STEM Education

STEM Education, 6(3): 467–495
DOI: 10.3934/steme.2026020
Received: 02 December 2025
Revised: 10 March 2026
Accepted: 17 April 2026
Published: 11 May 2026

https://www.aimspress.com/journal/steme

Research article

Variability in explanations of physical mechanisms among 10th-grade students when solving a STEM problem

Kenji López[1,]*, Marlenny Guevara[2], Valeria Cabello[3] and Jairo Montes[4]

[1] Pontificia Universidad Católica de Chile, Facultad de Educación, y Universidad del Valle, Facultad de Psicología, Cali – Colombia; jose.julian.kenji@gmail.com
[2] Universidad del Valle, Facultad de Psicología, Instituto de Investigación en Ciencias del Desarrollo, del Aprendizaje y Subjetividades (CIDEAS), Cali, Colombia; marlenny.guevara@correounivalle.edu.co
[3] Pontificia Universidad Católica de Chile, Facultad de Educación, Center for the Study of Policies and Practices in Education (CEPPE UC), Santiago de Chile, Chile; vmcabello@uc.cl
[4] Gaming & Learning, Cali, Colombia; jairoamontes@gmail.com

* **Correspondence:** Email: jose.julian.kenji@gmail.com; Tel: +57-313-765-2247.

Academic Editor: William Guo

Abstract: The development of scientific thinking during adolescence does not necessarily show a progressive linear sequence. The variability in students' explanations expressing scientific thinking organization requires a comprehensive examination in STEM education. The purpose of this study was to characterize such variability in 424 explanations provided by 53 10th-grade high school students when solving a STEM problem. Using a descriptive exploratory design, a microgenetic method was employed, with an intraindividual analysis of eight repeated measurements. Participants were assigned to solve a problem, which required linking physical mechanisms to achieve a goal in a virtual game. The complexity of participants' explanations in the game's dialogue windows was analyzed in relation to Kurt Fischer's theory. The five clusters showed nonlinear patterns. The first cluster had a score of 76.14 (n = 15) and showed a progressive increase. The second cluster scored 54.14 (n = 12) and showed intermediate fluctuations. The third cluster, characterized by low-complexity trajectories, had a mean of 39.54 (n = 10). The fourth cluster, 32.21 (n = 32.21), showed fluctuations between low and high scores. The fifth cluster, 17.55 (n = 7), showed a pattern of fluctuating scores throughout the sessions. These findings contribute to the field of research by considering the complexity of scientific thinking in problem-solving situations and by informing

science education to promote the understanding of physical concepts among secondary students.

Keywords: variability science education, thinking, problem-solving, explanations

1. Introduction

The present study examines the variability in the explanations constructed by students when solving a STEM problem based on physical mechanisms, as an approximation of their scientific thinking. Cognitive development exhibits different forms of progression and stability that are not linear but progressive. Additionally, regarding the development of scientific thinking, previous research indicates that a linear progressive sequence is not necessarily evident in adolescents [1]. This implies that there may be transitions and variability in the changes that occur during development.

Prior studies have highlighted the importance of designing science learning activities and assessments that respond to the various manifestations of cognitive variability among students [2]. Research suggests that science education benefits from temporally broad and detailed approaches, since multiple factors influence variability in thinking, such as the participant's personal characteristics, the context, the task to be solved, and the student's level of commitment or motivation, among others [3]. Without extended, process-oriented analyses of scientific thinking in educational contexts, interpretations may remain partial and fail to fully account for the interaction and dynamism involved in the development of students' thinking.

The promotion of scientific literacy and participation in science has been identified as a relevant objective in contemporary educational policy and research [4]. In the context of this study (Colombia), the Organization of Ibero-American States has advocated promoting and consolidating the study of integrated science and technology among young people. Consequently, STEM (science, technology, engineering, and mathematics) has emerged as a political-educational strategy [5]. Within this framework, STEM education has been described as a model that fosters creativity, motivation, and self-efficacy, particularly when implemented through inquiry-centered approaches [6].

The STEM approach seeks to develop students' skills in these disciplines and cultivate scientific thinking, preparing students to engage with technologically complex and globalized contexts [52]. An aspect frequently examined in STEM research is students' motivation toward science. Although science education is often linked to academic and professional trajectories, empirical studies report that some students express low interest or negative attitudes toward science classes [7–10]. In this regard, prior research has explored integrating STEM education with game-based and problem-based learning approaches as potential avenues to enhance engagement and motivation [50].

Variability has been studied using the microgenetic method [11,12]. Students produced written explanations across multiple problem-solving sessions that involved cognitively demanding tasks. These explanations were later coded to assess levels of complexity in scientific thinking [11]. The level of complexity of scientific thinking, as evidenced in students' explanations and responses while solving the problem, was coded.

Although prior microgenetic studies have examined the coordination of scientific concepts in small samples or single-case designs (e.g., [11,12,45]), and dynamic systems research has documented nonlinear developmental patterns in controlled or domain-specific contexts [31,43,58], more recent work has begun to explore these processes in virtual STEM environments. For instance, researchers documented progressive increases in the complexity of students' scientific reasoning across repeated sessions in a virtual problem-solving game, highlighting the role of task demands and interactive contexts in fostering more sophisticated reasoning [59]. However, despite these advances, there is still limited empirical evidence on how explanatory trajectories unfold in relation to the coordination of physical mechanisms, and on how variability in explanations can be characterized simultaneously at the intraindividual and interindividual levels within classroom-based STEM contexts using a theoretically grounded complexity framework.

By examining variability in students' explanations using a microgenetic approach, the present study aims to provide empirical evidence for ongoing discussions about developmentally sensitive science education. The aim is to characterize the explanations that 10th-grade students construct when solving a STEM problem involving physical mechanisms. Using a microgenetic method, an intraindividual analysis of the explanations constructed was conducted to answer the research question: How are the explanations constructed by 10th-grade students when solving a STEM problem involving physical mechanisms characterized by their intraindividual variability during the problem-solving process?

2. Theoretical frameworks

STEM education implies flexibility and curricular transversality [13]. Curricular transversality refers to the integration of knowledge from different curricular disciplines; in this case, science lessons that encourage students to develop competencies for establishing connections between disciplines rather than only within a specific subject [14]. Education that focuses on competencies requires the design of effective, attractive, technology-enriched educational environments that include problem-solving (especially when the solution is not apparent) [15]. Problem-solving is a process in which students understand the problem, formulate alternatives, plan and implement solutions, and explain and evaluate them [16,17]. Constructing explanations is a competence that helps make students' thinking visible when solving problem situations [18]. A previous study applied problem-solving tasks to 142 primary and secondary school students to assess their problem-solving skills. The study revealed that problem-solving skills are transferable and applicable not only to problems within a specific subject but also to cross-curricular situations, such as those in the STEM approach [15]. Another study, which evaluated a problem-solving intervention with 683 first- and second-year STEAM students, used questionnaires and observations. The results suggested that teachers can improve student engagement through non-routine problem-solving [19].

Students must become familiar with scientific concepts and procedures to develop science and technology learning. Furthermore, applying these scientific concepts and methods in practical situations and connecting them to students' prior knowledge are part of the constructivist approach to education, which holds that students actively construct cognitive structures by interpreting their experiences in specific situations [6]. Constructivist approaches to developing scientific and technological literacy seek to have students relate new concepts to their prior knowledge. Knowledge results from the process by which students take an active role and use prior cognitive structures to generate more complex knowledge [20]. This is consistent with the potential variability in the organization of scientific thinking, as the processes of relating prior knowledge to new knowledge can occur to varying degrees and at different times, depending on students' knowledge coordination [21]. This approach differs from the information processing approach, which considers knowledge as the complexity of information acquired, processed, and achieved through intentional actions [22].

In another study, 167 teachers followed a constructivist approach to teaching science and technology. Six elements were identified in three scales: 1) perceiving phenomena and formulating questions, 2) gathering information through tasks that involve research problems, 3) formulating hypotheses about possible answers to questions or solutions to problems, 4) examining evidence, 5) conducting experiments, and 6) analyzing the results obtained and drawing conclusions. They concluded that it was necessary to strengthen the elements of this approach in teaching scientific and technical content [20].

Developing explanations about physical phenomena is a process that expresses the scientific thinking of students and teachers [23,24]. Currently, in various educational contexts, scientific thinking through technology-supported activities is enhanced to develop scientific vocations [25–27]. However, explaining and using technology is required in all students' professional and academic lives and future workplaces [28]. Thus, constructing explanations in science education is a key element in learning and participating in the languages of science for both children and young people.

Scientific thinking involves both declarative knowledge (verbal concepts) and procedural knowledge (skills or performance) [29], such as experimenting, predicting, controlling variables, and explaining [30]. For instance, when faced with the accelerated movement of a car on an inclined plane, a student may respond such as: "The car rolls fast", "The car moves by force", or "The car moves faster when there is less friction". Although these explanations differ in their level of elaboration, the same student may produce them at different times depending on factors such as the task demands, the interaction context, or motivation.

The coexistence of explanations of varying complexity should not be interpreted as inconsistency but rather as an expression of intraindividual variability. From a microgenetic perspective, cognitive development is characterized by the coexistence and competition of multiple strategies within an individual's repertoire [35]. In this framework, variability constitutes an indicator of the process of change rather than a sign of error or regression. Development does not occur linearly, but rather involves fluctuations, partial reorganizations, and interactions between prior and emerging forms of thought [31,32].

In this study, variability is operationalized as the presence, coexistence, and intraindividual alternation of explanations that differ in conceptual complexity during the resolution of the same task. This definition is distinct from conceptual progression or regression. Progression implies a relatively stable directional shift toward more complex and consistent forms of explanation, while regression involves a predominant return to less elaborate levels of explanation. In contrast, variability does not presuppose direction or stability, but rather dynamic fluctuations within the student's explanatory repertoire, consistent with the principles of microgenetic analysis of cognitive change [32].

Scientific explanations can be statements of what phenomena are and why or how they occurred [30]. Observation and explanations are related processes. Observation is the recognition of relevant phenomena, while explanations systematically report what is observed and account for causal relationships [33]. Explanations are part of scientific thinking. One study analyzed how students construct explanations for physical phenomena. To do so, they posed 15 problems to two groups of high school students in grades 7–10 in a physics class. The problems involved pendulums with different masses and inclined planes. The students constructed their causal explanations. Initially, they mentioned isolated concepts, but over time, they made connections between magnitudes (force and energy) related to the reasons involved in certain problem situations [34].

Although the present task focuses on physical mechanisms (e.g., inclined plane, pulley, pendulum, catapult), the analysis of students' explanations is grounded in a developmental skill framework. Within this study, a mechanism is defined as a coordinated system of interacting components that involves forces, energy transfer, structural constraints, and sequential causal relations to produce a physical effect. Explaining a mechanism, therefore, requires more than naming isolated concepts; it involves coordinating magnitudes (e.g., force, distance, acceleration), specifying relationships among them, and organizing them into coherent causal chains.

