LUMAT
International Journal on Math, Science and Technology Education

Research article

Exploring Students' Scientific Reasoning During Virtual Game in STEAM Problem Solving

Kenji López Takegami[1,2], Valeria Cabello González[2], Marlenny Guevara Guerrero[1] and Jairo Montes González[3]

[1] Universidad del Valle, Colombia
[2] Pontificia Universidad Católica de Chile, Chile
[3] Gaming and Learning, Colombia

Abstract: The purpose of this study was to explore the complexity of scientific reasoning elicited during a virtual STEAM problem-solving activity by challenging questions posed to 53 tenth-grade high school students. Using a microgenetic approach, this study analyzed eight measurements from each student as they worked on a STEAM problem-solving game software, utilizing it as an interactive simulation. To measure the complexity of their scientific reasoning, researchers used a coding rubric based on the cognitive demand of questions that were self-administered through the interactive simulation. Multivariate analysis showed that the students' level of scientific reasoning complexity increased over time, shifting from descriptive reasoning in the early sessions to more complex causal reasoning in the final sessions. It is possible to interpret that this sophistication in scientific reasoning may be related to three points: 1) The demand for practical scientific knowledge regarding the virtual interactive simulation when solving the problem, 2) The challenging questions presented in the situation, and 3) The immersive nature not only of the simulation itself, but also of the repeated measures design of this science education activity. We discuss about the type of reasoning observed, progressively becoming more complex, and the transitions between lower to higher scores as a means of an ongoing process of learning opportunities for developing students' scientific reasoning in a virtual interactive simulation. Interaction with the designed game allowed students' scientific reasoning to increase progressively over the eight observation sessions.

Keywords: Multivariate analysis, Virtual interactive simulation, STEAM education, Scientific reasoning, Problem solving.

Correspondence: jose.julian.kenji@gmail.com

1 Introduction

Scientific reasoning plays a core role in the educational process, as it refers to a skill set of reasoning (e.g., inference, hypothesis, experimentation, explanations, etc.) required by students not only in daily life activities but also to face the challenges and solve problems in modern society. The objective of this study is to explore the complexity of scientific reasoning elicited during a virtual STEAM problem-solving activity by challenging questions posed to 53 tenth-grade high school students, involving physical mechanisms

Received 31 May 2025 / Accepted 14 November 2025 / Published 20 November 2025
Published by the University of Helsinki, Finland and LUMA Centre Finland.

in a game oriented towards STEAM education (science, technology, engineering, arts, and mathematics). Questions and the search for answers are important for the development of the psyche and in the process of teaching and learning science (Tort, 2005). Artificial intelligence has spurred the development of intelligent teaching and learning systems and interactive educational resources, which aim to develop knowledge in students through interactive processes (Maraza-Quispe et al., 2022; Sánchez et al., 2009). Examples of these systems are virtual problem-solving situations in which stimulating questions are used to foster student learning (Lee & Kinzie, 2011).

The development of scientific reasoning is complex and involves several classroom practices to promote it effectively. One way to promote scientific reasoning is through immersive virtual simulations. These situations allow students to put knowledge into practice, experiment, and apply knowledge in more authentic situations (Wu et al., 2021). Immersion in STEAM educational contexts fosters deeper engagement with scientific knowledge, the sciences, critical thinking, and problem-solving (Kaur et al., 2022; Piboon et al., 2024).

Another way to promote scientific reasoning is through answering the questions asked by the teacher, which have a specific educational intention and are characterized by being challenging, causal, and open-ended (Zeegers and Elliott, 2018). In contrast, closed and memory-based questions do not promote further elaboration by students (Eshach et al., 2013). Direct instructions and rote questions are characteristic of traditional education and continue to be present in learning environments designed to promote critical thinking (Michael & Michael, 2019). Consequently, educational test results show a deficiency in critical and scientific thinking at different educational levels. For example, Colombia's average science score on the PISA (Program for International Student Assessment) tests has shown improvement and stability in recent years. However, from 2006 to 2022, Colombia has consistently been below the Organization for Economic Cooperation and Development (OECD) average in both mathematics, reading, and natural sciences (ICFES, 2024). In this sense, a gap exists in research and intervention, which necessitates proposing situations that foster scientific learning and enhance students' scientific skills. Challenging questions guide students' cognitive progress, helping them to achieve higher levels of knowledge and understanding gradually. This highlights a gap and invites educators to refine their questioning techniques in pedagogical practice and to adopt dynamic assessments that reflect the students' learning process and encourage the development of scientific reasoning. Therefore, this study addresses the research question: What levels of students' scientific reasoning complexity do challenging questions evoke as they solve a virtual simulation in STEAM?

