online learning scenarios where personalized instructor feedback is often limited. To explore this hypothesis, our study is guided by the following research questions:

**RQ1: How does self-explanation data influence the detection of learning impasses in our proposed evaluative framework?**

This question aims to understand and analyze the role of self-explanation in identifying barriers to understanding mathematical concepts and assesses the accuracy of the prediction model within our evaluative framework.

**RQ2: Which factors are significant predictors of learning difficulties in online mathematics education?**

This explores which specific aspects of student engagement and self-explanation are most predictive of encountering learning challenges.

To address these questions, we developed an impasse evaluative framework utilizing a rubric-based approach tailored for mathematics education. This framework emphasizes extracting and analyzing features from self-explanations provided by learners, which serve as a rich source of insights into their conceptual understanding and thought processes. Our goal is to identify key patterns and markers indicative of learning difficulties, thereby contributing to enhancing the effectiveness of online mathematics education.

## Related work

### Effectiveness of self-explanations in addressing impasses

Self-explanation holds great promise for improving mathematics learning, with consistent support from previous research, particularly when paired with contrasting instructional examples (Renkl, 2017; Rittle-Johnson, 2017; Rittle-Johnson et al., 2017). In the context of mathematics, it plays a crucial role in advancing both conceptual and procedural knowledge. Conceptual knowledge relates to understanding abstract principles such as mathematical equivalence, while procedural knowledge is cultivated through problem-solving practices specific to particular contexts (Rittle-Johnson et al., 2001; Star, 2005).

Efforts have been made to integrate self-explanation into web-based learning environments, including the development of structured problem-solving tools like those created by Crippen and Earl (2007). McNamara et al. (2004) introduced iSTART, an interactive tutoring system that uses Natural Language Processing (NLP) techniques to assess and support self-explanations in reading comprehension. Similar systems have proven effective across diverse content domains (Jackson et al., 2010). Self-explanation contributes to learning through two key processes: firstly, it aids comprehension by streamlining knowledge integration (Chi, 2000), and secondly, it enhances the recognition and transfer of knowledge by directing learners' attention to the underlying structural aspects of content rather than surface-level features (McEldoon et al., 2013; Rittle-Johnson, 2006).

Self-explanation enables learners to infer causal relationships and conceptual connections, thereby deepening their understanding. Often, learners struggle to grasp the interconnections between learning units, leading to an inability to apply their knowledge to different tasks. However, through the process of self-explanation, learners establish connections between various learning elements, enhancing their comprehension.

Additionally, the act of self-explanation has been shown to have effects such as revealing areas of non-understanding for the learner (Siegler & Jenkins, 1989). In other words, we thought that utilizing self-explanation can aid in detecting these impasses. Impasses are obstacles that obstruct the comprehension of mathematical concepts. Proficient learners adeptly employ self-explanation during mathematical problem-solving, infusing each step with reflective self-explanation, which helps identify essential knowledge components. Conversely, less proficient learners may underestimate or ineffectively use self-explanation, resulting in a weaker grasp of mathematical concepts.

Recent advancements in digital technology have enabled the integration of self-explanation strategies into web-based learning platforms (Flanagan & Ogata, 2018; Nakamoto et al., 2021, 2022, 2024; Ogata et al., 2023). This shift towards digital education has prompted our exploration of the potential of self-explanations as a means to detect students' difficulties, or "impasse," in mathematics.

## Learning impasse

Learning impasses represent pivotal moments in a student's educational journey, particularly in the field of mathematics (VanLehn et al., 2003). Warli et al. (2020) emphasize the cognitive challenges students face in understanding complex mathematical concepts, particularly in group problem-solving contexts. These impasses often manifest as feelings of being stuck, error detection, or doubt, even when students are technically correct in their actions. In response to these difficulties, students might review previous knowledge, consult resources, or seek help from peers or instructors. These moments of struggle are pivotal in fostering genuine learning, as they lead to a deeper grasp of the subject, correction of misconceptions, and enhancement of problem-solving skills. These understanding aligns with the essence of our study on impasse detection and resolution in mathematics education. Extensive research on cognitive skill acquisition consistently underscores the prevalence of impasses in the learning process, particularly in the absence of tutors (VanLehn, 1987, 1990; VanLehn & Jones, 1993). Therefore, the automated detection of these mathematical impasse holds significant potential for advancing educational research.

This study delves deeper into the concept of learning impasses, particularly in the context of math self-explanation learning, and seeks to detect these impasses in automated learning environments. Through this study, we intend to contribute to the field by not only defining and identifying impasses in automated math learning environments but also by evaluating the role of self-explanations in overcoming these challenges.

## Learning context and premise conditions

### Math learning activities and collected logs in LEAF system

As a system platform, we used the LEAF platform (Flanagan & Ogata, 2018), which consists of a digital reading system named BookRoll and a learning analytics tool LAViEW where students and teachers can monitor and reflect on their learning. BookRoll captures the handwriting data as a series of vectors representing the coordinates and velocity of pen strokes, allowing realistic playback of the handwritten answers and fine-grained analysis of the students' answering process.

Students follow a structured sequence in their interaction with the system, as outlined in Table 1. They begin by reviewing math problems tailored to their educational level. Using stylus pens on a digital interface, students solve these problems, capturing their problem-solving approach as time series data. This data proves valuable for understanding stroke order, pauses, and handwriting patterns, aiding in anomaly detection. After solving the quiz, students utilize the LAViEW system to review their handwritten answers and explain their solution processes. Self-explanations are inputted chronologically, aligning them with the corresponding pen stroke data, ensuring a temporal association between self-explanations and the writing process. During these steps, the system meticulously logs various data points. This includes handwriting data, recorded as vectors to capture intricate details like stroke speed, pressure, and duration. This granularity in data collection provides insights into the cognitive processes of students during problem-solving and self-explanation phases.

Figure 1, cited from Nakamoto et al. (2024), describes handwritten answer playback and self-explanation input. The students input a sentence of explanation every time they think they have completed some step in their answers during the playback. Therefore, the self-explanation is temporally associated with the pen stroke data. The self-explanation of the

**Table 1** Student journey in the LEAF system

| Order | Activity | Description |
|---|---|---|
| 1 | View the Math Problems | Students start by reviewing math problems tailored to their educational level. |
| 2 | Solve the Math Problems Using Stylus Pens | Students use stylus pens on the digital interface to solve the problems. |
| 3 | Playback One's Answer | The system allows for the playback of recorded answers, enabling students to reflect on and evaluate their problem-solving process. |
| 4 | Input Self-Explanations | Students articulate their thought processes and reasoning by inputting self-explanations for their solutions. |
| 5 | Check the Standard Solution | Post-solution, students have access to standard solutions for comparison and analysis. |

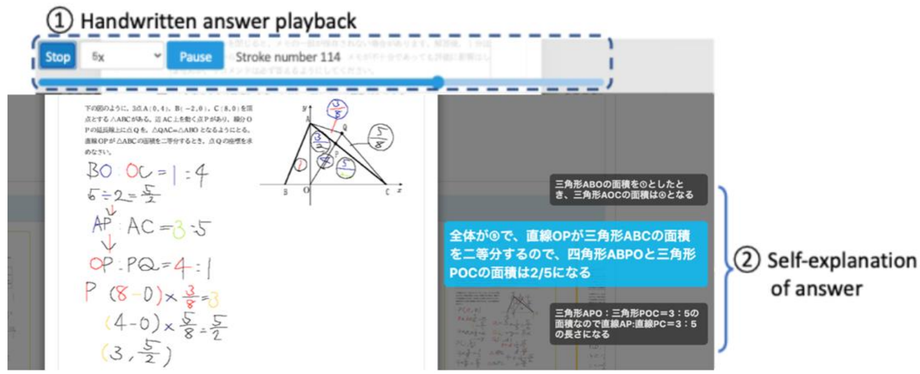


**Fig. 1** Handwritten answer review playback and self-explanation input user interface

answer contains the following from top to bottom: "If the area of triangle ABO is 1, the area of triangle AOC is 4. Since the whole is five and straight-line OP bisects the area of triangle ABC, the area of quadrilateral ABPO and triangle POC is 2/5. The area of triangle APO: triangle POC = 3:5, so the length of straight line AP: straight-line PC = 3:5."