Fischer's skill theory provides a theoretically appropriate framework for analyzing such mechanistic explanations because it conceptualizes development as the progressive coordination of representational units into increasingly complex hierarchical systems. From this perspective, higher levels of complexity correspond to integrating multiple variables into structured causal systems, which aligns directly with the cognitive demands of explaining chained physical mechanisms. Thus, the coding scheme derived from Fischer's levels captures not merely correctness, but the structural organization and coordination of concepts within students' explanations.

Furthermore, the task is situated within a STEM framework, not solely due to its digital platform. It requires integrating scientific concepts such as force, motion, and energy. It also involves technological mediation through interaction with a virtual simulation. In addition, it incorporates engineering-like problem-solving, as students must design functional chains of mechanisms to reach a goal. The task also demands mathematical reasoning, particularly in coordinating magnitudes and proportional relations. In this sense, the activity represents an integrated STEM problem-solving context rather than a traditional single-discipline physics exercise.

Cognitive developmental psychology has examined variability in students' thinking and performance trajectories using the microgenetic method. These analyses have improved our understanding of cognitive processes by examining performance trajectories during task resolution [35]. Trajectories refer to possible trends or regularities in knowledge that become evident over time during task performance [36]. Related to the previous example, a change in the trajectory of a student's responses can be observed when the student is asked why a car moves. Initially, the student may give a limited description, such as "The car moves". However, as they reframe their explanatory model, they may express one or more elaborate responses, including causes, for example, "It moves because of the engine" [37].

Researchers used the microgenetic method to analyze variability in the use of physics concepts [11]. Their study focused on how five 16- and 17-year-old students applied the concepts of force and dynamics in a practical scenario involving three physical mechanisms (a pendulum, a device, and a concave container for observing a ball's oscillations). Participants were assessed in 22 sessions, spread over six months, each lasting approximately 22 minutes. The results showed different patterns of coordination of physics concepts among the participants.

Other researchers analyzed the trajectories of control variables of 136 university students when solving a problem involving rectilinear motion at constant velocity [12]. In this scenario, students had to manage the variables of the car's velocity and duration as it traveled a given distance. The results indicate a diversity of procedures used by the students, and the performance of 10 participants suggests different approaches to problem-solving.

In the two studies described above, variability was evident. Repeated measurement processes revealed distinct trajectories and problem-solving skills across diverse groups. However, the studies did not detect characteristics of cognitive functioning within a discipline, which might yield a score indicating the student's success or failure on a particular test (see, for example, [38,39]). Taking trajectories into account reveals dynamic and varied processes that reflect the students' cognitive reorganization, moment by moment, as they solve the problems.

A previous study shows evidence of a nonlinear progression in scientific thinking [51]. A single measure of students' thinking may provide a limited view of their capabilities as it does not adequately account for teachers' significant efforts to support their students' learning in science. The valuable role of the microgenetic method in analyzing variability in scientific thinking due to its temporally broad and detailed approximation of participants' thinking processes has been reported in several studies [11,12,40,41].

In summary, the literature review indicates that some studies on students' explanations have focused on science-related problem-solving skills [14,15], demonstrating good problem-solving skills in high school and university students. Regarding the topics addressed, explanations and control of variables during experimentation exercises, as well as physical concepts such as speed and motion, have been explored [12]. Regarding the methodology, most studies used single measurements, showing a static view of the students' science skills [38,39]. Meanwhile, studies using microgenetic methods (repeated measures) have shown different trajectories of change in the complexity of coordination among physical concepts and problem-solving [11,30]. In contrast to previous studies that addressed explanations and concepts of physics separately, the present study seeks to explore both the explanations and the underlying concepts involved in the functioning of simple machines in the physics domain through a virtual problem-solving context. This approach seeks to address the need to strengthen science and technology-related content. For example, teachers declared this topic crucial for science education [20].

Consequently, in the current study, we seek to answer the following research question: How is variability characterized in the explanations provided by 10th-grade students when solving a STEM problem involving physical mechanisms? Addressing this question contributes to the discussion of the variability and nonlinearity of students' explanations when solving STEM problems. We know that the ability to explain and use technology is fundamental to all students' professional and academic lives. Therefore, it is essential to promote sound science education, given the dynamic nature of learning, which could contribute to better teaching. The results of this study are likely to guide the development of educational activities, evaluations, and interventions that are adaptable to the students' cognitive processes. Thus, more efficient education can be proposed, informed by research.

3. Methods

Given that prior research has shown that a single measurement captures the student's skills in a static, finished, or fulfilled way [38,39], and that repeated measures allow exploring the variability and complexity of the student's cognitive processes, other variables need to be taken into account. The types of questions asked when solving the problem, the complex and multivariate nature of the situation to be solved, and the types of explanations provided by students can provide clues for better understanding the processes of change and variability of students' scientific thinking.

As students' explanations might change slightly, the microgenetic method was selected as the primary analytic tool in the current study, following recommendations from previous research [18,42–45]. In practical terms, it involved eight observation sessions, allowing for an intraindividual analysis of explanations in a game situation oriented toward solving a STEM problem. The study design was non-experimental, exploratory-descriptive, with repeated measures in the same group of participants, using a microgenetic approach. Although it does not correspond to a classic long-term longitudinal design, the study involves intensive intraindividual follow-up over eight sessions distributed over four months. This type of design allows for the examination of change processes in relatively short periods through successive observations, fulfilling the conditions of the microgenetic method, as follows: 1) observations during periods of change until a certain stability is reached; 2) a large number of measurements; and 3) in-depth analysis of the study variables [46]. In this sense, the study is oriented toward the dynamic analysis of the intraindividual trajectories of explanations in the context of STEM problem-solving. By conducting eight observations over four months, the microgenetic method generates a framework that allows visualization of the different trajectories and the participants' explanations. The participants' explanations can vary across sessions, giving rise to dynamic trajectories. Line graphs were used to visualize the trajectories and the variability in the complexity of the explanations over time. An analysis of the trajectories of the 53 participants' explanations was conducted using clustering [21]. The clusters were based on distances along the trajectories in a two-dimensional field, and the centroids were defined using the K-means technique.

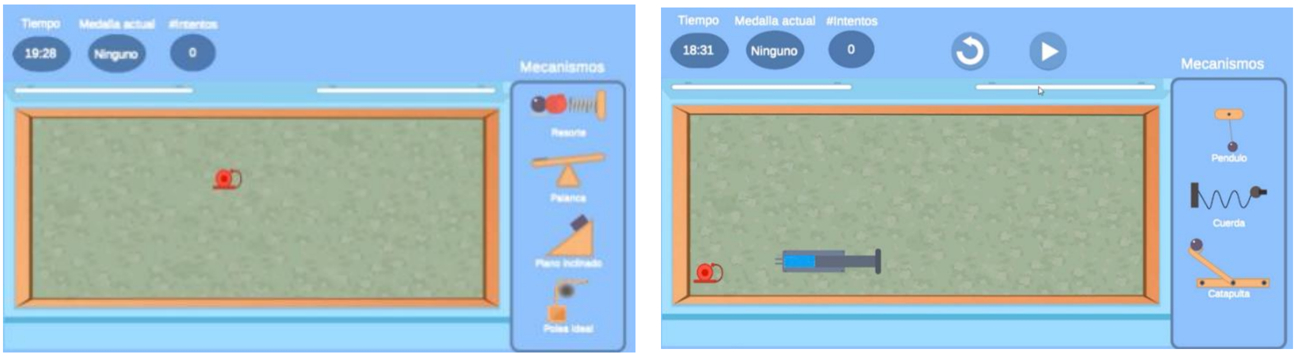
3.1. Participants

The sample consisted of 53 10th-grade students (28 women and 25 men, mean age = 15.4 years; SD = 0.63). Participants were selected through convenience sampling [47] at the secondary education level and, at the time of the study, were attending two private schools in Cali, Colombia. An essential factor in sample selection was participants' access to the internet and computers, given the virtual nature of the problem-solving situation, "Scientific Machine". Considering the limitations of some social contexts in providing these conditions, two private schools were selected for this study to ensure that participants had easy access to technological resources in their daily lives. Because the participants are minors, their parents signed informed consent forms, accepting their children's participation in the study. The students provided informed consent to participate voluntarily in the study. This research was governed by the deontological code of the psychologist in Colombia, specifically Law 1090 of 2006. It also had the ethical endorsement of the university that the researchers belong to.

During the data collection period (four months), participants continued attending their regular physics classes in accordance with the institutional curriculum for 10th grade. The study did not include additional didactic interventions or experimental manipulations related to these topics. Physics instruction may influence the trajectories of students' explanations. This potential influence is outside the scope of the present study. All 53 participants completed all eight sessions and provided explanations; there were no missed sessions or unanswered responses in the study.

3.2. Setting

For this study, a virtual problem situation was designed in the "Scientific Machine" format to examine the scientific thinking of high school students regarding physical concepts such as gravity, accelerated motion, and parabolic motion, among others (see Figures 1a and 1b).



(a) (b)

Figure 1. (a) Scientific Machine interface 1; (b) Scientific Machine interface 2.

The design of the problem situation considered physical concepts aligned with Colombian educational standards for tenth grade, in accordance with the basic standards of competencies in sciences of the Ministry of Education of Colombia. The Scientific Machine consists of activating a bell using the mechanisms available on the right side of the game screen. The problem situation has two interfaces that differ in their mechanisms. Interface 1 comprises a spring, a lever, an inclined plane, and a pulley (Figure 1a). Interface 2 is composed of a hydraulic piston, a rope, a pendulum, and a catapult (Figure 1b). The idea behind the problem situation is to generate a chain reaction among the available physical mechanisms that activate the bell, similar to a Rube Goldberg machine. To do this, the participant must use available physical mechanisms, perform calculations, coordinate simple machines to generate the chain reaction, and know physics concepts such as parabolic motion, gravity, speed, and force. The problem situation was presented individually to each participant. In the following link, you can see a video illustrating the interface of the problem-solving situation: https://www.youtube.com/watch?v=yXodGM_gJXI.

To ensure comparability between the two interfaces, both versions of the task were designed to maintain the same problem structure, goal conditions, and underlying physical principles. In both interfaces, participants were required to construct functional chains of mechanisms to achieve the same objective within the virtual environment. The available elements and their functional properties were equivalent, and the task required the coordination of the same core concepts, including force, motion, energy transfer, and proportional relations. The differences between the interfaces were limited to aspects of visual organization and interaction layout within the digital environment. Therefore, the conceptual demands and level of problem-solving difficulty remained constant across interfaces, allowing the analysis to focus on the variability in the complexity of students' explanations rather than on differences in task structure.