Based on a microgenetic design, an intra-individual analysis was carried out on the level of scientific reasoning complexity shown by participants in eight measurements. Scientific reasoning has been studied using microgenetic designs in previous studies, such as those by Guevara et al. (2016) and Dejonckheere et al. (2009). The virtual game is not only immersive, but also captures students' scientific reasoning (in actions and explanations), allowing for rigorous moment-by-moment analysis within and between sessions. This is innovative in that it will enable for process-based analysis, with a holistic approach that not only takes into account intra-individual factors but also allows for consideration of the inter-individual scientific friction among participants. Additionally, the combination of a virtual video game with challenging questions that require students to demonstrate causal explanations for their actions in the game is motivating. For instance, an analysis of 424 student responses to challenging questions was conducted. These questions were structured to require explanations, reflecting a specific cognitive demand. The level of scientific reasoning complexity demonstrated in the students' explanations within these responses was then coded. This study proposes two hypotheses. First, that students' level of scientific reasoning complexity will improve in the later sessions (higher scores). Second, that growth (change between students over time) in this reasoning will occur, and that this growth will be heterogeneous among students. A multivariate growth model was used for the analysis. The following section presents a theoretical review of the key concepts relevant to this study.

2 Theoretical Framework

This section outlines some of the skills that students need to succeed in the 21st century. Next, virtual and immersive learning experiences in education will be addressed. This section concludes with the role of challenging questions and scaffolding in scientific reasoning. The theoretical approach taken in this study focuses on the STEAM equation and Vygotsky's sociocultural theory. STEAM education promotes scientific skills in a cross-curricular manner from early education (Cabello et al. 2021; Sommer & Cabello, 2020). Likewise, challenging questions posed by researchers serve as scaffolding to enhance students' scientific engagement at more complex levels (Kawalkar & Vijapurkar, 2011).

Today's higher education environment and work context demand that students possess a wider range of competencies compared to previous generations (Martínez and Sánchez, 2018). Nowadays, skills in scientific reasoning (Manassero-Mas and Vázquez-Alonso, 2020), creative thinking, problem-solving skills (Cheng et al., 2017), critical reading (Bråten, 2017), mastery of ICT (Information and Communication Technologies) (Alderete et al., 2017), programming, data management (Gim, 2021), adaptability to change, STEAM skills (Castro, 2022; Dhitasarifa & Wusqo, 2024), and soft skills (Ahmad et al., 2019), among others, are required.

The interest in and need for science instruction at the elementary and secondary education levels was demonstrated by Castañeda Zapata et al. (2024) in a systematic review that included 41 studies published between 2018 and 2022. The challenge for education is not only evident in teaching processes but also in assessing these competencies as students progress through school (Blythe and Harré, 2020). In response to this demand, early childhood education has generated strategies for students to acquire the necessary knowledge and skills to face the challenges they will encounter in their personal, professional, and work lives (Tsankov, 2018). Scientific training allows for systematic and progressive knowledge. Despite the achievements in STEAM education, as mentioned above, international test results continue to show the need to foster science and scientific skills in students. Scientific reasoning is relevant to student learning because it is essentially the systematic and rigorous way in which knowledge is constructed and science is conducted. Scientific communication is a domain-general skill on which the construction of new knowledge is based, hence the importance of fostering it.

Assessment in school contexts has also changed over time. While initially content-centered and memory-based, assessment has shifted toward considering various contents and the learning process itself (Efremova et al., 2020). It has also moved from being centered on a single, isolated measurement element, usually summative and at the end of the formative process, to a continuous exercise that strengthens learning and provides feedback (Lau, 2015). Leveraging assessment as an input for learning is achieved to the extent that assessment is related to a reflective exercise on the learning process, meaning that it provides feedback (Hidri, 2021). The aforementioned understanding of assessment implies conducting assessments at different intervals, not just at the end of the formative process. The educational intentions of the formative process must be planned and meaningful for each curricular stage (Asale, 2017). Initially, an approach to the content may be needed, then learning new concepts to solve specific problems, and finally providing reasoned explanations. In addition to what has been mentioned regarding assessment, it is important to highlight the limitations of student access to virtual simulators. While there has been interest in and improvements to virtual education, students currently show low performance in science according to PISA test results (Barbare, 2004; Berezi, 2025). In this sense, assessment must be well articulated with the learning process; at the beginning, assessment can be more diagnostic, and later it must account for other competencies (McPhail, 2017). Additionally, assessment must also be designed for both real-life and virtual practical situations. That is, using challenging questions and fostering critical thinking skills. Importantly, it is not just about assessing at the beginning and at the end, as learning occurs throughout the training process. That is, learning is not achieved immediately at the end of the formative process, but rather gradually and with different degrees of complexity among students (Alonzo and Loughland, 2022).