## Premise conditions and the definition of impasse in math learning

Our approach to detecting impasses aligns with the definition proposed by VanLehn et al. (2003), where an impasse is seen as a point of cognitive struggle or uncertainty encountered by a student while trying to understand or apply a specific piece of knowledge in mathematics. Such impasses can manifest in various forms, including being stuck, detecting an error, or experiencing doubt or uncertainty even when the action is correctly performed. They prompt students to delve into their memory, consult external resources such as textbooks, or seek assistance from others. In line with this understanding, our research posits that the ability to produce detailed self-explanations is a crucial indicator of a student's comprehension of mathematical concepts. We find that difficulties in providing coherent self-explanations often signal these underlying cognitive impasses. Therefore, we rigorously analyze the quality of self-explanations, understanding that clear and thorough explanations are indicative of a solid grasp of concepts, while vague or incomplete ones point to areas of weak comprehension.

We closely examine instances where self-explanations by students in mathematics learning reveal significant impasses. A high-quality self-explanation is not only logically structured but also elucidates the student's clear pathway through the problem-solving process. This type of explanation demonstrates a deep understanding of the mathematical concept, showcasing the student's ability to integrate and apply their knowledge effectively. On the other hand, an ineffective self-explanation, characterized by a lack of coherence and clarity, points to potential learning impasses. These impasses might manifest as fragmented thoughts, unclear connections between ideas, or an inability to express the underlying mathematical reasoning. Such explanations often signal a superficial understanding or a disconnect in cognitive processing, where the student might be able to perform certain steps of the problem-solving process but fails to comprehend the overarching concepts or principles.

## Proposed method

### Rubric-based evaluative framework

Our study, in addressing the concept of learning impasses, seeks to develop a rubric-based evaluative system, aimed at providing an objective and precise assessment of student reasoning in mathematics. Rubrics, as defined by Andrade (2000) and Arter and Chappuis (2007), are vital assessment tools designed to classify the quality of student reasoning, performance, or outputs across a range from outstanding to deficient. These rubrics serve as guidelines to assess students' work, with general rubrics offering a broad overview of performance levels, and task-specific rubrics detailing the mathematical elements crucial for each performance level. Particularly suitable for tasks that allow multiple solutions or strategies (Thompson & Senk, 1998), our study employs simplified rubrics to evaluate the step-by-step problem-solving ability of students, with a focus on facilitating systematic judgment by the system.

Verschaffel et al. (1999) defined five essential steps in mathematical problem-solving: drawing representations, listing elements, simplifying figures, executing calculations, and assessing solutions. Self-explanation, although not directly linked to a particular step, is crucial in revealing the knowledge components involved in these steps. We hypothesize that students navigate these steps, with self-explanation being a key aspect of their procedural knowledge. Tables 2 and 3 present the rubric definitions and sample answers as proposed by Nakamoto et al. (2024), which we have applied in our theoretical framework for detecting learning impasses. Table 2, in particular, defines key terms utilized in this paper and Table 3 illustrates sample answers that ideally represent the required knowledge components at each step of solving the quiz. Through this rubric-based approach, we aim to provide a clearer understanding of students' mathematical problem-solving processes and to identify where and why they encounter learning impasses. This framework not only aids in assessing students' current understanding but also helps in pinpointing areas that require further instructional focus.

**Table 2** Description of words

| Name | Definitions |
|---|---|
| Rubric | Can-do descriptors that clearly describe all the essential elements of the quiz and are used to create labels and sample self-explanations for scoring. Ordinal Scale (1-4). |
| Labels | Labels consist of true or false for each of the rubrics 5 steps, subsequently referred to as "correct step" or "incorrect step" answers. In particular, "Incorrect Step" signifies the point at which the student got stuck. |
| Sample Sentences of Self-explanations | Model answers of self-explanations prepared according to the 4-step rubric number |

**Table 3** Rubrics and sample sentences of self-explanation

| Number | Rubric | Sample Sentences of Self-explanations |
|---|---|---|
| Step 1 | Be able to find the equation of a linear function from two points. | Substituting the y-coordinate of p into the equation of the line AC. |
| Step 2 | Be able to find the equation of the line that bisects the area of a triangle. | Find the area of triangle ABC, and then find the area of triangle OPC. |
| Step 3 | Be able to represent a point on a straight line using letters (P-coordinates). Be able to represent a point on a straight line using letters (Q-coordinate). | With the line OC as the base, find the y-coordinate of p, which is the height. p's coordinate is (t,-1/2t+4).<br>Since the coordinates of P are (3,5/2), the line OP is y=5/6 and the coordinates of Q are placed as (t,5/6). |
| Step 4 | Be able to formulate an equation for area based on relationships among figures. | Finally, the area of △QAC was found from the areas of △AQO and △OQC, and the coordinates of Q were found. |

## Labeling impasses by human assessment

To assess students' comprehension of rubric components, we proceed with the creation of labels of problem-solving step impasses through meticulous human assessment. This assessment process involves the manual review of students' handwriting and self-explanations, culminating in the assignment of labels that signify students' grasp of the rubric components.

Our approach extends beyond the conventional binary evaluation of responses as merely correct or incorrect. It resonates with the perspective of Carroll and Kay (1988) by prioritizing the cognitive processes underpinning student responses. We aim to probe into the cognitive depths of students who may rely on rote memorization or have a cursory understanding of mathematical concepts but struggle to articulate their thought processes. This nuanced approach transcends simple correctness assessment, striving to uncover the underlying reasons for student responses. It aligns with modern pedagogical research, as advocated by Hattie et al. (2009), which stresses the significance of a profound comprehension of student learning mechanisms. This approach crucially differentiates between the presence and absence of comprehension barriers at each stage of the learning process. This representation serves dual purposes:

**Correct Step:** We showcase a scenario where the learner demonstrates correct understanding. This is indicated by coherent self-explanations and consistent handwriting patterns, typically leading to a binary value of ‘1’ in our assessment, signifying no detected impasse.

**Incorrect Step:** Errors in the handwritten solution and structural inadequacies in the self-explanations, especially in tackling a geometric concept, point to a ‘0’ in our binary system, highlighting the specific area of the learner’s difficulty.

With the acquired datasets in hand, our next focus was the creation of gold-standard self-explanations for our model. This process involved a systematic deconstruction of each quiz into distinct steps. The self-explanations thus developed served as benchmarks against which we evaluated students’ responses. These model self-explanations played a pivotal role in our study, not only in terms of their comprehensiveness but also in representing the expected knowledge.