To ensure the reliability and validity of this study, diverse actions were implemented. First, the Scientific Machine involves physical topics corresponding to the participant's grade level. It presents issues students already know about in a design that tries to familiarize them. This decision was made to assess the validity of the measurement of explanations that do not generate difficulties in participants' understanding. Second, two physics expert teachers suggested and endorsed the physical mechanisms and variables considered for each mechanism (weight, speed, force, etc.). The experts' review of this learning scenario was a means of working toward construct validity. They identified the concepts and mechanisms in the Scientific Machine, and we calculated the inter-rater agreement as the first reliability indicator. The experts were 100% in agreement with identifying the concepts in 6 of the eight mechanisms. In the hydraulic piston, Expert 2 did not identify horizontal launch as a concept specific to this mechanism, resulting in 87.5% agreement. In the rope, Expert 1 did not identify the concepts of wave speed and medium characteristics, resulting in 87.5% agreement on the concepts (see Table 1).

Table 1. Agreement of concepts recognized by experts in the Scientific Machine.

Mechanism	Concepts recognized by Expert 1	Expert 2	% of agreement
Inclined plane	7	7	100%
Piston	8	7 Unidentified concept: horizontal launch	87.5%
Spring	5	5	100%
Lever	6	6	100%
Pulley	4	4	100%
Catapult	7	7	100%
Pendulum	5	5	100%
Rope	5 Unidentified concept: relationship between wave speed and characteristics of the medium	6	87.5%

Third, the Scientific Machine was designed by a systems engineer, taking into account the instructions of the two physics expert teachers. It ensured the software functioned correctly and stored the information accurately, which was necessary for the rigor of further analysis. Fourth, the items asking the participants to solve the Scientific Machine were approved by the two physics expert teachers. Fifth, a pilot test was conducted with a sample of 25 students to assess their understanding of the prompts and their ability to register data. This test verified that the problem situation was well structured, items were clear, and the necessary results could be compiled and stored safely. To ensure construct validity of the rubric, we adapted the categories based on a well-recognized prior work (Kurt Fischer's theory of skills, 1997) [48] and checked against other rubrics used to analyze explanations as a proxy in the learning sciences [11]. The coding process was carried out in several stages. First, the unit of analysis was defined as each verbal explanation the student provided during problem solving. Subsequently, the operational criteria for each rubric category were discussed, and prototypical examples and cases were analyzed. An initial independent coding phase was performed on a subset of data to refine the categorical definitions and address any ambiguities. Lastly, to validate the rubric defined as a coding scheme for this study, a double-coding procedure was conducted by two trained researchers. We used the Kappa test to establish the level of agreement between the coders when rating 20% of the explanations. A Kappa coefficient ($\kappa = 0.957$)

was obtained. An “almost perfect” agreement between observers was reached (0.81–1.00) [49]. Discrepancies between coders identified during the double-coding procedure were discussed until consensus was reached, and the final decisions were incorporated as definitive criteria for coding the entire dataset.

3.3. Data collection procedure

The two Scientific Machine interfaces were presented alternately, starting with interface one, then applying interface two, and repeating this pattern until session 8. This means that interface one was given in sessions 1, 3, 5, and 7, while interface two was implemented in sessions 2, 4, 6, and 8. When introducing each game interface, each participant was told that the objective was to activate the bell (red bell) that would appear on the game screen. Participants were told to drag physical mechanisms on the right side of the game screen onto the game board with the cursor, where they could be linked to activate the bell. Participants were given time and indicators of the number of attempts. The Undo button allowed them to return the mechanisms to the right window.

In contrast, the Play button allowed them to activate and observe the mechanisms and the sequence of movements determined by their configuration on the screen. Participants were given 20 minutes to solve the virtual problem situation with unlimited attempts. If time ran out, the participant’s performance was saved up to that point. Each participant was asked to answer the following questions, presented in a text box on the screen: “Why did you build the Scientific Machine this way? Why do you think it was not possible to activate the bell? Why do you think it was possible to activate the bell?” The questions were to be answered during each session. Of the three explanations offered by participants in response to the three questions, one—the most complex—was selected for each session. This means that each participant had eight explanations per session throughout the study. A total of 1272 explanations ($3 \times 8 \times 53$) were recorded in the study, one for each question in each session. Since the most complex explanation was selected per session, 424 explanations ($1 \times 8 \times 53$) were coded. The participants’ data and written explanations were recorded in an Excel file. The sessions were conducted at 15-day intervals during a 4-month data collection process.

Regarding the number of sessions, previous studies using the microgenetic method suggest that, starting with four observations, it is possible to clearly identify trajectories of change and stabilization in cognitive processes, especially when measurements are taken during periods of active transformation [11,18,40]. However, in the present study, eight sessions were chosen to increase analytical sensitivity and capture intraindividual fluctuations, possible reorganizations, and nonlinear patterns in the complexity of the explanations. Eight sessions were considered sufficient to capture variability in the complexity of students’ explanations across the problem-solving process. The repeated measures across eight sessions provided multiple opportunities to observe changes, regressions, and advances in the organization of scientific reasoning within individuals. Moreover, by considering explanations as the unit of analysis, the eight sessions generated a sufficiently large corpus of responses for the cluster analysis. In studies of cognitive processes in learning contexts, sequences of approximately 6–10 sessions are commonly used to capture dynamics of change while minimizing potential fatigue or attrition effects [1,2]. As for the 15-day interval between sessions, this decision was based on methodological and pedagogical criteria. From a methodological standpoint, the time spacing enabled observation of cognitive reorganization processes that transcend the immediate effects of repetition or short-term memory. From a practical standpoint, the interval was designed to minimize potential fatigue or demotivation associated with the repeated presentation of the same problem, thereby promoting sustained student participation throughout the study.

3.4. Data analysis

For the analysis, a coding scheme was used to determine the complexity of the explanations based on Kurt Fischer's theory of abilities (see Table 2). Students' explanations reflect the declarative knowledge of scientific thinking demonstrated during the game and were coded according to their level of complexity. A cluster analysis of the different variability trajectories of the students' explanations in the study was conducted.

The coding scheme presented in Table 2 is a direct operationalization of Kurt W. Fischer's (1997) theory of skills. Consistent with this framework, the level of complexity assigned to each explanation was not defined by the amount of information provided or the conceptual correctness of the content, but rather by the cognitive structure underlying the explanation. Specifically, the degree of hierarchical coordination among the elements mentioned by the student was analyzed. Sensorimotor-level explanations describe the overall functioning of the mechanism without differentiating internal relationships.

Table 2. Coding scheme for declarative knowledge of scientific thinking, based on Kurt Fischer's theory of skills (1997).

Category	Example explanations	Complexity level score
Sensorimotor: Explanation that describes the mechanism(s) as a whole, based on its general physical or operating characteristics	"Because I calculated the spaces and distance well, each, so that each mechanism was activated, and the bell rang." "The machine activated when the ball fell."	1
Simple representation: Explanation that describes specific attributes of/between mechanism(s)	"The lever needed more force to ring the bell." "The distance wasn't enough."	2
Relationship of representations: Explanation in which two or more relationships or specific attributes of/between mechanism(s) are coordinated	"The 62° angle of the catapult caused the ball to reach the piston to activate it. The high piston pressure was adequate to activate the chain." "The pressure was enough because the distance was short."	3
Abstract representation: Explanation in which several relationships of/between mechanism(s) are coordinated	"When the lever arm is in favor of the pulley, it needs less force, so when it falls on the lever arm, there is greater torque and the load increases." "The relationship between distance and force shows a mechanical advantage that allowed the mechanism to activate."	4

Simple representations refer to specific, isolated attributes. Representational relationships involve the explicit coordination of two or more attributes or causal relationships. Abstract representations integrate multiple relationships under more general principles (e.g., force, pressure, torque). Thus, the coding focused on the structural organization of the explanation as an empirical indicator of the skill level demonstrated in the game situation, ensuring alignment between the model's theoretical levels and the observable indicators in the students' discourse.

To increase transparency in the coding process, a decision rule was established for borderline cases. Explanations were coded as score 2 when students mainly described isolated elements, actions, or observable features of the mechanisms without explicitly coordinating them conceptually. Score 3 explanations included explicit relations between two elements or physical concepts (e.g., describing how one mechanism affects another). Score 4 explanations were assigned when students articulated causal relations involving multiple mechanisms or coordinated representations that explained how the system operates as a whole.

Once the complexity level of the participants' explanations in each session was established, the percentage of that complexity was estimated. This allowed us to present the results in a condensed form and visualize the variability in participants' trajectories across the eight sessions. A two-factor ANOVA was performed to determine whether there were differences among interfaces of the problem situation. This analysis aimed to compare whether the means of two or more normal probability distributions with the same variance were equal [53]. The statistical assumptions for the application of this test were met. The two factors involved were the interface (1 or 2) and the scores for participants' explanations across eight sessions. A cluster analysis of the different variability trajectories of the students' explanations in the study was conducted.

In all sessions, participants provided at least one explanation related to the task's resolution. No missing responses or sessions without explanatory verbal production were identified. Consequently, the database was complete for all participants and measurement points. It was not necessary to apply imputation procedures or statistical strategies for handling missing data.

3.5. Ethical considerations

This study followed the Ethical Guidelines of Universidad del Valle. Accordingly, the planned data collection strategies posed no physical, moral, mental, emotional, or social risks to the participants, now or in the future. The study did not expose them to situations that risked their dignity or physical or emotional integrity. Furthermore, none of the data collection instruments or procedures involved any type of discrimination (based on gender, creed, nationality, ethnicity, or socioeconomic status).

Only those who voluntarily and explicitly expressed their interest in participating through informed consent were included as participants in this study. As the participants were minors, their parents signed the authorization for them to participate in the study. The participants signed the assent for their participation in the study. Similarly, all participants were free to withdraw from the study at any time they deemed appropriate. Absolute confidentiality of the collected information was maintained and guaranteed. Participants' names were not disclosed; codes were used. The study was approved by the ethics committee, as stated in the approval certificate (Ethics Committee of the Universidad del Valle: protocol 108-021).

The participants' database, their explanations during the study, and their informed consent forms were stored by the researchers. This information is contained in digital files in Excel and Word format, stored on the researchers' personal computers. Access to and manipulation of the information were restricted to the researchers and authors of this article.

4. Results

The purpose of the ANOVA was to establish whether there were differences between the explanations of the two interfaces of the task. The ANOVA analysis suggested an interaction between the interface (1 or 2) and the participants' explanation scores, with an $F(1, 10) = 0.921$. The results showed a nonsignificant main effect of the interface, $F(1, 52) = 0.01$, $p = 0.921$, $\eta^2p < 0.001$, indicating that there were no overall differences in the complexity of the explanations between interfaces. This indicates no significant interaction between the interface and the scores obtained in the sessions. To examine differences in performance across sessions and interfaces, a repeated-measures analysis of variance (ANOVA) was conducted. The analysis reports the full statistical parameters, including F values, degrees of freedom, p-values, and partial eta squared (η^2p) as a measure of effect size. Assumptions associated with repeated-measures analyses were evaluated before interpretation. Sphericity was assessed using Mauchly’s test. When the assumption of sphericity was violated, Greenhouse–Geisser corrections were applied to adjust the degrees of freedom.