International assessment tests measure the competencies as mentioned earlier (Araya-Castillo et al., 2023). In response to the results obtained, each country makes efforts to promote the development of scientific and technological competencies in its students. This is done to improve the quality of education provided to its students (Brown, 2022). One way to foster learning and assess these required student competencies is through problem-solving scenarios and challenging questions. A study by Bargiela et al. (2022) exemplifies this; they investigated the role of teacher-posed questions in developing critical thinking skills and dispositions among 30 students. The results suggest that the teacher’s questions in science class activate students’ skills in explanation, inference, and analysis. This finding highlights that when students perceive the questions as challenging, it positively influences their learning process and is linked to their academic engagement. Reinforcing the importance of engagement, Álvarez-Pérez et al. (2021) found that students with greater academic engagement tend to achieve better grades.

In addition to the previous studies, Sánchez et al. (2022) used microgenetic analysis to show that positive outcomes in learning processes can be understood as being mediated and explained by Vygotsky's sociocultural framework. These authors found that participant learning arises from mediation over time and is part of the social history of psychological development. This epistemological stance is consistent with Bargiela et al. (2022), as their work similarly views student engagement as a means to foster scientific competencies. The three aforementioned studies highlight the role of challenging questions in the learning process and development of scientific reasoning in students. They also show the relevance of microgenetic analysis in the study of the development of human cognition. The above points demonstrate that scientific reasoning, being systematic, can be supported within the sociocultural framework; for example, posing challenging questions and engaging in virtual games can be cultural indicators that promote student learning. For its part, the consideration of microgenetic studies shows the possibility of analyzing scientific friction as a process that is not only intra-individual but also allows for considering the inter-individual.

1.2 Virtual and Immersive Situations in STEAM Education

Immersive technology education, or immersive learning, is a pedagogical approach that uses technology such as virtual reality, augmented reality, and other virtual environments to create engaging and participatory learning experiences (Maldonado et al. 2023; Prince, 2022). This approach seeks to immerse students in simulated or virtual environments so they can explore, interact, and learn in a practical and meaningful way (Beck et al., 2024; Dengel & Mazdefrau, 2018). Engineering and mathematics contribute to STEAM education by providing the cognitive and methodological tools necessary to critically understand, analyze, and transform the world. Specifically, engineering allows us to apply scientific knowledge to design tangible solutions to environmental problems. Mathematics, for its part, serves as the common language of science and technology: enabling us to represent phenomena, model data, and formulate predictions (Belbase et al. 2021). In general, STEAM education is linked to scientific reasoning skills, as these underlie the problem-solving process and support the integration of knowledge domains (Guzey, 2015). In this sense, the STEAM perspective not only strengthens problem-solving, critical thinking, creativity, and innovation, but also the use of scientific reasoning skills (Madelson & Seifert, 2017).

In fact, soft methodologies are an effective approach for designing courses related to practical technologies. Furthermore, the use of immersive technologies improves learning performance in acquiring fundamental knowledge and applied knowledge (Wu et al., 2021). These findings illustrate the importance of technological education in teaching applied knowledge and skills in the context of STEAM education. The role of engineering and its optimization in academic immersion is highlighted. Engineering learning can be enhanced in immersive situations, as extended processes such as projects, problem-solving, design, and testing require iterations and improvements (Soroko, 2024). That is, attempts, practice, and experimentation allow us to build on previous knowledge. This will enable students to consolidate scientific reasoning as they perform variable control, such as when adjusting models to fit a specific design.

Likewise, a study by Kaur et al. (2022) concluded that the role of immersive education in technology is beneficial for students. Specifically, the authors note that employing technological tools in the educational context improves student participation by integrating interactive content into sessions. This, in turn, translates into improved learning and student motivation in STEAM education. Another study that relates virtual situations with educational immersion is that by Piboon et al. (2024). In this study, a virtual studio learning environment was created based on the STEAM education concept to foster students' scientific creativity. Seventy-five secondary school students participated and underwent an immersive educational environment for eight weeks. Assessments were conducted before, during, and after the implementation of the virtual environment. The results suggest that the average score was higher after the intervention, with a significance level of 0.05.

In summary, the prior studies highlight the benefits of virtual environments and immersive STEAM education. These approaches foster experiential learning by actively involving students in practical experiences and reflection, thereby building scientific knowledge and fundamental competencies for their future careers. Within the framework of STEAM education, immersive learning is consolidated as a strategy that enhances the development of essential competencies for scientific reasoning. By fostering skills such as problem-solving, design thinking, critical thinking, creativity, and data analysis, this approach supports cognitive processes inherent to the empirical cycle, including observation, hypothesis formulation, experimentation, and the interpretation of results. In this way, immersive learning not only strengthens the interdisciplinary integration characteristic of the STEAM approach but also promotes the systematic construction of scientific thinking in students (Kumar and Deák, 2023). Next, aspects related to scaffolding, challenging questions, and their combined role in developing scientific reasoning will be considered.

1.3 The Role of Challenging Questions and Scaffolding in Scientific Reasoning

Questions and the search for answers are essential in scientific development, as well as in the processes of teaching and learning science (Tort, 2005). A challenging question is not just any type of question; it is one characterized by a specific educational intention, one that promotes understanding, higher-order cognitive processes, scientific thinking, and assesses competencies (Márquez & Roca, 2006).