Figure 2 displays an image of the handwritten answer along with its associated self-explanations. Evaluators reviewed these elements to create a ground truth, classifying each part as correct or incorrect. In Figure 2, the self-explanation is outlined as follows. The term “Stroke No.” refers to a number that is automatically generated by the system.

Stroke 34: Calculated the area of triangle ABC.
Stroke 60: Calculated the area of triangle ABO.
Stroke 182: Determined the coordinates of point P and derived the equation of line OP.
Stroke 201: Expressed the coordinates of point Q in terms of variables.
Stroke 291: Determined the value of k.
Stroke 316: Substituted the value of k and obtained the answer.

## Feature extraction method

### Feature 1: Self-explanation quality indicators

#### *Settings*

The primary objective of our problem setting is to establish a comprehensive framework that effectively identifies when and where students encounter cognitive barriers during mathematical problem-solving. To achieve this, we compare students’ self-explanations

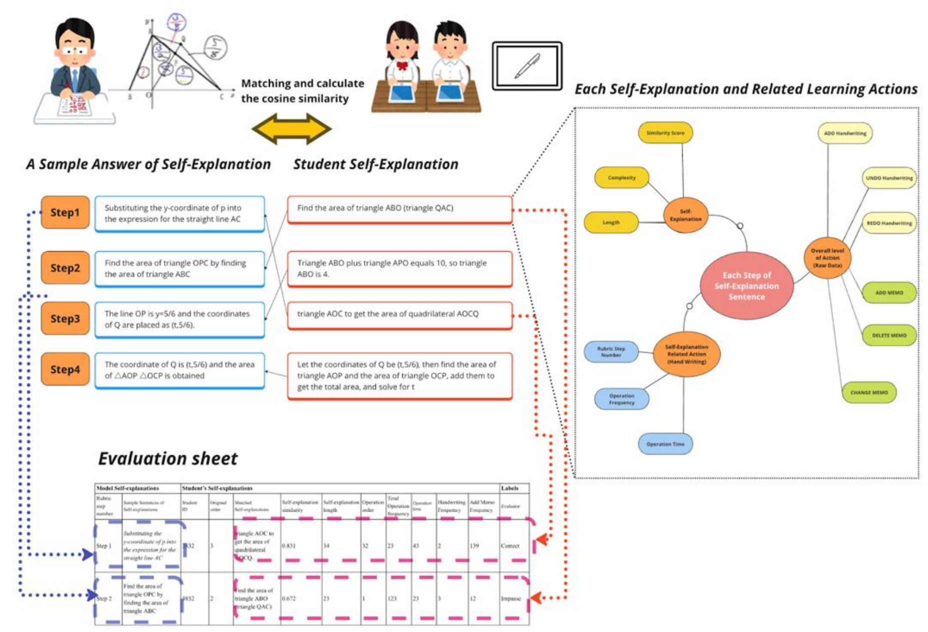


**Fig. 3** Overview of impasse detection analysis and self-explanation related features

with model answers, employing a similarity scoring mechanism. This mechanism quantitatively assesses the alignment of each student's self-explanation with the exemplar sentences provided for each step of the rubric, as illustrated in Figure 3. The data collected from this comparison are subjected to rigorous statistical analysis, which not only validates the efficacy of our impasse detection method but also aids in refining our model. This refinement process involves pinpointing the most predictive features and enhancing the feedback mechanism for better educational outcomes.

Our decision to standardize the problem-solving process into four distinct steps is grounded in an extensive analysis of junior high school mathematics curricula, incorporating the rubric steps proposed by Nakamoto et al. (2021, 2024). In junior high school mathematics, problem-solving typically involves multiple steps, each representing a unique unit of knowledge or skill. These steps can vary in number, from two to six or more, depending on the complexity of the mathematical concept. For our study, we have deliberately chosen to limit the problem-solving steps to four. This decision is based on the standard level of junior high school mathematics curricula and the common problem-solving approaches observed at this educational stage. For each problem P, we assume the presence of four essential knowledge elements, denoted as K1, K2, K3, and K4, with A(P) representing the self-explanatory text created by a student. This relationship is formulated as follows:

$$A(P) = [A(K1), A(K2), A(K3), A(K4)] \tag{1.1}$$

### *Method*

In our methodology, both the sample sentences and the students' self-explanations are transformed into vectors using BERT (Bidirectional Encoder Representations from Transformers). We then compute cosine similarity to align the rubric steps with the students' explanations. The weighted average of these similarity scores is used to determine a rubric-based score for each student, resulting in five distinct scores corresponding to each rubric step. These scores are instrumental for further analysis and model development. Our approach incorporates a range of language models for text representation. These include traditional models like TF-IDF (Salton & Buckley, 1988) and advanced transformer-based models such as BERT (Devlin et al., 2019) and Sentence BERT (SBERT; Reimers & Gurevych, 2019). BERT, built upon the transformer architecture (Vaswani et al., 2017), is renowned for its effectiveness in natural language processing tasks and has been widely applied in educational technology (Yang et al., 2021). SBERT, an adaptation of BERT, excels in generating sentence embeddings and is particularly suited for evaluating the diverse self-explanations of students.

When a student creates a self-explanation S(P), we calculate the similarity for each knowledge element using a function Sim and perform block division using a function Split. The association of each knowledge element (K1 to K4) in the student's self-explanation block L with A(P) is determined by similarity. For instance, combinations such as (K1, L4), (K2, L2), and so forth, are considered. The self-explanation in block L is then linked to handwritten log information, as shown below:

$$\mathrm{Sim}(P) = [\mathrm{Sim}(S(K1), A(K1)), \mathrm{Sim}(S(K2), A(K2)), \mathrm{Sim}(S(K3), A(K3)), \mathrm{Sim}(S(K4), A(K4))] \tag{1.2}$$

## Feature 2: Handwriting data associated with each self-explanation

In our research, we employ a comprehensive approach that combines handwritten data with self-explanation texts to profoundly understand students' learning processes, building upon prior studies in handwriting detection (Flanagan et al., 2022; Iiyama et al., 2017; Kishi & Miura, 2018; Ochoa et al., 2013; Nakamoto et al., 2021). Figure 3 describes the overview of the Image of feature extraction associated with one Self-explanation of a student. This method involves analyzing handwritten data, including pen strokes and inputs, to reveal the temporal and procedural aspects of problem-solving, crucial for identifying impasses by observing the sequence and methodology students employ in tackling mathematical problems. Integrating these insights with self-explanation data allows us to form a holistic view of how students learn and pinpoint areas of difficulty. We adopt a methodology inspired by Kishi and Miura (2018), preparing model solutions for each quiz that encapsulate essential steps and knowledge components for correct answers. We then compare students' self-explanations to these models and identify the most relevant self-explanation sentences. Lastly, we generated the handwriting data associated with a most relevant self-explanation sentence as shown in Figure 3 and Table 4. The detail of feature extraction method for the key factor is as follows.

$$\text{Total Activity Duration} = \sum_{i=1}^{n-1}(t_{i+1} - t_i) \tag{2.1}$$

where $t_i$ is the timestamp of the $i$-th operation and $n$ is the total number of operations.

$$\text{Operation Count} = n \tag{2.2}$$

where $n$ is the total number of operations.

$$\text{Explanation Interaction Level} = \text{Count}(\text{Strokes}) \tag{2.3}$$

where Strokes are the individual operations for a self-explanation.