Consequently, there were no significant differences between the two interfaces. The above result shows vital reliability, as both sets assess participants' explanations and physical concepts in similar ways. Therefore, successive measurements could be administered session by session, even if the sets were staggered. Since there were no significant differences among the game scenarios, the analyses presented below focus on accounting for the complexity of explanations using a temporal criterion to answer the question: How is the variability in the explanations provided by 10th-grade students characterized when solving a STEM problem involving physical mechanisms?

To examine whether the two task interfaces produced different levels of explanation complexity across sessions, a repeated-measures analysis was conducted. The results indicated no statistically significant interaction between interface and session. However, in line with current reporting recommendations, effect sizes and confidence intervals were examined in addition to p-values. The observed effect sizes were small, suggesting that any potential differences between interfaces were minimal in practical terms. Confidence intervals around the estimates also indicated substantial overlap between conditions. Therefore, rather than claiming strict statistical equivalence, the findings suggest that both interfaces functioned similarly in supporting students’ explanations within the context of the task. A visual representation and description of the second interface have been included to facilitate comparison between the task environments.

The groups were established by analyzing the explanatory trajectories of the participants during eight-game sessions while they solved the virtual problem situation, the Scientific Machine, using the group technique (K-means). Data on the complexity of verbalizations were rated on a scale of 1 to 4, with 4 indicating the highest level of complexity (see Table 2) and analyzed using TANAGRA 1.4.48 software [54]. The clustering technique brings together similar data trajectories. As a result, five trajectory groupings were identified based on changes in the complexity of participants' explanations across the game sessions. The number of clusters and their sizes were determined based on the trends in the identified trajectories, in this case, 5. The different trajectories of the explanations were grouped according to the weights of each cluster, and the silhouette scores were calculated (see Table 3). Subsequently, each identified cluster was described, analyzed, and then ordered from highest to lowest number of cases, according to the observed trend in variability.

Cluster analysis was conducted to identify patterns of trajectories in the complexity scores across the eight sessions. Although the complexity scale ranges from 1 to 4 and therefore represents ordinal data, the K-means algorithm was used as a pragmatic approach to identify general trajectory patterns across repeated measures. To examine the robustness of the clustering solution, an additional clustering procedure using the k-medoids algorithm was conducted. This method is more robust for ordinal data and less sensitive to extreme values. The resulting grouping structure showed patterns consistent with the clusters identified using K-means, supporting the stability of the identified trajectories.

Table 3. Cluster quality indices for different cluster solutions.

Clusters	No. of participants	Average silhouette (original scale)	Average silhouette
2	12	068	68.0
3	10	0.76	76.104
4	9	0.71	61.2
5	7	0.55	55.3

Note. The silhouette index ranges from −1 to 1. Values closer to 1 indicate better separation between clusters. For readability, silhouette values are also presented multiplied by 100.

The final solution of five clusters was selected after comparing different partitions derived from K-means analysis and examining both statistical and conceptual criteria. Statistically, cluster sizes and silhouette indices (Table 3) were considered, indicating adequate levels of cohesion and separation. Although Cluster 5 presented a lower silhouette index, its highly differentiated trajectory pattern, characterized by pronounced oscillations and recurring peaks at levels of abstraction, justified its retention as an independent group. Conceptually, solutions with fewer clusters tended to merge temporally distinct trajectories, reducing the analysis's sensitivity to capture the microgenetic variability observed throughout the eight sessions. Consequently, retaining five clusters preserves the structural diversity of the explanatory trajectories and a more accurate description of the differential dynamics in the complexity of scientific thinking manifested by the participants.

The quality of the cluster solutions was examined using internal cohesion criteria (intra-cluster sum of squares) and relative separation between groups, comparing solutions with different numbers of centroids (see Table 3). In the absence of global fit indices, typical of parametric models, a solution that balanced parsimony, statistical stability, and theoretical interpretability was prioritized. This strategy is consistent with microgenetic studies that seek to identify emerging patterns in individual trajectories without assuming predefined latent structures.

Table 4 shows the mean complexity scores per session. This result reveals a progressive increase throughout the sessions. Overall, these results demonstrate a general upward trend in the complexity of the explanations.

Table 4. Mean complexity scores per session.

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7	Session 8
Means per session	1.67	1.69	2.07	2.35	2.66	2.45	3.13	3.4

Figure 2 illustrates the variability trajectories of explanation complexity for participants in Cluster 1. This type of graph shows the temporal changes in explanation complexity across the eight-game sessions. As a result, the explanation trajectories exhibit contrasting patterns; some show an increase in complexity, and others show a decrease across sessions. In the initial sessions, this pattern of variability shows a high peak of explanations (53%) that account for descriptions (score 1) about physical concepts. However, this decreases to half in the intermediate sessions, disappears in session 7 (0%), and is presented at low level in session 8 (7%). Across sessions, explanations focused on descriptions (score 1) decreased, and explanations focused on abstract representations (score 4) increased. In summary, for Cluster 1, the types of explanations provided by participants indicate a transition from understanding the physical mechanisms of the task, focused primarily on describing physical concepts, to understanding that emphasizes abstract representations. In other words, the participants varied in their approaches to establishing causal relationships among the physical mechanisms that configure and operate the Machine, thereby allowing it to solve the problem.

Figure 3 shows the different trajectories of explanations across complexity levels. This is shown for the 12 participants of Cluster 2. Temporal changes in the complexity of explanations are evident throughout the eight gaming sessions. As a result, the trajectories of explanations exhibit a W-shaped pattern. In the initial sessions, this pattern of variability shows a high peak in explanations (75%) characterized by descriptions (score 1) about physical mechanisms. However, this peak decreases until it disappears in session 6. Throughout the sessions, while explanations focused on descriptions (score 1) decreased, explanations focused on relations of simple representations (score 3) and abstract representations (score 4) increased.

In summary, for Cluster 2, the type of explanations provided by participants indicates variability with transitions from understanding the physical mechanisms of the task, focused mainly on describing physical mechanisms, to understanding, focused on relations of simple representations and abstract representations. Figure 3 shows the above, where the trajectories of the explanations tend to go upward. In addition, the result indicates that the participants established relationships between the physical mechanisms and their functioning to solve the problem.

Figure 4 shows trajectories of explanations by complexity levels for the 10 participants of Cluster 3. This type of graph shows the temporal changes in the complexity of the explanations throughout the eight game sessions. As a result, the trajectories of the explanations show a U-shaped pattern. In the initial sessions, this pattern of variability shows a high peak of explanations (70%) that account for descriptions (score 1) of the physical mechanisms. However, the peak decreases until it disappears in session 6. While the explanations focused on descriptions (score 1) decreased throughout the sessions, the reduced ones concentrated on relationships between simple representations (score 3) increased.

In summary, for Cluster 3, the type of explanations provided by the participants indicates a variability from understanding the physical mechanisms of the task, focused mainly on the description of physical mechanisms, to understanding and concentrating on relationships between simple representations. This indicates that participants established relationships between physical mechanisms and their functioning to solve the problem. In comparison, the variability of the explanations for Cluster 1 is more pronounced across the eight sessions. In contrast, for Cluster 2, the variability of the explanations is more concentrated between sessions 4 and 8.

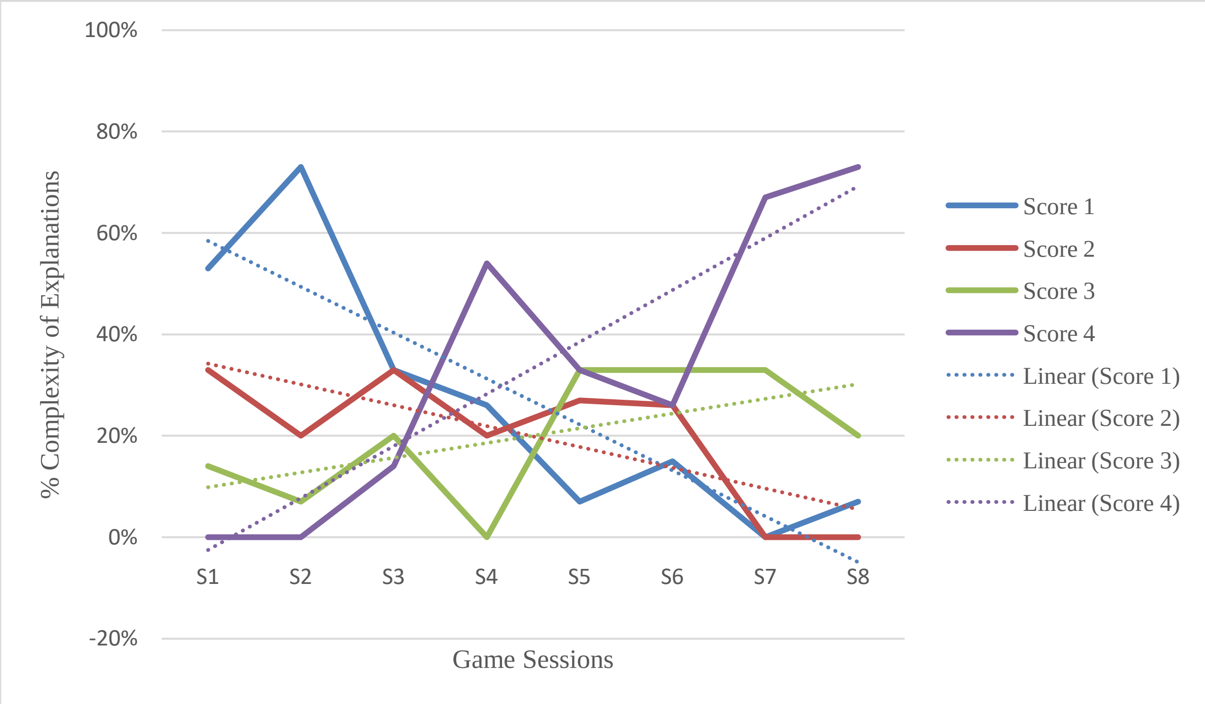


Figure 2. Trajectories of the complexity of explanations: Cluster 1.