According to Márquez and Roca (2006), challenging questions have specific characteristics: 1) They promote knowledge production and foster the acquisition of creative thinking, 2) They are student-centered, prompting students to mobilize their own knowledge rather than being oriented towards a topic that they may or may not implicitly know, 3) They are contextualized, 4) They provide clues about the model, theory, or concepts involved, 5) They pose a clear demand to the student. In summary, according to Tsankov (2018), answering a problem-based question demands the following cognitive demands: the student enters an active position, creates explanations, plans the application of knowledge and skills in solving the problem, and improves self-efficacy. Challenging questions are characterized not only by being open-ended and causal but also by being particularly effective at stimulating learning. They are also characterized, according to Rojas et al. (2020), by having a problem-based structure.

Next, it is important to review empirical evidence regarding the role of questions. Van Vondel et al. (2017) analyzed the instructional skills required of teachers that lead to the construction of scientific thinking. In classroom interaction, the teacher can give different instructions that lead to different expressions in students. The authors highlighted four instructions: 1) Giving an order, 2) Providing information, 3) Asking a closed or memory-based question, and 4) Proposing a stimulating follow-up through open-ended questions. The authors above also suggest that students develop scientific understanding more effectively when educational approaches include stimulating follow-up activities and challenging questions.

In another study, Saysal (2019) investigated what cognitive demands might be present in teachers' questions during science classes. To do this, they recorded class sessions of a science teacher with their 26 sixth-grade students. Two coding schemes were used: one analyzed the teacher's discursive functions when asking questions, and another analyzed the demands integrated into the questions. Primarily, the teacher's questions promoted cognitive demands such as making predictions, hypothesis formulation, making inferences from observations, and evaluating evidence.

Additionally, Alanazi et al. (2024) investigated the effectiveness of guided discovery and scaffolding strategies in improving problem-solving skills in physics students. The authors designed a quasi-experiment with a pre- and post-intervention assessment. The results show that guided discovery and scaffolding improve students' problem-solving skills. This is not the case for those who receive instruction in the context of traditional teaching and without technological immersions.

For the purpose of investigating how teachers use questions to scaffold thinking and help students construct scientific knowledge, Chin (2007) analyzed 36 instructional sequences from six teachers in seventh-grade science classes. The author proposed a framework that included four types of questions: Socratic, verbal puzzles, semantic network, and framing. The questions contributed to various forms of student thinking, such as recalling information, generating ideas, applying concepts, making comparisons, formulating hypotheses, predicting results, giving explanations, analyzing data, making inferences, evaluating information, and making connections between ideas. In the present study, the challenging questions correspond to the Socratic type used by Chin (2007).

2 Method

A microgenetic design consisting of eight observation sessions was used. This allowed for an intra-individual analysis of the participants' scientific reasoning evoked by the challenging question throughout the immersive virtual simulation. According to Crowley and Siegler (1991), this method consists of individually studying a group of subjects during a series of successive observations. This design has three fundamental properties: 1) conducting observations following the principle of developmental change (psychological, in this context), extending until stability is reached, 2) a high density of observations, and 3) in-depth analysis of the variables under study. Exploring the complexity of scientific reasoning that arises during a virtual STEAM problem-solving activity presented to participants can be achieved through successive observations that allow us to observe the processes of change in the participants' scientific reasoning. However, the research design, lacking a pre-post test or comparison group, constitutes a limitation in establishing the causes of the advances in scientific reasoning. This is the reason why in this study, we did not try to estimate the effects of a specific intervention.

2.1 Sample

Fifty-three 10th grade students (28 female and 25 male) between 15 and 17 years old (Mage = 15.4 years; SD = 0.63) from Cali, Colombia, participated. All students solved the problem eight times in an immersive virtual simulation. The recorded unit of analysis is the level of scientific reasoning complexity demonstrated by the participants in their responses in each session when answering the challenging questions posed by the software. For the analysis, a rubric, which is described later, was used.

2.2 Ethical Considerations

This study followed the Ethical guidelines of the University of Valle. Accordingly, the planned data collection strategies did not present physical, moral, mental, emotional, or social risks for the participants, neither now nor 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).