$$\text{Explanation Detail Level} = \sum \text{CharCount}(\text{Self-Explanation}) \tag{2.4}$$

where CharCount is the function that counts the number of characters.

$$\text{Explanation Complexity} = \text{Count}(\text{Nodes}) \tag{2.5}$$

where Nodes are the number of morphologically analyzed words using MeCab.

$$\text{Explanation Sequence} = (s_1, s_2, \ldots, s_n) \tag{2.6}$$

where $s_i$ is the $i$-th stroke in the sequence.

$$\text{Text Similarity Score} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \tag{2.7}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the vector representations of the self-explanation and reference sentence, respectively.

$$\text{Operation Segment Duration} = \sum_{i=1}^{k}(t_{\text{end},i} - t_{\text{start},i}) \tag{2.8}$$

where $t_{\text{start},i}$ and $t_{\text{end},i}$ are the start and end times of the $i$-th operation segment and $k$ is the total number of segments.

$$\text{Accurate Activity Duration} = \sum \text{new\_time} \tag{2.9}$$

where new_time is the recalculated time intervals using an improved method.

$$\text{Rubric Step Number} = \text{Ordinal}(\text{Scale}) \tag{2.10}$$

where Scale typically ranges from 1 to 4, representing the rubric steps.

**Table 4** Description of features for analysis

| Feature Group | Feature Name | Description | Mean | SD |
|---|---|---|---|---|
| Self-Explanation Quality Indicators | Self-Explanation Engagement Level | The count of strokes or actions taken associated with each self-explanation, indicating the degree of user engagement. | 0.06 | 0.19 |
| | Self-Explanation Length | The character count of each self-explanation, serving as an indicator of the explanation's level of detail. | 0.08 | 0.22 |
| | Self-Explanation Complexity | The number of nodes (words) in each self-explanation, determined through morphological analysis using MeCab, indicating the complexity of the explanation. | 0.09 | 0.22 |
| | Self-Explanation Sequence | The order of strokes (operations) in each self-explanation, important for understanding user operation patterns. | 0.25 | 0.24 |
| | Self-Explanation Similarity Score | A score based on cosine similarity between self-explanations and reference sentences, indicating the similarity of the text. | 0.08 | 0.22 |
| Self-Explanation Related Action (Generated Handwriting Features) | Total Activity Duration | The total sum of time intervals between consecutive operations for each user, reflecting the overall duration of user activity(Include time intervals). | 0.10 | 0.23 |
| | Generated Handwriting Features Count | The total count of operations performed by each user, indicating the level of user activity. | 0.30 | 0.25 |
| | Total Activity Duration | The total sum of time intervals between consecutive operations for each user, reflecting the overall duration of user activity. | 0.22 | 0.21 |
| | Operation Segment Duration | A segment of time data associated with specific user operations, segmented based on user interactions, indicating the temporal aspects of user interactions. | 0.23 | 0.23 |
| | Accurate Activity Duration | The total of calculated times, reflecting a more accurate measure of user activity duration(Exclude time intervals. only the time for operations). | 0.30 | 0.22 |
| | Rubric Step Number | An ordinal scale (typically ranging from 1 to 4) that quantifies the number of rubric steps. | 2.50 | 1.12 |
| Overall Level of Action (Raw Operational Data) | ADD_HW_MEMO | Addition of hardware-related memo. | 0.04 | 0.17 |
| | UNDO_HW_MEMO | Undo hardware-related memo. | 0.09 | 0.22 |
| | CLOSE | Close an item or section. | 0.21 | 0.38 |
| | OPEN_RECOMMENDATION | Open a recommended item. | 0.05 | 0.18 |
| | TIMER_PAUSE | Pause a timer. | 0.01 | 0.10 |
| | CLOSE_RECOMMENDATION | Close a recommended item. | 0.03 | 0.17 |
| | ADD MEMO | Addition of a general memo. | 0.02 | 0.13 |
| | QUIZ_ANSWER_CORRECT | Provide a correct answer in a quiz. | 0.04 | 0.19 |
| | PAGE_JUMP | Jump to a specific page. | 0.08 | 0.24 |
| | QUIZ_ANSWER | Provide an answer in a quiz. | 0.01 | 0.09 |
| | CHANGE MEMO | Modify an existing memo. | 0.00 | 0.00 |
| | DELETE_MEMO | Delete a memo. | 0.04 | 0.17 |
| | CLEAR_HW_MEMO | Clear hardware-related memos. | 0.04 | 0.17 |
| | REDO_HW_MEMO | Redo a hardware-related memo. | 0.09 | 0.22 |
| | ADD_RECOMMENDATION | Add a recommendation. | 0.21 | 0.38 |

## Experimental settings

### Datasets overview

During the data collection period from January 1, 2020, to December 31, 2022, our study amassed a significant dataset comprising 900 answers. The dataset was derived from the engagement of 93 unique users who interacted with a set of 23 different quizzes. Initially, all students were in the 1st grade of junior high school at the onset of the study. From April 1, 2021, the students transitioned into the 2nd grade, and from April 1, 2022, the students transitioned into the 3rd grade. While there are no explicit criteria indicating the students' precise level of academic ability, as a reference, the average score on the year-end mathematics exam was 74.92 out of 100, with a standard deviation of 11.74.

The quizzes covered a range of topics, each coded with a unique identifier reflecting both the topic and the specific focus within the curriculum. The topics included:

**Linear Functions:** Analyzing relationships and constructing graphs based on linear equations.

**Quadratic Functions:** Exploring the properties and applications of quadratic equations in various contexts.

**Geometry and Similarity:** Investigating properties of geometric shapes and their similarity through rigorous problem-solving.

**Quadratic Equations:** Delving into the solutions and real-world applications of quadratic equations.

**Pythagorean Theorem:** Applying the theorem to solve problems involving right triangles.

**Properties of Circles:** Understanding the geometric properties of circles and their implications in different scenarios.

**Symmetric Expressions:** Exploring the use of symmetric expressions in algebraic contexts.

**Square Roots:** Examining the concepts and calculations involving square roots.

**Utilization of Simultaneous Equations:** Solving complex problems using systems of equations, with a focus on applications that involve algebraic solutions and real-world scenarios.

The self-explanation answers were meticulously documented and categorized based on two key criteria: the presence of step numbers (1 to 4) in the rubric and the identification of a learning impasse ("Impasse" or "Not Impasse"). There were 581 responses at Step 1, 488 at Step 2, 382 at Step 3, and a further reduction to 314 at Step 4. Conversely, responses labeled as 'Impasse' exhibit an ascending pattern with the advancement of rubric steps, starting with 319 responses at Step 1, 412 at Step 2, 518 at Step 3, and eventually reaching 586 at Step 4. This data suggests that as the complexity of tasks increases with each rubric step, the incidence of impasses also rises. This indicates a potential accumulation of learning challenges that students encounter as they progress through the steps. The ground-truth impasse labels of each step were created by one academic assistant and the results were reviewed by one of the authors.

### Hierarchical logistic regression model

Table 5 describes the hierarchical Analysis Features Group. The use of hierarchical logistic regression analysis in our study is a strategic choice aimed at dissecting the impact of various predictors on the likelihood of learning impasses in an online educational setting. This method allows us to sequentially introduce variables related to self-explanation activities, engagement metrics, and interaction patterns to the model and assess their individual and collective contribution to the variance in the dependent variable, which in this case could be the occurrence of an impasse. By accounting for other variables, this technique clarifies the unique predictive value of each factor and its effectiveness in identifying student learning behaviors.