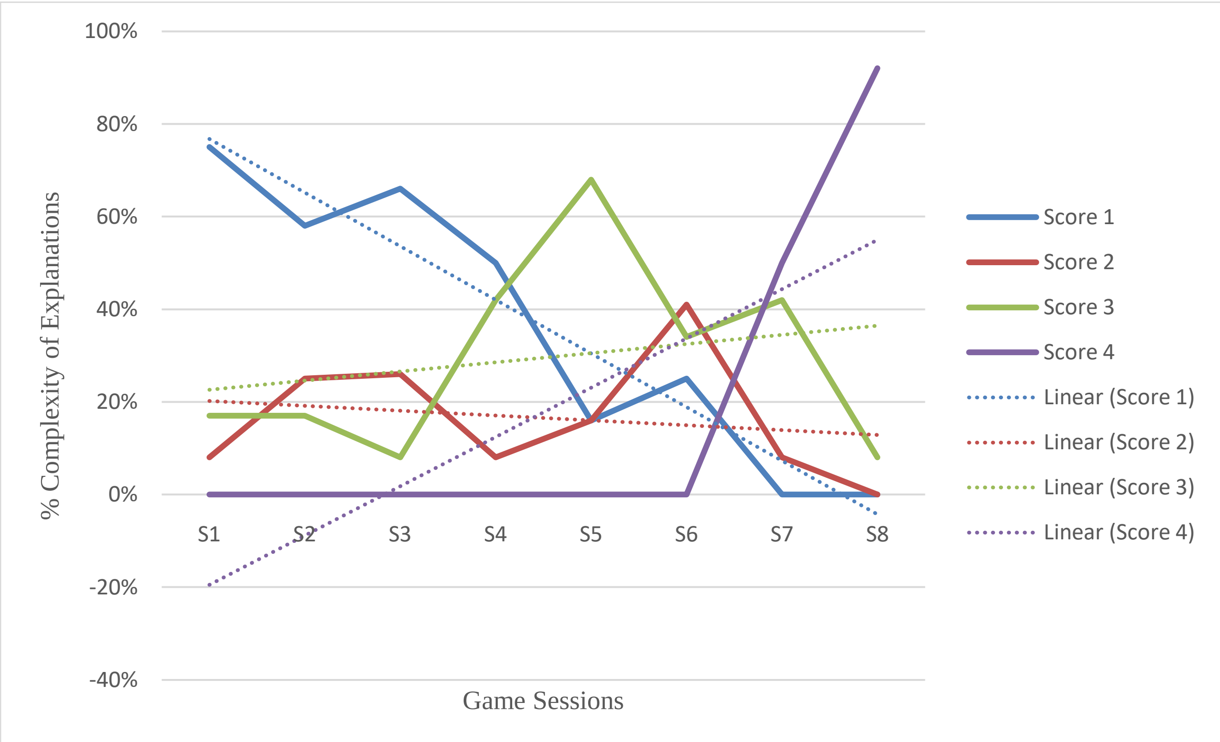


Figure 3. Trajectories of the complexity of explanations: Cluster 2.

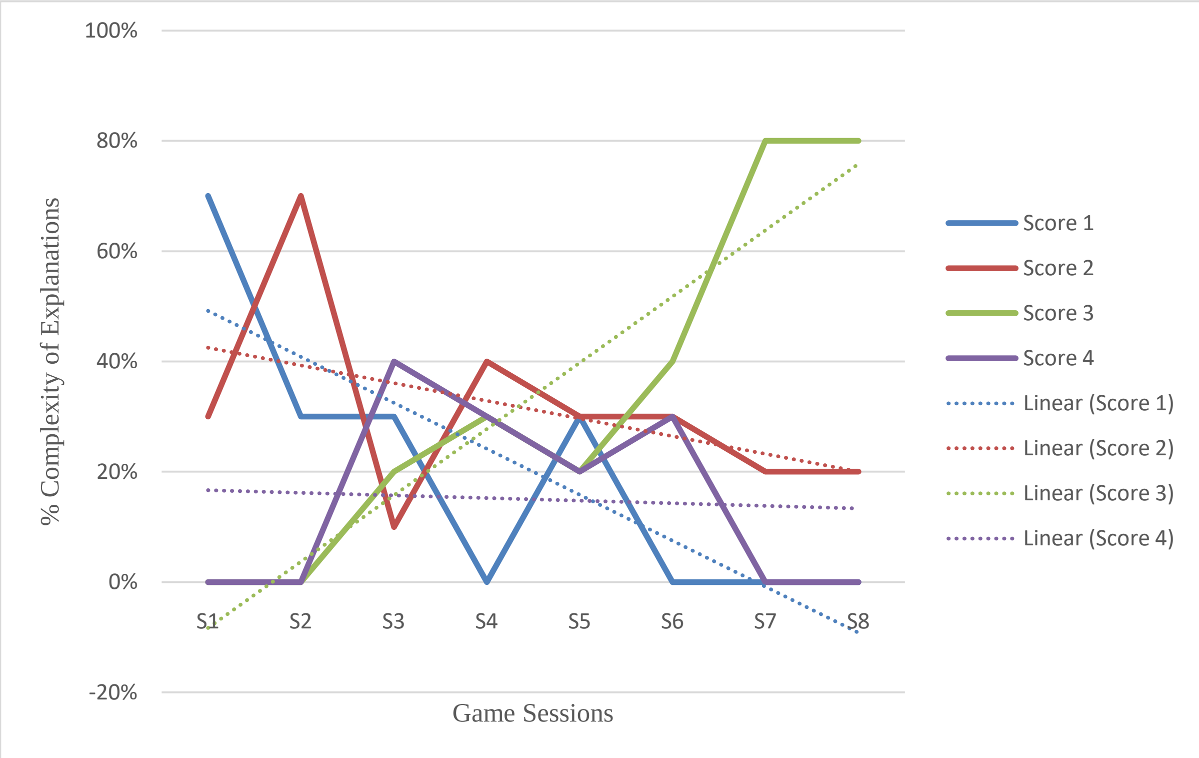


Figure 4. Trajectories of the complexity of explanations: Cluster 3.

Figure 5 shows trajectories of explanations by complexity level for the nine participants in Cluster 4. This type of graph shows the temporal changes in the complexity of explanations across the eight-game sessions. As a result, the trajectories of explanations exhibit an M-shaped pattern. In the initial sessions, this pattern of variability shows a high peak of explanations (44%) characterized by descriptions (score 1) of physical mechanisms, which decreases and is low in session 8 (23%). Across sessions, explanations focused on descriptions (score 1) decreased, while explanations focused on simple representations (score 2) and on relationships between simple representations (score 3) increased.

In summary, for Cluster 4, the type of explanations provided by participants indicates variability from understanding the physical mechanisms of the task, primarily focused on describing physical concepts, to understanding focused on simple representations (score 2) and relationships between representations (score 3) that combine several physical mechanisms involved in the problematic situation. As a result, participants in this group did not provide explanations focused on abstract representations (score 4).

Figure 6 shows the trajectories of explanations by complexity level for the five participants in Cluster 5. This graph shows the temporal changes in the complexity of explanations across the eight-game sessions. As a result, the trajectories of explanations show a zigzag pattern. In the initial sessions, this pattern of variability showed a high level of explanations (42%) that accounted for abstract representations (score 4) about physical mechanisms, with peaks in sessions 3 (57%), 7 (57%), and 8 (87%). In summary, for Cluster 5, the type of explanations provided by participants indicates significant variability with transitions from understanding the physical mechanisms of the task, primarily focused on describing and offering simple representations of the physical mechanisms, to focused understanding of relationships that involve simple and abstract representations. In other words, participants in this group shifted their focus toward establishing relationships between mechanisms and the justifications for the physical mechanisms underlying the functioning of the problematic situation. This group exhibited the most variable changes in the complexity of explanations across the eight sessions.

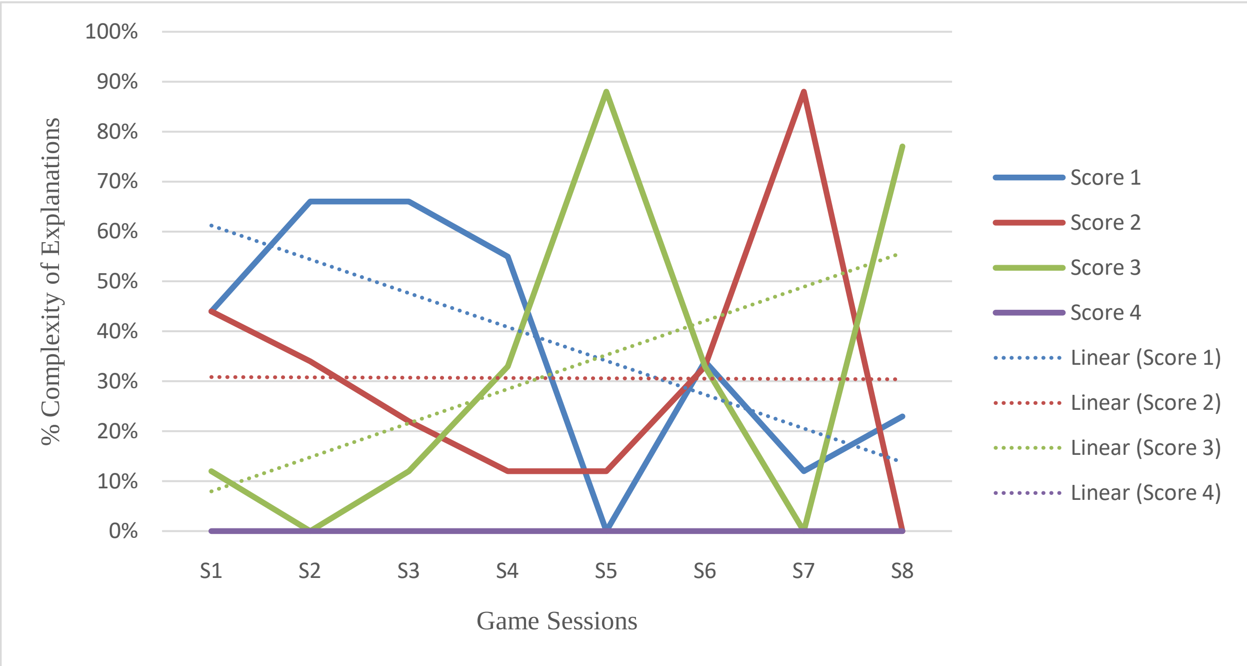


Figure 5. Trajectories of the complexity of explanations: Cluster 4.