2.3 Instruments

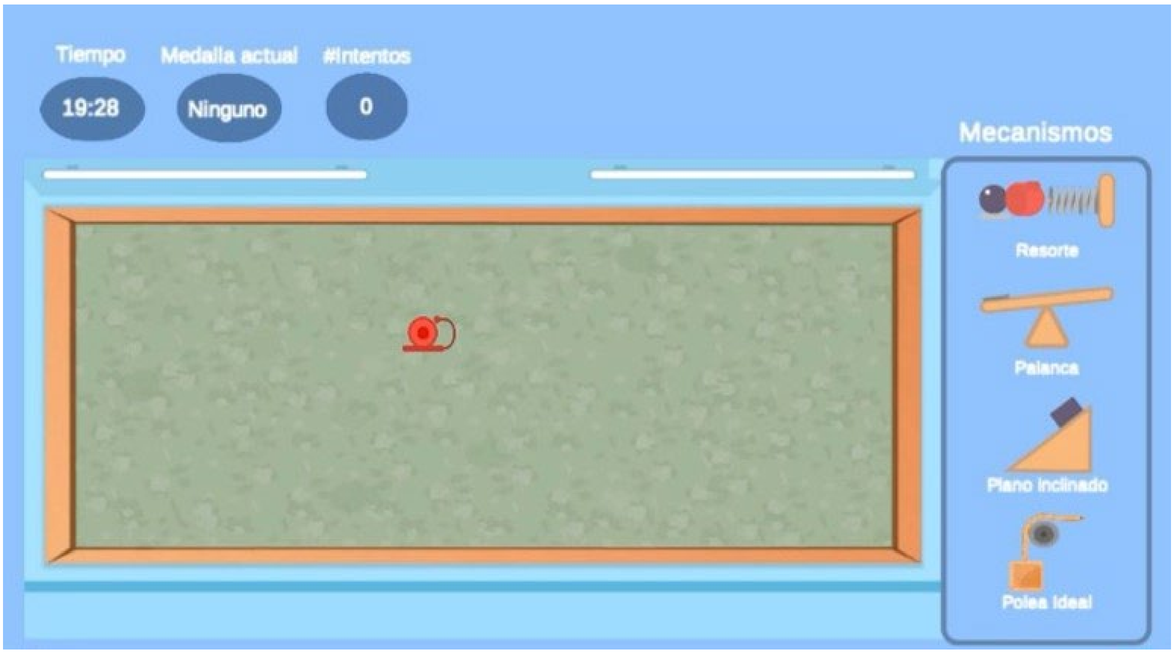
An immersive virtual simulation, the "Scientific Machine," was developed to assess the cognitive demands placed on secondary school students when responding to challenging questions concerning physical mechanisms. This situation was created for this study due to the importance of virtual scenarios in facilitating learning processes. Furthermore, due to its gamification, it is novel for students (Boytchev & Boytcheva, 2020). When designing the situation, it was considered consistent with the science education content that the participants were studying in their respective school year. Furthermore, as a practical and contextualized problem-solving situation, it allowed participants to experiment, formulate hypotheses, and control variables, actions, and skills essential in scientific reasoning. The designed problem-solving situation also offers a technical advantage: it allows participants' responses to be saved and downloaded for later analysis and coding.

The design took into account the physical mechanisms (simple machines) that are taught according to the Colombian educational curricular standards for tenth grade, as per the Ministry of National Education. The physical mechanisms involved were: spring, lever, inclined plane, pulley, hydraulic piston, rope, pendulum, and catapult. The Scientific Machine consists of activating a bell on the screen by linking together as many of the four available physical mechanisms as possible.

The problem situation has two sets. The first set is composed of the following physical mechanisms: spring, lever, inclined plane, and pulley. The second set is composed of the following physical mechanisms: hydraulic piston, rope, pendulum, and catapult. The sets were presented in an alternating manner; set 1 was applied first, followed by set 2, with this dynamic repeated up to session 8. That is, set 1 was presented in sessions 1, 3, 5, and 7, while set 2 was implemented in sessions 2, 4, 6, and 8. Figure 1 shows the interface of the problem situation with the physical mechanisms of set 1. The Scientific Machine, as designed for the present study, conforms to the Colombian natural sciences and physics curriculum established for 10th-grade secondary education. By combining several physical mechanisms, the Scientific Machine requires students to use systemic and critical thinking skills to understand and articulate the relationship between the mechanisms involved (i.e., pulley, inclined plane, lever, spring, hydraulic piston, pendulum, catapult, and rope). In other words, both solving the designed situation and answering the challenging questions involved in the virtual game require the use of scientific reasoning skills and integration of knowledge in STEAM areas (i.e., formulating hypotheses, identifying variables that influence the results of an experiment, proposing models, using tools to solve problems,

designing and creating solutions to problems, and performing mathematical operations related to speed and gravity). The following link shows a video illustrating the interface of the Scientific Machine problem-solving situation:
https://www.youtube.com/watch?v=yXodGM_gJXI.

Figure 1. "The Scientific Machine".



As an instruction, participants were told that the physical mechanisms located on the right side of the game screen should be dragged with the cursor to the game board and linked together to activate the bell. In each game session, each participant had to answer the following questions, which appeared in a text box on the screen:

- Why did you build the Scientific Machine that way?
- How do you think the Scientific Machine will work to activate the bell?
- Why do you think it was not possible to activate the bell?
- Why do you think it was possible to activate the bell?