Our approach is methodically aligned with the dichotomous analytical framework of "overall level of analysis" and "specific task analysis," as established by Jovanović et al. (2021), and draws from the methodologies of Allen et al. (2015). This dual perspective ensures a comprehensive examination of student behavior, capturing both the broad learning environment and the intricate details of specific tasks.

## Results

### Correlation analysis

The purpose of the correlation analysis was to empirically test the relationships between self-explanation features and the correctness of answers (Ground Truth), directly relating to our first research question about the influence of self-explanation on detecting learning

**Table 5** The hierarchical Analysis Features Group

| Feature Group / Model Name | $M_S$ | $M_G$ | $M_A$ | $M_{S+G}$ | $M_F$ (Final) |
|---|---|---|---|---|---|
| Self-Explanation Quality Indicators | ○ | | | ○ | ○ |
| Self-Explanation Related Action (Generated Handwriting Features) | | ○ | | ○ | ○ |
| Overall Level of Action (Raw Operational Data) | | | ○ | | ○ |
| Total Features | 5 | 6 | 15 | 11 | 26 |

impasses. By employing Spearman correlation analysis, we observed how closely student self-explanations mirror reference materials and whether the detail in their explanations correlates with their understanding of mathematical concepts.

### *RQ1: How does self-explanation data influence the detection of learning impasses in our proposed evaluative framework?*

**Self-Explanation Similarity Score:** A positive correlation of 0.2527 between "Ground_Truth" and "Self-Explanation Similarity Score" suggests that greater similarity to reference material often corresponds to better comprehension. While this correlation is statistically significant, it is relatively low, indicating that other factors may also play significant roles in comprehension.

**Self-Explanation Length:** The length of self-explanations showed a positive correlation of 0.2434 with "Ground_Truth", indicating that more detailed explanations tend to signify deeper understanding. Although the correlation is modest, it highlights the potential benefit of encouraging students to provide detailed explanations.

**Self-Explanation Engagement Level:** The correlation between "Ground_Truth" and "Self-Explanation Engagement Level" is 0.0909. According to the guidelines by Schober et al. (2018), this indicates low practical value. This suggests that engagement alone, as measured in this study, may not be a predictor of comprehension.

### *RQ2: Which factors are significant predictors of learning difficulties in online mathematics education?*

**Self-Explanation Sequence:** A significant negative correlation of -0.2522 between "Ground_Truth" and "Self-Explanation Sequence" points to certain structuring methods of explanations possibly being less effective or overly complex. This suggests that the way students organize their explanations might sometimes hinder their understanding.

**Rubric Step Number:** The increasing complexity of material, indicated by a negative correlation of -0.2007 with "Rubric_Step_Number," correlates with a decrease in correct answers. This indicates that as the material becomes more challenging and requires students to combine multiple pieces of knowledge, their performance declines. This suggests that later steps in the quizzes, which are more complex, present greater difficulties for students.


**Fig. 4** The correlation analysis

## Hierarchical logistic regression analysis

In this section, we analyze the efficacy of various predictors in identifying learning impasses in middle school students' mathematical comprehension through hierarchical logistic regression. The analysis is organized under three subheadings: (1) Analysis of Self-Explanation Features, (2) Exploring General Learning Operations and the Impact of Memo and Quiz Actions, and (3) Comprehensive Analysis in the Final Model. Each subheading represents a distinct model or combination of models to examine different aspects of student engagement and self-explanation features, offering a holistic understanding of the factors influencing learning impasses.

### *Analysis of self-explanation features ($M_S$, $M_{S+G}$, $M_F$)*

In our study, examining the efficacy of various predictors in identifying impasses in middle school students' mathematical comprehension, we employed distinct models that focused on self-explanation features and general learning activities. The analysis, segmented into different models—$M_S$, $M_{S+G}$, and $M_F$—revealed insightful findings on the predictors' influence on detecting comprehension difficulties, as indicated by their respective AUC values.

(1). **Model $M_S$ (AUC = 0.740):** This model focused on self-explanation features. Four self-explanation features demonstrated a statistically significant association with successful impasse detection:

- Self-Explanation Length: Odds Ratio: 15.69
- Self-Explanation Similarity Score: Odds Ratio: 8.75
- Self-Explanation Engagement Level: Odds Ratio: 1.63
- Self-Explanation Sequence: Odds Ratio: 0.12

These positive coefficients corroborate the hypothesis that a higher level of detail in self-explanations aligns with better comprehension and course success.

(2). **Combined Model $M_{S+G}$ (AUC = 0.777):** This model integrated generated handwriting features counts with self-explanation features, enhancing the predictive power of each feature group. The significant predictors included:

- Self-Explanation Length: Odds Ratio: 18.44
- Self-Explanation Similarity Score: Odds Ratio: 10.11
- Generated Handwriting Features Count: Odds Ratio: 2.30
- Rubric Step Number: Odds Ratio: 0.62

This integration suggests that both qualitative aspects of self-explanation and quantitative measures of operational engagement are pivotal in identifying learning impasses.

(3). **Model $M_F$ (AUC = 0.802):** This model achieved the highest AUC among the self-explanation focused models, underscoring the significance of combining various predictors:

- Self-Explanation Length: Odds Ratio: 24.76
- Self-Explanation Similarity Score: Odds Ratio: 9.15
- Generated Handwriting Features Count: Odds Ratio: 15.98

This model's high AUC value reinforced the importance of a multifaceted approach, combining both qualitative and quantitative elements of student engagement to predict learning impasses effectively.

### *Exploring general learning operations and the impact of memo and quiz actions ($M_G$, $M_A$)*

(1). The MG model (AUC = 0.655) highlighted the role of operational factors such as:

- Generated Handwriting Features Count: Odds Ratio: 3.43
- Total Activity Duration: Odds Ratio: 1.37

Although this model exhibited a lower AUC compared to self-explanation focused models, it still underlined the relevance of general engagement levels in learning activities. The model's findings align with the notion that higher operational engagement is positively associated with mathematical comprehension and course success.

(2). The $M_A$ model (AUC = 0.603) centered on memo-related actions and quiz answers, revealing a nuanced perspective on impasse detection:

- Quiz Answers Correct: Odds Ratio: 2.18
- Add Memo: Odds Ratio: 0.56
- Add HW Memo: Odds Ratio: 0.53

While quiz answers were significant predictors, memo actions had a lesser impact, indicating a more complex relationship with learning comprehension.

### *Comprehensive analysis in final model ($M_F$)*

The final model, $M_F$ (AUC = 0.802), gathered significant predictors from the previous models, providing a holistic view of the factors influencing impasse detection. This model's high AUC value reinforced the importance of a multifaceted approach, combining both qualitative and quantitative elements of student engagement to predict learning impasses effectively.