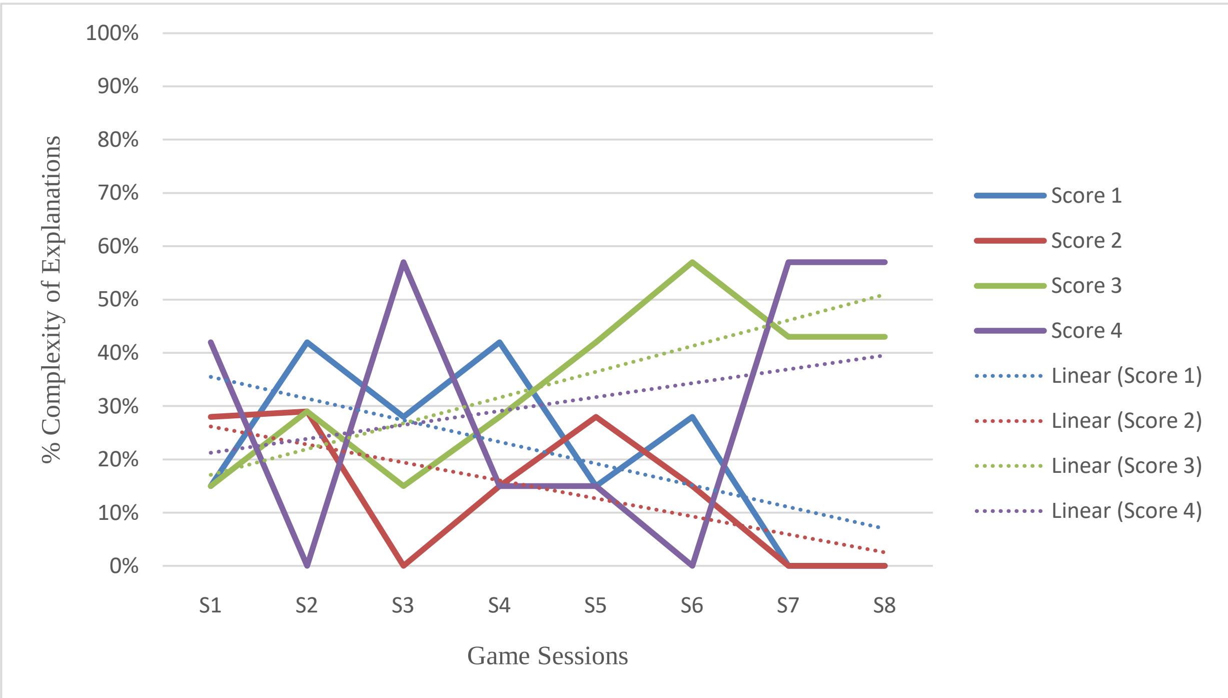


Figure 6. Trajectories of the complexity of explanations: Cluster 5.

To detail the explanations given by a participant, three of the explanations from participant P30 during sessions 1, 5 and 8 are shown below; explanations are presented verbatim in quotation marks as written by participant P30:

- Session 1, Score 1: “Because I calculated the spaces and distances well, as well as the force, so that each of the mechanisms was activated, and thus the bell rang.”
- Session 5, Score 3: “The plane inclined by the slope and gravity drops the block and activates the bell, the spring with the accumulated force activates the pulley; and the weight of the lever box shifts the lever arm in its favor, so that the end was longer and could activate the bell. In that space, the lever cannot be used to balance [anything].”
- Session 8, Score 1: “The catapult is not in the path of the rope ball.”

In the specific case of these three explanations, a peak is evident; it starts low, increases, and ends with a low score (cluster 1 participant). The above shows that the participants' processes can account for variability and transitions throughout the sessions. Linear and progressive trajectories are not necessarily followed. The clusters describe participants' processes that can show different possible trajectories of explanations, which vary, oscillate, and show advances and setbacks throughout the sessions. Explanations are participants' abstractions about the empirical evidence of their experimentation. The examples of explanations demonstrate diverse levels of understanding. They are also sufficient to reflect the complexity of the participants' understanding. Therefore, the participants' explanations demonstrate diverse levels of understanding and reveal the nonlinear dynamics of cognition. In future research, it would also be interesting to examine the metacognitive processes evident in participants' written explanations in the virtual game.

In the previous example, if an isolated session (1 or 3) was considered, the interpretation of the participant's performance would differ. Also, it is expected that a high score (session 5) should not be followed by a low score (session 8); that is, there is a tendency to assume that the complexity of participants' explanations should increase progressively as the sessions pass. The above highlights the importance of the extended, detailed analysis of the microgenetic method, since the participants' explanations reveal variability and different dynamic trajectories, as evidenced by various patterns identified across clusters. These results call attention to professionals and researchers to consider interventions and evaluations with multiple measures, detailed over time, to better understand what happens and how changes in the development of communication skills arise in students.

5. Discussion

This study aimed to characterize the explanations of 53 10th-grade high school students as they solved a problem involving the use and understanding of physical mechanisms. The results highlighted two key empirical findings: (1) students identified relevant physical concepts when explaining how mechanisms operated, and (2) five distinct patterns of variability in explanatory complexity were observed across sessions.

Regarding the identification of physical concepts, students’ explanations included references to force, torque, movement, and acceleration, among others. This identification was observed more frequently in Sessions 4, 5, and 6, where explanations were categorized as simple representations (clusters 4 and 5), following earlier sessions characterized primarily by descriptive accounts of physical properties (e.g., "it is tall", "it is big"). Across sessions, the mean complexity scores increased overall, and more explanations incorporated abstract representations and causal coordination among components (clusters 1 and 2). These descriptive results indicate a shift in the structural organization of explanations over time.

In relation to previous research, [34] reported that children aged 7–10 constructed causal explanations including physical concepts such as mass and acceleration when solving problem situations. The present findings are consistent with the idea that declarative components of scientific thinking emerge early. However, the observed increase in explanatory complexity across sessions may suggest continued development in how older students integrate formal scientific concepts in dynamic contexts. This interpretation should be understood as theoretically informed rather than as direct evidence of developmental mechanisms. These findings also align with arguments emphasizing the relevance of sustained STEM education sensitive to short-term processes of change [3,55].

Being an open, problematic situation without an obvious answer, the Scientific Machine required students to articulate explanations as part of the problem-solving process. The results show that participants produced explanations framed in terms of physics concepts. While prior literature suggests that problem-solving techniques extend across STEM domains [15,19], the present study does not directly test transfer across domains; therefore, this connection should be interpreted with caution.

The results also show fluctuations in the trajectories of explanatory complexity, including temporary increases and decreases across sessions. This pattern is consistent with constructivist perspectives, suggesting that improvements in complexity may occur over short periods and in nonlinear ways [20]. From moment to moment, students displayed variability in how they approached the problem. This variability could be interpreted as evidence of active reorganization processes; however, the study design does not allow direct observation of the underlying cognitive mechanisms.

Beyond describing nonlinear trajectories, it is important to consider possible mechanisms that may explain temporary decreases in explanatory complexity. The nonlinear trajectories observed in the explanations do not necessarily represent simple regression in understanding. Temporary decreases in explanatory complexity may emerge from several mechanisms related to the dynamics of problem solving. For example, students may sometimes produce shorter explanations as a strategic response while focusing their cognitive resources on manipulating the mechanisms in the virtual environment. In other moments, attention may shift from verbal explanation toward the construction and testing of mechanical configurations required to achieve the task goal. Additionally, coordinating multiple physical mechanisms in the task may increase cognitive load, which can lead students to temporarily simplify their verbal explanations. From a microgenetic perspective, such fluctuations are expected in learning processes and may reflect the competition and coordination of multiple strategies, rather than a loss of knowledge.

Two task-related aspects may help explain these patterns. First, the guiding questions posed during the sessions served as scaffolds, prompting participants to elaborate on their explanations. Second, the multivariate nature of the Scientific Machine required coordination of multiple physical relations. These features may have supported explanatory elaboration, although this remains an interpretative inference rather than a directly tested causal claim.

Regarding variability in explanations, the results showed non-homogeneous trajectories across sessions, particularly in cluster 5, which displayed zigzag patterns. Empirically, this indicates alternating increases and decreases in complexity levels. This pattern could be interpreted as dynamic back-and-forth reorganization, in which more complex explanations are sometimes replaced by less complex ones. Such fluctuations may reflect integration of new information or shifts in resolution strategies [31], although these mechanisms were not directly measured.

The different trajectories observed across the five clusters throughout eight sessions were captured through the microgenetic method. This design enabled repeated measurement of explanatory complexity and the identification of inter- and intraindividual variability patterns. A single measurement would not have revealed these fluctuations, but rather a static snapshot of performance [38,39]. Similar patterns of variability in scientific reasoning have been described in previous studies [42–44] and in recent research conducted in virtual STEM environments. For example, researchers reported progressive increases in the complexity of students' scientific reasoning across repeated sessions, although the present findings extend this work by revealing more heterogeneous and fluctuating explanatory trajectories, particularly by identifying cluster-based patterns that capture both stability and variability in students' explanations [59].

Authors have emphasized the importance of providing practical rehearsal opportunities to support learning processes [3]. The present results descriptively show that students' explanations followed diverse trajectories over time. This suggests that cross-sectional assessments might overlook dynamic processes observable through longitudinal sampling [9,56]. However, this implication should be viewed as a methodological consideration derived from the observed patterns rather than as a tested comparison.

The identification of physical concepts in a problem situation was also reported in a 22-session microgenetic study [11], in which participants displayed different coordination patterns of conceptual resources. The zigzag-shaped trajectories observed in the present study resemble variability patterns described in that research. Both studies document nonlinear changes across sessions. Nevertheless, cluster patterns are domain- and task-sensitive, and development is influenced by contextual factors, including task characteristics and interaction formats [42,45,57]. Therefore, the present findings should be interpreted within the specific conditions of this study.

The findings indicate that explanatory complexity unfolded through nonlinear trajectories characterized by intraindividual fluctuations. This pattern may have implications for educational assessment and instructional design. For example, temporary decreases in complexity could be interpreted as transitional restructuring processes rather than necessarily as deficits. However, this interpretation is theoretical and was not directly evaluated through independent measures of cognitive restructuring.

It is also important to distinguish between the complexity of explanations and success in solving the task. While more complex explanations may indicate a higher level of conceptual coordination, successful completion of the task may depend on additional factors, such as experimentation with mechanisms, trial-and-error strategies, or procedural exploration within the virtual environment. Therefore, explanatory complexity should be interpreted as an indicator of the organization of conceptual reasoning rather than a direct measure of task performance. Future studies could examine the relationship between explanatory complexity and performance indicators such as time on task, number of attempts, or successful system activation, which may provide a more comprehensive understanding of how conceptual reasoning and problem-solving performance interact in STEM learning contexts.

Regarding pedagogical scaffolding, the variability observed across clusters may suggest the importance of aligning instructional support with the structural level of students' explanations. When descriptive explanations predominate, teachers might encourage coordination of causal relations; when relational coordination is evident, scaffolding may shift toward abstraction and generalization. These implications should be considered as possible applications derived from the findings rather than experimentally tested outcomes.

Given the microgenetic nature of the design, repeated exposure to the same general task format could potentially generate practice effects. However, several elements qualify this possibility: (1) configurations of mechanisms varied between sessions, (2) functional relationships differed, and (3) coordination demands changed. Moreover, the empirical trajectories did not show homogeneous linear progression but exhibited differentiated patterns, including regression and fluctuations. This variability is not fully consistent with a simple cumulative practice effect, although the possibility cannot be entirely ruled out.