In the present study, participants' responses were explored and evaluated to identify the cognitive demands evoked by these questions when solving the Scientific Machine, taking into account the approaches of Caspari et al. (2018) and Moreira et al. (2019). Responses were coded for scientific reasoning complexity: a score of 1 was given for descriptions of physical mechanisms (indicating lower complexity), while a score of 5 was assigned for causal explanations describing the behavior of these mechanisms (indicating maximum complexity). Table 1 shows the rubric for evaluating the complexity of the scientific reasoning used.

Table 1. Rubric for evaluating cognitive demands required by the challenging question.

Cognitive Demands	Score
Reasoning based on descriptions: Explanations include references to the properties of the physical concepts and physical mechanisms of the problem situation, without a causal link between the properties of the phenomenon being analyzed.	1
Reasoning based on simple causal relationships: Explanations refer to a single cause or property of the physical mechanisms or to a physical concept involved in the problem situation.	2
Reasoning based on complex causal relationships: Explanations integrate several properties of the mechanisms or physical concepts.	3
Partial mechanistic reasoning: Explanations describe the internal behavior of the device's components.	4
Global mechanistic reasoning: Explanations describe the overall behavior of the device.	5

The requested information is included. The immersive virtual game "Scientific Machine" was designed by a systems engineer, based on the physical principles and variables involved in each physical mechanism (i.e., pulley, inclined plane, lever, spring, hydraulic piston, pendulum, catapult, and rope), which were previously individually validated by two physics experts. Furthermore, the structure and parameters of a problem-solving situation were provided by the first author to design the virtual game. This ensured that the software functioned correctly and stored information accurately, which was necessary for the rigor of the subsequent analysis. A pilot test was also conducted with a sample of 20 students to verify their understanding of the instructions and data recording. This test verified that the designed situation was well-structured, that the elements were clear, and that the necessary outputs could be collected and stored securely. For the construct validity of the coding scheme, we adapted the categories based on the recognized prior work of Caspari et al. (2018) and Moreira et al. (2019) and compared them with other schemes used to analyze explanations as a direct indicator of science learning (Bertrand and Namukasa, 2020; Domenici, 2022). The coding scheme is shown in Table 1; this coding scheme may support future implementations of the study or provide new insights to other researchers. Finally, to validate the coding scheme defined for this study, two trained researchers conducted a double coding procedure. An inter-rater reliability test (Kappa test) was used to determine the level of agreement between coders when evaluating 20% of the responses. A Kappa coefficient of 94.3% was obtained, which represents almost perfect agreement (Krippendorff, 2011).

2.4 Analytical Strategy across

To explore the complexity of scientific reasoning that challenging questions evoked in students, we analyzed responses from eight problem-solving sessions involving physical mechanisms. For this analysis, we employed an ordinal logit model, also known as a cumulative logit model, which included a random effect for students (e.g., Feldman & Rabe-Hesketh et al., 2012). This model provides a flexible means to assess if the proportion of student responses demonstrating higher scientific reasoning complexity increases from session to session. Moreover, it facilitates an evaluation of potential differences in student responses between physical mechanisms: spring, lever, inclined plane, and pulley from set 1 (sessions 1, 3, 5, and 7) and the physical mechanisms: hydraulic piston, rope, pendulum, and catapult from set 2 (sessions 2, 4, 6, and 8).

In formal terms, the model used can be expressed with the following equation:

$$logit\{\Pr(y_{ij} > s)\} = \theta_j - \tau_s$$

In this model, the score that student 'j' receives in each session 'i', in category 's' of the scientific reasoning rubric, depends on the student-specific intercepts j (θ_j), minus the thresholds for each category change assigned by the rubric across all sessions (τ_s). The term θ_j represents the students' propensity to provide responses that exhibit more or less complex scientific reasoning. This term is a normal random effect, or latent variable, with a zero mean and variance of σ^2 For its part, the term τ_s informs us about the cumulative probability odds that the sample as a whole presents, across all sessions, of providing responses with higher or lower scientific reasoning (1 vs 2, 3, 4, 5; 1, 2 vs 3, 4, 5; 1, 2, 3 vs 4, 5; 1, 2, 3, 4 vs 5). In this study, this model is called model 1.

This response model can be expanded to include covariates that allow the key hypotheses to be evaluated. The central idea of this study is that students in later sessions have greater odds of presenting higher proportions of complex scientific reasoning responses, in contrast to the initial sessions. We evaluated this hypothesis with model 2, which is expressed as follows:

$$logit\{\Pr(y_{ij} > s)\} = \gamma\ sessions + \delta\ set1 + \theta_j - \tau_s$$

In this model, the term γ indicates the expected logit increase for each additional session in which the sample participates. If γ>0, we would have evidence to support the hypothesis. For its part, the term δ, would indicate if, on average, there are differences in students' propensities to provide responses with higher or lower scientific reasoning. If δ>0, it would indicate that the physical mechanisms of the first set are easier than those of set 2; in contrast, if δ<0, it would indicate that the physical mechanisms of the second set are easier for the group of students.