## Discussion

### Effectiveness of a multifaceted evaluative framework in identifying learning impasses

In our study, we used hierarchical logistic regression to better understand learning challenges in online secondary school mathematics education. Our results show that combining different types of data can effectively identify these challenges. Initially, our

**Table 6** Results for the effect models with the indicators. Only significant indicators are shown

| Model | | Predictors | Odds Ratio | Coeff. | Std Error | t Value | p-value |
|---|---|---|---|---|---|---|---|
| $M_S$ | AUC = 0.740 | **Self-Explanation Length** | **15.69** | 2.75 | 0.21 | 13.09 | <0.001 |
| | LogL = -2133.1 | **Self-Explanation Similarity Score** | **8.75** | 2.17 | 0.19 | 11.50 | <0.001 |
| | AIC = 4278.1 | **Self-Explanation Engagement Level** | **1.63** | 0.49 | 0.18 | 2.72 | 0.010 |
| | BIC = 4315.3 | Self-Explanation Sequence | 0.12 | -2.10 | 0.15 | -13.83 | <0.001 |
| $M_G$ | AUC = 0.655 | **Generated Handwriting Features Count** | **3.43** | 1.23 | 0.15 | 8.38 | <0.001 |
| | LogL = -2360.8 | Rubric_Step_Number | 0.65 | -0.42 | 0.03 | -13.46 | <0.001 |
| | AIC = 4731.7 | Accurate Activity Duration | 0.49 | -0.71 | 0.24 | -2.88 | 0.004 |
| | BIC = 4762.6 | | | | | | |
| $M_A$ | AUC = 0.603 | **QUIZ_ANSWER_CORRECT** | **2.18** | 0.78 | 0.21 | 3.71 | <0.001 |
| | LogL = -2433.0 | ADD_MEMO | 0.56 | -0.58 | 0.11 | -5.29 | <0.001 |
| | AIC = 4898.0 | ADD_HW_MEMO | 0.53 | -0.64 | 0.15 | -4.13 | <0.001 |
| | BIC = 4997.0 | | | | | | |
| $M_{S+G}$ | AUC = 0.777 | **Self-Explanation Length** | **18.44** | 2.91 | 0.22 | 13.43 | <0.001 |
| | LogL = -2027.7 | **Self-Explanation Similarity Score** | **10.11** | 2.31 | 0.20 | 11.79 | <0.001 |
| | AIC = 4075.4 | **Generated Handwriting Features Count** | **2.30** | 0.83 | 0.24 | 3.48 | <0.001 |
| | BIC = 4137.2 | Rubric_Step_Number | 0.62 | -0.48 | 0.04 | -13.58 | <0.001 |
| | | Self-Explanation Sequence | 0.13 | -2.02 | 0.16 | -12.86 | <0.001 |
| $M_F$ | AUC = 0.802 | **Self-Explanation Length** | **24.76** | 3.21 | 0.23 | 13.94 | <0.001 |
| | LogL = -1941.8 | **Generated Handwriting Features Count** | **15.98** | 2.77 | 0.31 | 8.82 | <0.001 |
| | AIC = 3933.7 | **Self-Explanation Similarity Score** | **9.15** | 2.21 | 0.20 | 10.87 | <0.001 |
| | BIC = 4088.4 | **Accurate Activity Duration** | **2.03** | 0.71 | 0.31 | 2.31 | 0.021 |
| | | ADD_MEMO | 0.75 | -0.28 | 0.13 | -2.15 | 0.031 |
| | | Rubric_Step_Number | 0.60 | -0.51 | 0.04 | -14.10 | <0.001 |
| | | OPEN | 0.52 | -0.65 | 0.28 | -2.34 | 0.019 |
| | | Self-Explanation Sequence | 0.26 | -1.34 | 0.18 | -7.34 | <0.001 |
| | | ADD_HW_MEMO | 0.06 | -2.84 | 0.26 | -10.88 | <0.001 |

model, which focused only on self-explanation features like the length of explanations, showed good accuracy (AUC = 0.740). This finding highlights that self-explanation traits are important markers of students' understanding and potential learning barriers, addressing Research Question 1.

In contrast, models that focused on operational aspects, such as the number of operations and total activity duration, showed moderate accuracy (AUC = 0.655). This suggests that operational engagement also plays a significant role in the learning process, addressing Research Question 2. While operational metrics alone showed moderate accuracy, combining them with self-explanatory data significantly improved the model's precision (AUC = 0.777). This shows the benefits of integrating these data types.

Most notably, our comprehensive model, which combined self-explanation and operational factors, achieved the highest accuracy (AUC = 0.802). This result confirms the effectiveness of a holistic approach that includes detailed operational, engagement, and self-explanation data. Overall, our findings highlight the important roles that both self-explanatory attributes and operational engagement play in identifying learning challenges in secondary school mathematics.

Interestingly, none of the $M_A$ features had a strong impact in the $M_F$ model. This can be explained by how the ground truth was established, primarily relying on self-explanation data. Self-explanation features directly influence the model's predictions, making them more prominent. In contrast, memo-related actions and quiz answers, which are indirectly related to the ground truth, have less impact when included with self-explanation features. This happens because the predictive power of the self-explanation features overshadows the contributions of the indirectly related features.

## Significant predictors of learning difficulties in online mathematics education in $M_F$

The 'Self-Explanation Length' emerged as a significant predictor with an odds ratio of 24.76. This means that students who provide longer, more detailed explanations are likely to understand complex mathematical concepts better. Self-explanation is critical in enhancing understanding by requiring students to articulate their thought processes, which reinforces their learning. Previous research has shown that detailed self-explanations promote better integration of new knowledge (Chi, 2000) and enhance both conceptual and procedural understanding in mathematics (Rittle-Johnson, 2017). This finding supports the idea that detailed self-explanations lead to deeper cognitive processing. This directly addresses RQ1, as it shows how detailed self-explanations can help identify and overcome learning impasses by revealing students' thought processes and areas of misunderstanding.

Additionally, the 'Self-Explanation Similarity Score' showed a substantial odds ratio of 9.15, reinforcing our methodological approach. This high odds ratio indicates a strong link between the similarity of a student's self-explanation to reference material and their understanding of mathematical concepts, addressing Research Question 1. The high predictive value of the similarity score suggests that students who can closely align their explanations with accurate reference material achieve better understanding. This aligns with the notion that high-quality self-explanations facilitate the recognition and transfer of knowledge by focusing on the structural aspects of content (Rittle-Johnson, 2006).

However, 'Self-Explanation Complexity' and 'Self-Explanation Sequence' did not significantly associate with step correctness, suggesting that these aspects might not align well with actual comprehension. This finding is relevant to RQ2, as it helps identify which aspects of self-explanation are less predictive of learning difficulties, thereby refining our understanding of effective self-explanation strategies. The ‘Rubric Step Number’ suggests that later stages in problem-solving, which involve more complex tasks, are more challenging for students, which is supported by a negative correlation coefficient of -0.2007, indicating that more difficult problems are associated with lower correctness rates.

The ‘Generated Handwriting Features Count’ was a significant predictor with an odds ratio of 15.98, highlighting the importance of active engagement in learning. This indicates that more operations reflect greater student involvement with the learning material, aligning with RQ2. Active engagement, as evidenced by frequent handwriting operations, indicates a higher level of interaction with the learning material, which is essential for effective learning. This supports findings that emphasize the role of active problem-solving and manipulation of information in understanding complex concepts (McNamara et al., 2004).

Handwriting features, including stroke count and timing patterns, were found to indicate cognitive effort and problem-solving, supporting VanLehn’s (1990) concept of impasse-driven learning. ‘Accurate Activity Duration’ also showed a positive relationship, suggesting that factors like the time spent on an activity and the number of pen strokes significantly influence learning outcomes. This finding is relevant to both RQ1 and RQ2, as it shows that sustained effort and engagement are crucial for overcoming learning impasses and improving comprehension.