Another limitation of the present study is that the explanations analyzed were written productions. Consequently, individual differences in writing skills could have influenced the assigned complexity scores. Although the coding system was designed to identify levels of conceptual complexity based on structural and relational criteria of the content, the participants' ability to express their reasoning in writing may have influenced the explicitness of these relationships. Therefore, some of the variability observed in the trajectories could reflect not only differences in conceptual complexity but also variations in writing skills.

A challenge for future research is to identify patterns of scientific explanations and concepts from the Scientific Machine (or similar problem-solving tasks) across different grades and social contexts while retaining repeated measurements. It may also be valuable to examine how performance in open-ended STEM problems articulates with formal physics instruction. Additionally, exploring tangible-material tasks and collaborative learning contexts could provide further insight into the dynamics of explanatory development.

6. Conclusions and implications

Considering student performance as a dynamic process requires continuous monitoring of students' understanding rather than treating performance as definitive states of "known" or "not known". This perspective is relevant to student evaluation and learning assessment, as it aligns with evidence that cognitive performance is fluctuating and context-dependent. Systematic observation of classroom performance at different time points may allow for a more precise characterization of students' problem-solving processes and cognitive strategies, which can vary even over short intervals. Accordingly, structured, repeated observations using diverse formats can support the analysis of the dynamic evolution of students' cognitive abilities. One possible direction for future research is to examine challenging, contextualized situations that facilitate interaction between experts and novices, where scaffolding and feedback can be provided during and after the learning process. Such instructional supports were not implemented in the present study but could be explored to determine their influence on trajectories of explanatory complexity. Project-based learning situations in science, incorporating both formative and summative assessments at multiple time points, may offer opportunities to document changes in students' understanding over time.

As a problem-solving approach, incorporating the Scientific Machine into STEM classrooms may provide opportunities to develop explanations and apply physical concepts, given its open-ended structure, which allows students to approach situations involving physical mechanisms in varied ways. In addition, such tasks may create conditions for the development of transversal processes, including perseverance, within complex learning contexts.

The variability in 10th-grade students' explanations using physical concepts began with descriptions of physical concepts, followed by the establishment of relationships between them and causal explanations about the functioning of mechanisms. As the sessions progressed, students offered more complex explanations addressing relationships between simple and abstract representations of physical mechanisms. This pattern suggests a reorganization of understanding across the problem-solving sessions. These findings indicate that students' science learning processes may involve moment-to-moment changes in the structure and complexity of their explanations, as reflected in the different trajectories identified in the cluster analysis. Science learning contexts thus present multiple opportunities to construct and reorganize explanations, particularly when students engage with challenging questions and problem-solving situations that involve different cognitive demands.

Students explained physical mechanisms involving force, force ratios, distance, and accelerated motion, all of which are present in the Scientific Machine. Furthermore, the 10th-grade students' explanatory paths when solving the task were characterized as dynamic, reflecting a non-sequential and nonlinear process (variability patterns), as observed in the identified clusters. The study design was non-experimental and cross-sectional, exploratory in nature and descriptive in scope, and used a microgenetic method, as participants' explanations were derived from their interactions with the game across repeated sessions. In particular, the microgenetic analysis characterized different trajectories of complexity in explanations across sessions of solving the Scientific Machine, providing a process-oriented approach that enabled close analysis of individual cases and comparisons to identify interindividual patterns. Within this framework, the study contributes empirical evidence regarding the use of the Scientific Machine as a context for examining the development and variability of scientific explanations. Specifically, the task required students to address complex and multivariate problems related to physical mechanisms, engage with information and communication technologies, manage virtual applications, complete tests, and formulate hypotheses, processes associated with components of scientific literacy. The findings also situate scientific thinking within a STEM education context, given its connection to core physical concepts addressed in the task.

Several limitations related to sampling and generalizability must be acknowledged. The participants came from private educational institutions with regular access to technological resources. This context may be associated with particular educational conditions, such as greater exposure to digital tools and structured academic support. Likewise, specific socio-educational characteristics may have influenced both familiarity with virtual environments and the development of scientific explanations.

Access to technology is particularly relevant in this study, as the problem-solving situation required interaction with a digital interface. Students with sustained prior exposure to similar tools may exhibit different engagement patterns or levels of fluency compared to students from contexts with limited technological access. Consequently, the trajectories of explanatory complexity observed here may not fully represent those that would emerge in schools with different infrastructural or socioeconomic conditions. Future research should replicate this design in public institutions and in contexts with different levels of technological access to examine the robustness and transferability of the trajectory patterns identified in diverse educational environments.

In conclusion, the results underscore the importance of examining change processes in explanations and problem-solving over time, particularly when contrasted with single-time point assessments. Traditional cross-sectional approaches may not fully capture students' dynamism, variability, and transitions across levels of explanatory complexity. A process-oriented perspective may therefore provide a complementary lens for understanding the development of scientific thinking in educational contexts, particularly when students engage in repeated problem-solving activities that allow the observation of trajectories in explanatory complexity.

Author contributions

Kenji López: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing; Marlenny Guevara: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing; Valeria Cabello: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing; Jairo Montes: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Use of Generative AI tools declaration

No artificial intelligence was used.

Acknowledgements

We would like to thank the Faculty of Education of the Pontificia Universidad Católica de Chile for their support in the translation of this article.

This work was supported by ANID – MILENIO – NC2021_014 and FONDECYT project 1221716, directed by Professor Valeria Cabello González, Pontificia Universidad Católica de Chile.

Conflict of interest

The authors declare that they do not have any conflicts of interest.

Ethics declaration

The study was approved by the ethics committee, as stated in the approval certificate (Ethics Committee of the University of Valle: protocol 108-021).

References

1. Ben-Soussan, T., Croce, R., Horvat, M., McCullick, B., Pesce, C., Tomporowski, P.D., et al., Variability of practice as an interface between motor and cognitive development. *International*

Journal of Sport and Exercise Psychology, 2019, 17(2): 133–152. https://doi.org/10.1080/1612197X.2016.1223421

2. Vieira, R.M. and Tenreiro-Vieira, C., Fostering scientific literacy and critical thinking in elementary science education. *International Journal of science and mathematics education*, 2016, 14(4): 659–680. https://doi.org/10.1007/s10763-014-9605-2
3. Yıldız Taşdemir, C. and Güler Yıldız, T., Science learning needs of preschool children and science activities carried out by teachers. *Journal of Turkish Science Education*, 2024, 21(1): 82–101. https://doi.org/10.36681/tused.2024.005
4. Rushton, E.A. and Reiss, M.J., From science teacher to 'teacher scientist': exploring the experiences of research-active science teachers in the UK. *International Journal of Science Education*, 2019, 41(11): 1541–1561. https://doi.org/10.1080/09500693.2019.1615656
5. Díaz-Levicoy, D.A., Ferrada, C., Puraivan, E. and Salgado Orellana, N., Análisis bibliométrico sobre educación STEM. *Revista espacios*, 2019, 40(08).
6. García, J.G., El constructivismo en la educación y el aporte de la teoría sociocultural de Vygotsky para comprender la construcción del conocimiento en el ser humano. *Dilemas Contemporáneos: Educación, Política y Valores*, 2020.
7. Hadzigeorgiou, Y. and Schulz, R.M., Engaging Students in Science: The Potential Role of "Narrative Thinking" and "Romantic Understanding." *Frontiers in Education*, 2019, 4: 454767. https://doi.org/10.3389/feduc.2019.00038
8. Lager, A. and Lavonen, J., Engaging Students in Scientific Practices in a Remote Setting. *Educ. Sci.*, 2023, 13(5): 431. https://doi.org/10.3390/educsci13050431
9. Sarı, U., Duygu, E., Şen Ömer, F. and Kırındı, T., The Effects of STEM education on scientific process skills and STEM awareness in simulation based ınquiry learning environment. *Journal of Turkish Science Education*, 2020, 17(3): 387–405. https://doi.org/10.36681/tused.2020.34
10. Shirazi, S., Student experience of school science. *International Journal of Science Education*, 2017, 39(14): 1891–1912. https://doi.org/10.1080/09500693.2017.1356943
11. Brock, R. and Taber, K.S., Making Sense of 'Making Sense' in Science Education: A Microgenetic Multiple Case Study. In *Cognitive and Affective Aspects in Science Education Research: Selected Papers from the ESERA 2015 Conference*, 2017, 157–166. Springer International Publishing. https://doi.org/10.1007/978-3-319-58685-4_12
12. Abello, R., Castaño, J. and Escobar, H., Trayectorias de control y covariación de variables como expresión del cambio cognitivo en la solución de un problema. *Universitas Psychologica*, 2016, 15(1): 281–302. https://doi.org/10.11144/Javeriana.upsy15-1.tccv
13. Aróstegui, J.L., A New Education for a New Era: Transdisciplinarity, Transversal Competences and an Eclectic Approach to Evaluation. En: Ortega-Sánchez, D. (ed). *Controversial Issues and Social Problems for an Integrated Disciplinary Teaching*, 2022, 53–64. https://doi.org/10.1007/978-3-031-08697-7_5
14. Mora, S.Z.J., La transversalidad curricular: algunas consideraciones teóricas para su implementación. *Revista boletín redipe*, 2018, 7(11): 65–81.
15. Nikolay, T., The transversal competence for problem-solving in cognitive learning. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, 2018, 6(3): 67–82. https://doi.org/10.5937/ijcrsee1803067T
16. Resing, W.C., Bakker, M., Pronk, C.M. and Elliott, J.G., Progression paths in children's problem solving: The influence of dynamic testing, initial variability, and working memory. *Journal of Experimental Child Psychology*, 2017, 153: 83–109. https://doi.org/10.1016/j.jecp.2016.09.004