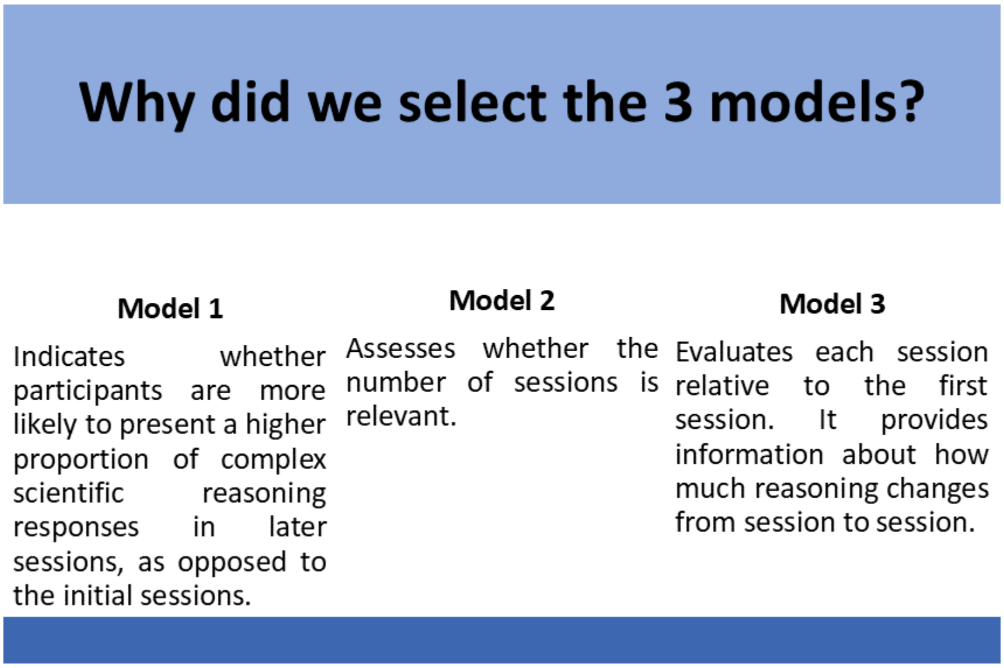
Model 3 was then added to the sequence of fitted models. Model 3 is less structured and allows for a session-by-session comparison of the relative differences in students' responses, in relation to the first session (included as a reference). In this case, we included dummy variables for each subsequent session (session s= 2, the rest of the sessions = 0). This was done to obtain a set of seven coefficients that indicate which sessions show higher scientific reasoning. This model can be expressed as follows:

$$logit\{\Pr(y_{ij} > s)\} = \beta_{2-8}\, sessions_{2-8} + \theta_j - \tau_s$$

In general terms, this model is less parsimonious than the previous model. Still, it is more flexible as it allows us to illustrate, session by session, the relative change that the sample shows with respect to the first session. In this model, it is expected that the coefficients for later sessions are greater than the coefficients for initial sessions ($\beta_t < \beta_{t+1}$).

Consequently, for each session, the odds associated with achieving each level of scientific reasoning complexity (as defined in the study's rubric) can be determined. Furthermore, hypothesis tests regarding how these odds change or grow for each complexity level per session can be conducted. Figure 2 illustrates the rationale behind selecting the three models mentioned above.

Figure 2. Justification for using the models.



3 Results

This study did not analyze the connection between the instructional process and the observed improvement in scientific literacy, as this is beyond the scope of the study. The study aimed to explore the complexity of scientific reasoning that emerges during a virtual STEAM problem-solving activity, utilizing challenging questions. The questions served as the pedagogical resource for exploring scientific reasoning.

3.1 Model Fit

The graded response model implemented indicates if students present scientific reasoning responses of greater complexity as they progress to the next session. What is represented is the probability that the level of reasoning examined is at or above a threshold as a function of the level of the latent trait. This section answers the questions: To what extent does scientific reasoning grow across the number of sessions in a STEAM situation? And in which sessions was there a more marked growth rate? The results indicate that a linear approximation based on the number of sessions is a reasonable approximation.

Similarly, model 3, where we included each session as a dummy variable, also shows a favorable fit to the data (LRT = 159.94 (7), $p < 0.001$) [The 7 in parentheses is the degrees of freedom of the likelihood ratio test]. This indicates that knowing which session is being evaluated is relevant for understanding the degree of complexity presented by the sample's responses. Taken together, both models agree in affirming that the number of sessions and the evaluated sessions are associated with the complexity of students' responses. The model fit indicators can be seen in Table 2.

Table 2. Model fit indicators.