## Qualitative evaluation in impasse detection outcomes

Our model works well in certain situations. For example, when students write detailed and well-structured self-explanations, the model effectively identifies what is missing or unclear in their understanding. It is particularly good at recognizing gaps in students’ knowledge when their explanations are comprehensive.

However, there are challenges when students cannot explain well. In such cases, the model tends to interpret this as a learning impasse, especially early in the problem-solving process. This suggests that our model is more effective for students who can clearly explain their thinking but struggles to accurately detect issues for those who are not used to writing self-explanations or lack such skills. To improve the model’s utility, it needs adjustments to better handle cases where the student’s explanation is less detailed or clear. Additionally, supporting students’ motivation and ability to write self-explanations is crucial. These adjustments would provide a reliable tool for identifying learning challenges across a wider range of student responses and problem-solving stages.

**Table 7** Quiz of Pythagorean theorem

| Question | Step | Sample Answer | Students' Self-explanation | Ground Truth | Pred |
|---|---|---|---|---|---|
| In the provided diagram, two semicircles with AC and BC as diameters are present. The chord AQ of the larger semicircle is tangent to the smaller semicircle at point P. Given that ∠APC=120° and the radius of the smaller semicircle is 6, calculate the area of the shaded region. Express your answer in the form of √3 - π. | 1 | To determine the shaded area, first divide the figure into various sections. Starting with △AOQ, using the dimensions obtained from the left diagram, the base and height can be determined, resulting in an area of $81r^3/4$. | To determine the length of side BC in triangle BPC, considering the triangle ratios of 90 degrees, 60 degrees, and 30 degrees, we draw a perpendicular line PK from point P to line BC. We then calculate the length using the area of triangle BPC. | Correct Step (1) | Correct Step (1) |
| | 2 | For the sector OQC, with a central angle of 60 degrees, the area is calculated as 9 × 9 × π × 1/6, which equals 27/2π. | Utilizing the triangle ratios for 90 degrees, 60 degrees, and 30 degrees, we determine the length of KC (the radius of the large circle). Next, we calculate the area of sector KPC. | Correct Step (1) | Incorrect Step (0) |
| | 3 | Additionally, the area of the smaller semicircle is calculated as 6 × 6 × π × 1/2, resulting in 18π. | We find the area of triangle AQK using ratios. | Incorrect Step (0) | Incorrect Step (0) |
| | 4 | Therefore, the area of the inclined section is given by $81r^3/4$ - 9/2π. | Finally, we calculate the total area of sector KPC plus triangle AQK minus the semicircle BC. | Incorrect Step (0) | Incorrect Step (0) |
| The question reads: "Please find the radius of the circumscribed circle for a triangle with sides of lengths 5, 6, and 7." | 1 | Draw auxiliary lines in the triangle inscribed in a circle. Designate the circumcenter as O, and let E be the intersection point between AO and the circumcircle. Also, designate the foot of the perpendicular dropped from vertex A to the base BC as D. | Find a similar triangle when considering the height with the base of the triangle as 7. | Incorrect Step (0) | Incorrect Step (0) |
| | 2 | Prove that the triangles △ABE and △ADC inside the circumcircle are similar. | Subsequently, solve it proportionally. | Incorrect Step (0) | Incorrect Step (0) |
| | 3 | Using the Pythagorean theorem, determine the height of △ABC. Let h be the height, BD=x, and find the values of x and h. | | Incorrect Step (0) | Incorrect Step (0) |
| | 4 | From △ABE ∽ △ADC, determine the radius based on the ratio of corresponding sides. With AB:AD=AE:AC, derive the radius R of the circumcircle as R=35r6/24 using the similarity ratio. | | Incorrect Step (0) | Incorrect Step (0) |

## Limitations and future work

Our study has a few key limitations. First, we rely heavily on the presence and quality of self-explanations. We assumed that well-written self-explanations mean there are no learning barriers, while poor explanations signal a problem. However, not all students are equally skilled at explaining their thoughts in writing, which can lead to misjudging their understanding. Some students might skip self-explanations because they feel confident in the material or lack the training to write detailed explanations. This reliance on self-explanation data can introduce biases. Additionally, our model's effectiveness depends on identifying consistent patterns in student behavior to predict learning impasses. This approach assumes that all students have similar learning styles and problem-solving methods, which is not always true. Poorly articulated or incomplete explanations can significantly reduce the model's accuracy in detecting impasses. Finally, our research focused on middle school mathematics, specifically on linear equations and functions, categorized into four steps. However, mathematical content is diverse and complex, requiring a more flexible approach to analyzing different concepts and strategies. The findings from this specific dataset and context may not apply to other settings or domains in mathematics (Ikawati et al., 2020). Future research should aim to broaden the scope to include a wider range of topics and educational environments, improving the model's adaptability and relevance.

Despite these limitations, our study has practical applications for educational settings. Features like 'Self-Explanation Length' and 'Self-Explanation Similarity Score' can be integrated into intelligent tutoring systems to provide real-time feedback and personalized learning paths. By monitoring these features, educators can identify students who need extra support and adjust their teaching strategies accordingly.

**Abbreviations**

BERT: Bidirectional Encoder Representations from Transformers; BERTScore: BERT based text generation evaluation Score; NLP: Natural Language Processing; LEAF: Learning and Evidence Analytics Framework; LAViEW: Learning Analytics Visualizations & Evidence Widgets; $M_S$: Model of Self-Explanation Quality Indicators; $M_G$: Model of Generated Handwriting Features; $M_A$: Model of Overall Level of Action; $M_{S+G}$: Model of Self-Explanation Quality Indicators & Generated Handwriting Features; $M_F$: The Final Model.

**Acknowledgments**

Not applicable.

**Authors' contributions**

RN conducted the created machine learning models and data analysis and drafted the initial manuscript. BF, TY, YD and KT provided insight, and editing of the manuscript. HO provided supervision for this research. All authors read and approved the final manuscript.

**Authors' information**

Ryosuke Nakamoto is currently a doctoral student in the Graduate School of Informatics, Kyoto University. His research focuses on using the collected self-explanations and examining how they can be used to support learning, including identifying students' stuck points and feedback to students in mathematics.

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.

Yiling Dai is a Program-Specific Researcher at the Academic Center for Computing and Media Studies, Kyoto 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.

Taisei Yamauchi is a doctoral student in the Graduate School of Informatics, Kyoto University. His research interests include mining to improve students' learning engagement and achievement with topic-based learning pattern analysis.

Kyosuke Takami is a senior researcher at Education Data Science Center, National Institute for Educational Policy Research (NIER) in Japan. He has prior expertise in the application of neuroscience in education and received a PhD degree from Osaka University. He has been a high school teacher for more than 9 years. His research interests include: learning analytics, educational data mining and educational technologies.

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 JSPS Grant-in-Aid for Scientific Research (B) JP20H01722 and JP23H01001, (Exploratory) JP21K19824, (Early Career) JP23K17012, (A) JP23H00505, and NEDO JPNP20006.