17. Al Muhdhar, M., Fitriani, A., Susilo, H. and Zubaidah, S., The Effects of Integrated Problem-Based Learning, Predict, Observe, Explain on Problem-Solving Skills and Self-Efficacy. *Eurasian Journal of Educational Research*, 2020, 20(85): 45–64. https://doi.org/10.14689/ejer.2020.85.3
18. Guevara, M., Rojas Ospina, T. and van Geert, P., Understanding transfer from a dynamic system approach: Two studies of children using problem-solving tasks. *Cogent Education*, 2020, 7(1). https://doi.org/10.1080/2331186X.2020.1814512
19. Evans, T., Klymchuk, S., Murphy, P.E., Novak, J., Stephens, J. and Thomas, M., Non-routine mathematical problem-solving: creativity, engagement, and intuition of STEM tertiary students. *STEM Education*, 2021, 1(4): 256–278. https://doi.org/10.3934/steme.2021017
20. Zuljan, D., Valenčič Z.M. and Pejić P.P., Cognitive constructivist way of teaching scientific and technical contents. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, 2021, 9(1): 23–36. https://doi.org/10.23947/2334-8496-2021-9-1-23-36
21. Xu, F., Towards a rational constructivist theory of cognitive development. *Psychological Review*, 2019, 126(6): 841–864. https://doi.org/10.1037/rev0000153
22. Agarkar, S.C., Influence of learning theories on science education. *Resonance*, 2019, 24(8): 847–859. https://doi.org/10.1007/s12045-019-0848-7
23. Geelan, D.R., Learning to communicate: Developing as a science teacher. *Australian Science Teachers Journal*, 1996, 42(1).
24. McLure, F., The Thinking Frames Approach: Improving High School Students' Written Explanations of Phenomena in Science. *Res Sci Educ*, 2023, 53: 173–191. https://doi.org/10.1007/s11165-022-10052-y
25. Fischer, K.W., Bullock, D.H., Rotenberg, E.J. and Raya, P., The dynamics of competence: How context contributes directly to skill. In *Development in context*, 2014, 93–117. Psychology Press.
26. Hatch, J.B., Jensen, J.L., Neeley, S. and Piorczynski, T., Learning scientific reasoning skills may be key to retention in science, technology, engineering, and mathematics. *Journal of College Student Retention: Research, Theory & Practice*, 2017, 19(2): 126–144. https://doi.org/10.1177/1521025115611616
27. Ratheeswari, K., Information communication technology in education. *Journal of Applied and Advanced research*, 2018, 3(1): 45–47. https://doi.org/10.21839/jaar.2018.v3iS1.169
28. Hasanah, D., Samsudi, S., Khafid, M. and Mahmud, A., Vocational High School Teachers Continuous Professional Learning. *International Conference on Science, Education, and Technology*, 2023, 9(1): 176–183.
29. Klahr, D. and Zimmerman, C., Development of Scientific Thinking. *Stevens' Handbook of Experimental Psychology*, 2018. https://doi.org/10.1002/9781119170174.epcn407
30. Land, S.M. and Zimmerman, H.T., Supporting children's place-based observations and explanations using collaboration scripts while learning-on-the-move outdoors. *Intern. J. Comput-Support. Collab. Learn*, 2022, 17: 107–134. https://doi.org/10.1007/s11412-022-09366-w
31. Lewis, M.D., The promise of dynamic systems approaches for an integrated account of human development. *Child Development*, 2000, 71: 36–43. https://doi.org/10.1111/1467-8624.00116

32. van der Aalsvoort, G.M., van Geert, P. and Steenbeek, H.W., Microgenetic methodology: Possibilities with regard to research on learning and instruction. *Investigating Classroom Interaction*, 2009, 203–227. Brill. https://doi.org/10.1163/9789087907624_011
33. Frøyland, M. and Remmen, K., Students' use of observation in geology: towards "scientific observation" in rock classification. *International Journal of Science Education*, 2019, 42(1): 113–132. https://doi.org/10.1080/09500693.2019.1704914
34. De Carvalho, A.M.P. and Paulo, S., Building up Explanations in Physics Teaching. *International Journal of Science Education*, 2004, 26(2): 225–237. https://doi.org/10.1080/0950069032000052072
35. Siegler, R.S., Microgenetic Analyses of Learning. In D. Kuhn, R. S. Siegler, W. Damon, & R.M. Lerner (Eds.), *Handbook of child psychology: Cognition, perception, and language*, 2006, 464–510. John Wiley & Sons Inc.
36. Filippi, R., Ceccolini, A., Periche-Tomas, E. and Bright, P., Developmental trajectories of metacognitive processing and executive function from childhood to older age. *Quarterly Journal of Experimental Psychology*, 2020, 73(11): 1757–1773. https://doi.org/10.1177/1747021820931096
37. Márquez, C., Roca, M., Sardà, A., Pujol, R. and Gómez, A., La construcción de modelos explicativos complejos mediante preguntas mediadoras. *Investigación en la Escuela*, 2004, 53: 71–81.
38. Jufrida, J., Basuki, F.R., Kurniawan, W., Pangestu, M.D. and Fitaloka, O., Scientific literacy and science learning achievement at junior high school. *International Journal of Evaluation and Research in Education*, 2019, 8(4): 630–636. https://doi.org/10.11591/ijere.v8i4.20312
39. Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Klager, C., Bradford, L., et al., Improving science achievement—Is it possible? Evaluating the efficacy of a high school chemistry and physics project-based learning intervention. *Educational Researcher*, 2022, 51(2): 109–121. https://doi.org/10.3102/0013189X211067742
40. Copur-Gencturk, Y. and Atabas, S., A microgenetic analysis of teachers' learning through teaching. *IJ STEM Ed*, 2024, 11: 29. https://doi.org/10.1186/s40594-024-00488-1
41. Fossa, P., Cortés, C., Molina, M.E., Barros, M., Marcotti, C.M., Sprovera, I., et al., Microgenetic Analysis of Thought Trajectories: A Mixed Design. *Integr. psych. behav*., 2022, 56: 630–652. https://doi.org/10.1007/s12124-021-09633-9
42. Guevara, M., *Peer Interaction and Scientific Reasoning Processes in Preschoolers: An intra-individual approach* [Doctoral dissertation, University of Groningen], 2015, Groningen, The Netherlands.
43. van Geert, P.L.C., Dynamic systems approaches and modeling of developmental processes. In J. Valsiner, & K. Connoly (Eds.), *Handbook of developmental psychology*, 2003, 640–672. SAGE Publications Inc. https://doi.org/10.4135/9781848608306.n28
44. Montes, J., *Variabilidad y experimentación: Una aproximación desde los sistemas dinámicos no lineales al razonamiento científico* [Doctoral dissertation, Universidad del Valle], 2013, Cali - Colombia.
45. van der Steen, S., Steenbeek, H., Van Dijk, M. and Van Geert, P., A process approach to children's understanding of scientific concepts: A longitudinal case study. *Learning and Individual Differences*, 2014, 30: 84–91. https://doi.org/10.1016/j.lindif.2013.12.004

46. Chen, L., A Brief Review of Microgenetic Method in Second Language Acquisition Research. In *7th International Conference on Social Science and Higher Education (ICSSHE 2021)*, 2021, 100–106. Atlantis Press. https://doi.org/10.2991/assehr.k.211122.065
47. Golzar, J., Noor, S. and Tajik, O., Convenience Sampling. *International Journal of Education & Language Studies*, 2022, 1(2): 72–77.
48. Fischer, K.W. and Bidell, T.R., Dynamic development of action and thought. *Handbook of child psychology*, 2006, 1: 313–399. https://doi.org/10.1002/9780470147658.chpsy0107
49. Abraira, V., El índice kappa. *SEMERGEN - Medicina de Familia*, 2001, 27(5): 247–249. https://doi.org/10.1016/S1138-3593(01)73955-X
50. Fortus, D. and Touitou, I., Changes to students' motivation to learn science. *Disciplinary and Interdisciplinary Science Education Research*, 2021, 3(1): 1–14. https://doi.org/10.1186/s43031-020-00029-0
51. Johnson, T.M., Byrd, K.O. and Allison, E.R., The impact of integrated STEM modeling on elementary preservice teachers' self-efficacy for integrated STEM instruction: A co-teaching approach. *School Science and Mathematics*, 2021, 121(1): 25–35. https://doi.org/10.1111/ssm.12443
52. Aguilera, D., Lupiáñez, J., Perales, J. and Vílchez, J., Objetivos de la educación STEM. Revisión sistemática. *Actas electrónicas del XI Congreso internacional en investigación en didáctica de las ciencias*, 2021, 1939–1942.
53. Potvin, P.J. and Schutz, R.W., Statistical power for the two-factor repeated measures ANOVA. *Behavior Research Methods, Instruments*, & Computers, 2000, 32: 347–356. https://doi.org/10.3758/BF03207805
54. Siburian, L., Data Mining Memprediksi Kebutuhan Vaksin Imunisasi dengan Menggunakan Metode Naive Bayes (Studi kasus UPT Puskesmas Teladan). *Resolusi: Rekayasa Teknik Informatika dan Informasi*, 20221, 1(5): 282–290.
55. Grimalt-Álvaro, C. and Couso, D., ¿Qué sabemos del posicionamiento STEM del alumnado? Una revision sistemática de la literatura. *Revista de Investigación Educativa*, 2022, 40(2): 531–547. http://dx.doi.org/10.6018/rie.467901
56. Hellgren, J.M. and Lindberg, S., Motivating students with authentic science experiences: changes in motivation for school science. *Research in Science & Technological Education*, 2017, 35(4): 409–426. http://dx.doi.org/10.1080/02635143.2017.1322572
57. Hurtado, J.P.A., Zapata, J.A.V. and Jaimes, M.E.B., Variabilidad y cambio en la comprensión de deseos en niños preescolares: un estudio microgenético. *Universitas Psychologica*, 2021, 20: 1–14. https://doi.org/10.11144/Javeriana.upsy20.vccd
58. Czyż, S.H., Variability of practice, information processing, and decision making—how much do we know? *Frontiers in Psychology*, 2021, 12: 639131. https://doi.org/10.3389/fpsyg.2021.639131
59. López Takegami, J.J.K., Cabello González, V., Guevara Guerrero, M. and Montes González, J.A., Exploring Students' Scientific Reasoning During Virtual Game in STEAM Problem Solving. LUMAT: *International Journal on Math, Science and Technology Education*, 2025, 13(3): 1. https://doi.org/10.31129/LUMAT.13.3.2795

Author's biography

Dr. José Julián Kenji López Takegami: Psychologist. Specialist in Educational Psychology. Master's in Education. Doctor of Psychology. Doctor of Education. Professional Profile: Professor and Researcher, Santiago de Cali University.

Dr. Marlenny Guevara Guerrero: Psychologist. Master's in Psychology. Ph.D. in Developmental Psychology, University of Groningen, Netherlands. Master's in Psychology, University of Valle, Cali, Colombia. Psychologist, University of Valle, Cali, Colombia. Professional Profile: Associate Professor, Faculty of Psychology, Universidad del Valle, Cali, Colombia (2016-Present). Member of the Institute for Research in Development Sciences, Learning and Subjectivities (CIDEAS), Universidad del Valle, Cali, Colombia.

Dr. Valeria Cabello González: Psychologist. Master's in Psychology with a specialization in Educational Psychology from the Pontifical Catholic University of Chile. Doctor of Philosophy in Educational Psychology, University of Dundee, Scotland. Professional Profile: Research Associate at the Center for the Study of Policies and Practices in Education (CEPPE UC). Associate Professor, School of Education, Pontifical Catholic University of Chile.

Dr. Jairo Andrés Montes: Psychologist. Master's in Psychology. PhD in Developmental Psychology and Master's in Educational Psychology from Universidad del Valle, and Psychologist from Pontificia Universidad Javeriana, Cali. Professional Profile: He has worked at educational institutions as a professor, consultant, and researcher in areas related to the study of cognitive development, specifically in the development of scientific thinking and higher-order thinking skills.