**Availability of data and materials**
Not applicable.

## Declarations

**Competing interests**
The author declares no competing interests.

**Author details**
[1] Graduate School of Social Informatics, Kyoto University, Japan

[2] Center for Innovative Research and Education in Data Science, Institute for Liberal Arts and Sciences, Kyoto University, Japan

[3] Academic Center for Computing and Media Studies, Kyoto University, Japan

[4] Education Data Science Center, National Institute for Educational Policy Research (NIER), Japan

Received: 12 January 2024 Accepted: 30 June 2024
Published online: 1 January 2025 (Online First: 23 July 2024)

**References**

Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Are you reading my mind?: Modeling students' reading comprehension skills with natural language processing techniques. In J. Baron, G. Lynch, N. Maziarz, P. Blikstein, A. Merceron & G. Siemens (Eds.), *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 246–254). ACM. https://doi.org/10.1145/2723576.272361

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, *57*(5), 13–19.

Arter, J., & Chappuis, J. (2007). *Creating and recognizing quality rubrics*. Pearson/Merrill Prentice Hall.

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*, *101*(1), 70–87. https://doi.org/10.1037/a0013247

Brown, J. S., & VanLehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, *4*, 379–426. https://doi.org/10.1016/S0364-0213(80)80010-3

Carroll, J. M., & Kay, D. S. (1988). Prompting, feedback and error correction in the design of a scenario machine. *International Journal of Man-Machine Studies*, *28*, 11–27. https://doi.org/10.1016/S0020-7373(88)80050-6

Chi, M., Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477.

Chi, M. T. H. (2000). Self-explaining: The dual processes of generating inference and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. 5., pp. 161–238). Erlbaum.

Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, *11*(4), 389–415.

Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers & Education*, *49*(3), 809–821. https://doi.org/10.1016/j.compedu.2005.11.018

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics.

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

Flanagan, B., Majumdar, R., & Ogata, H. (2022). Early-warning prediction of student performance and engagement in open book assessment by reading behavior analysis. *International Journal of Educational Technology in Higher Education*, *19*, 41. https://doi.org/10.1186/s41239-022-00348-4

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. https://doi.org/10.4324/9780203887332

Iiyama, M., Nakatsuka, C., Morimura, Y., Hashimoto, A., Murakami, M., & Minoh, M. (2017). Detecting answer stuck point using time intervals of pen strokes. *Transactions of Japanese Society for Information and Systems in Education*, *34*(2), 166–171.

Ikawati, Y., Al Rasyid, M. U., & Winarno, I. (2020). Student behavior analysis to detect learning styles in Moodle learning management system. In *Proceedings of 2020 International Electronics Symposium* (pp. 501–506). IEEE. https://doi.org/10.1109/IES50839.2020.9231567

Jackson, G. T., Guess, R. H., & McNamara, D. S. (2010). Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science*, *2*, 127–137. https://doi.org/10.1111/j.1756-8765.2009.01068.x

Jovanović, J., Saqr, M., Joksimovíc, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computers & Education*, *172*, 104251. https://doi.org/10.1016/j.compedu.2021.104251

Kishi, K., & Miura, M. (2018). Detecting learners' weak points utilizing time intervals of pen strokes. *International Journal of Learning Technologies and Learning Environments*, *1*(1), 61–77. https://doi.org/10.52731/ijltle.v1.i1.295

McEldoon, K. L., Durkin, K. L., & Rittle‐Johnson, B. (2013). Is self‐explanation worth the time? A comparison to additional practice. *British Journal of Educational Psychology*, *83*(4), 615–632. https://doi.org/10.1111/j.2044-8279.2012.02083.x

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 222–233. https://doi.org/10.3758/BF03195567

Nakamoto, R., Flanagan, B., Dai, Y., Takami, K., & Ogata, H. (2022). An automatic self-explanation sample answer generation with knowledge components in a math quiz. In M. M. Rodrigo, N. Matsuda, A. I. Cristea & V. Dimitrova (Eds.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium. AIED 2022. Lecture Notes in Computer Science, vol 13356* (pp. 254–258). Springer, Cham. https://doi.org/10.1007/978-3-031-11647-6_46

Nakamoto, R., Flanagan, B., Dai, Y., Takami, K., & Ogata, H. (2024). Unsupervised techniques for generating a standard sample self-explanation answer with knowledge components in a math quiz. *Research and Practice in Technology Enhanced Learning*, *19*, 016. https://doi.org/10.58459/rptel.2024.19016

Nakamoto, R., Flanagan, B., Takami, K., Dai, Y., & Ogata, H. (2021). Identifying students' stuck points using self-explanations and pen stroke data in a mathematics quiz. In M. M. T. Rodrigo et al. (Eds.), *Proceedings of the 29th International Conference on Computers in Education* (pp. 521–530). Asia-Pacific Society for Computers in Education.

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. In J. Epps, F. Chen, S. Oviatt, K. Mase, A. Sears, K. Jokinen & B. Schuller (Eds.), *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (pp. 583–590). ACM. https://doi.org/10.1145/2522848.2533789

Ogata, H., Flanagan, B., Takami, K., Dai, Y., Nakamoto, R., & Takii, K. (2023). EXAIT: Educational eXplainable Artificial Intelligent Tools for personalized learning. *Research and Practice in Technology Enhanced Learning*, *19*, 019. https://doi.org/10.58459/rptel.2024.19019

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In K. Inui, J. Jiang, V. Ng & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 3982–3992). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

Renkl, A. (2017). Learning from worked-examples in mathematics: Students relate procedures to principles. *ZDM*, *49*(4), 571–584. https://doi.org/10.1007/s11858-017-0859-3

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*(1), 1–15.

Rittle-Johnson, B. (2017). Developing mathematics knowledge. *Child Development Perspectives*, *11*(3), 184–190. https://doi.org/10.1111/cdep.12229

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM*, *49*(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*, 346–362. https://doi.org/10.1037//0022-0663.93.2.346

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.

Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies*. Lawrence Erlbaum Associates, Inc.

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, *36*, 404–411.

Thompson, D. R., & Senk, S. L. (1998). Using rubrics in high school mathematics courses. *The Mathematics Teacher*, *91*, 786–793.

VanLehn, K. (1987). Learning one subprocedure per lesson. *Artificial Intelligence*, *31*(1), 1–40. https://doi.org/10.1016/0004-3702(87)90080-4

VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. The MIT Press.

VanLehn, K., & Jones, R. M. (1993). Learning by explaining examples to oneself: A computational model. In S. Chipman & A. Meyrowitz (Eds.), *Cognitive models of complex learning* (pp. 25–82). Kluwer Academic.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *The Journal of the Learning Sciences*, *2*(1), 1–59.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, *21*(3), 209–249. https://doi.org/10.1207/S1532690XCI2103_01

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 5998–6008.

Verschaffel, L., De Corte, E., & Vierstraete, H. (1999). Upper elementary school pupils' difficulties in modeling and solving nonstandard additive word problems involving ordinal numbers. *Journal for Research in Mathematics Education*, *30*(3), 265–285. https://doi.org/10.2307/749836

Warli, Cintamulya, I., & Rahayu, P. (2020). Scaffolding process based on students diagnostic difficulties in proving group problems by using mathematics mapping. *Journal of Physics: Conference Series*, *1422*, 012012. https://doi.org/10.1088/1742-6596/1422/1/012012

Yang, A. C. M., Chen, I. Y. L., Flanagan, B., & Ogata, H. (2021). From human grading to machine grading: Automatic diagnosis of e-Book text marking skills in precision education. *Educational Technology & Society*, *24*(1), 164–175.

## Publisher's Note

The Asia-Pacific Society for Computers in Education (APSCE) remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.