	Model 1	**Model 2**	**Model 3**
Log Likelihood	-672.26	-596.09	-592.28
LRT		152.34(3) ***	159.94(7) ***
AIC (Akaike Information Criterion)	1354.51	1208.17	1208.57
BIC (Bayesian Information Criterion)	1374.76	1240.57	1257.16
Observations	424	424	424
Students	53	53	53

Note. LRT = *likelihood ratio test* of the models in contrast to the null model (Model 1).
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

3.2 Session Number Coefficient

The students' level of scientific reasoning complexity improves (scores are higher) in the later sessions; this is shown by the estimated Y value of 0.54 ($p < 0.001$). Heterogeneous growth (within-student change) throughout the sessions is also observed for the student sample, as indicated by the values in Table 3.

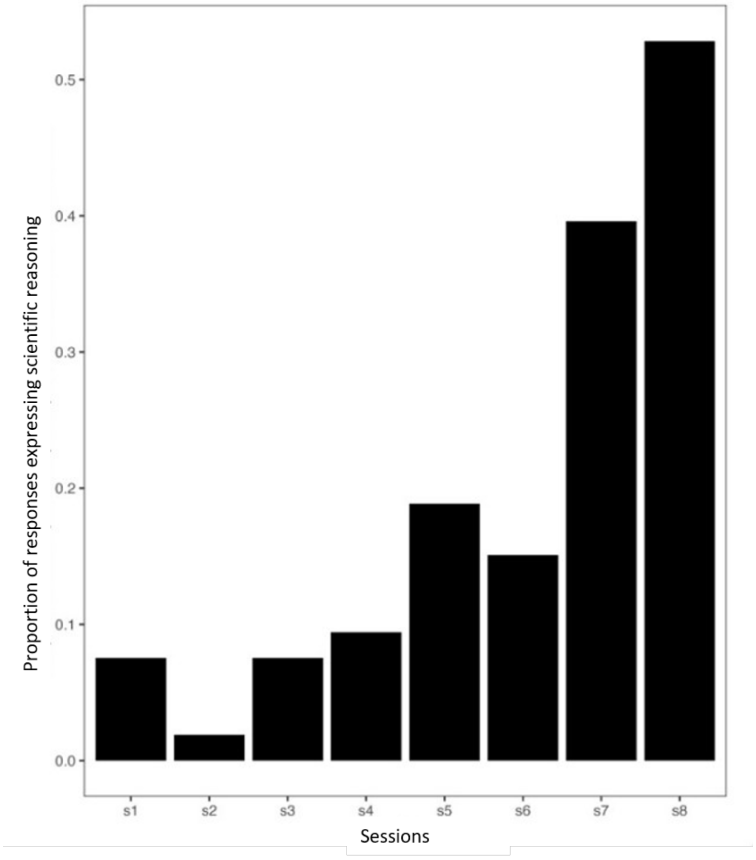
Table 3. Results of the Estimated Models: Estimated responses, standard errors, and P-value for the three models.

		Model 1			Model 2			Model 3		
Parameter	**Terms and variables**	**E**	**EE**	**P<**	**E**	**EE**	**P<**	**E**	**EE**	**P<**
τ_1	Threshold 1	-1.13	(0.12)	***	0.72	(0.23)	**	-0.00	(0.29)	
τ_2	Threshold 2	-0.40	(0.11)	***	1.71	(0.24)	***	0.99	(0.29)	***
τ_3	Threshold 3	0.65	(0.11)	***	3.14	(0.28)	***	2.45	(0.32)	***
τ_4	Threshold 4	1.47	(0.13)	***	4.15	(0.31)	***	3.48	(0.34)	***
γ	Number of sessions				0.54	(0.05)	***			
δ	Machine Set[a]				0.68	(0.19)	***			
λ	Interaction				-0.04	(0.08)				
β_2	Session2							-0.28	(0.38)	
β_3	Session3							1.09	(0.38)	**
β_4	Session4							1.21	(0.38)	**
β_5	Session5							2.14	(0.37)	***
β_6	Session6							1.47	(0.37)	***
β_7	Session7							3.14	(0.39)	***
β_8	Session8							3.44	(0.41)	***
	Variance of θ_j	0.07			0.33			0.34		

E = Unstandardized estimate on *logit* scale, SE = Standard error of the estimate in parentheses
P< = *Wald test* p-value *** p < 0.001; ** p < 0.01; * p< 0.05
[a] *dummy* variable where set 1 = 1, and set 2 = 0.
[b] Term of interaction between the number of sessions and the machine set.

Figure 3 shows the distribution of responses with a scientific reasoning score of 5 among students during the eight sessions. It is evident that the level of scientific reasoning grows throughout the sessions. The estimated time parameter of 0.56 from model 2 is consistent with what is shown in Figure 2, where a progressive and ascending growth of responses with a score of 5 is observed. As sessions progress, the accumulation of high scores becomes more probable than that of low scores. The results for each session are always higher than those obtained in the previous session.

Figure 3. Distribution of responses with a score of 5 during the eight sessions.



In Figure 2, it can be seen that the proportion of scores of 5 grows during the first six sessions, but in the last two sessions, this growth is even greater. The prior figure is a graphical representation of the obtained results that allow model 3's results to be confirmed.

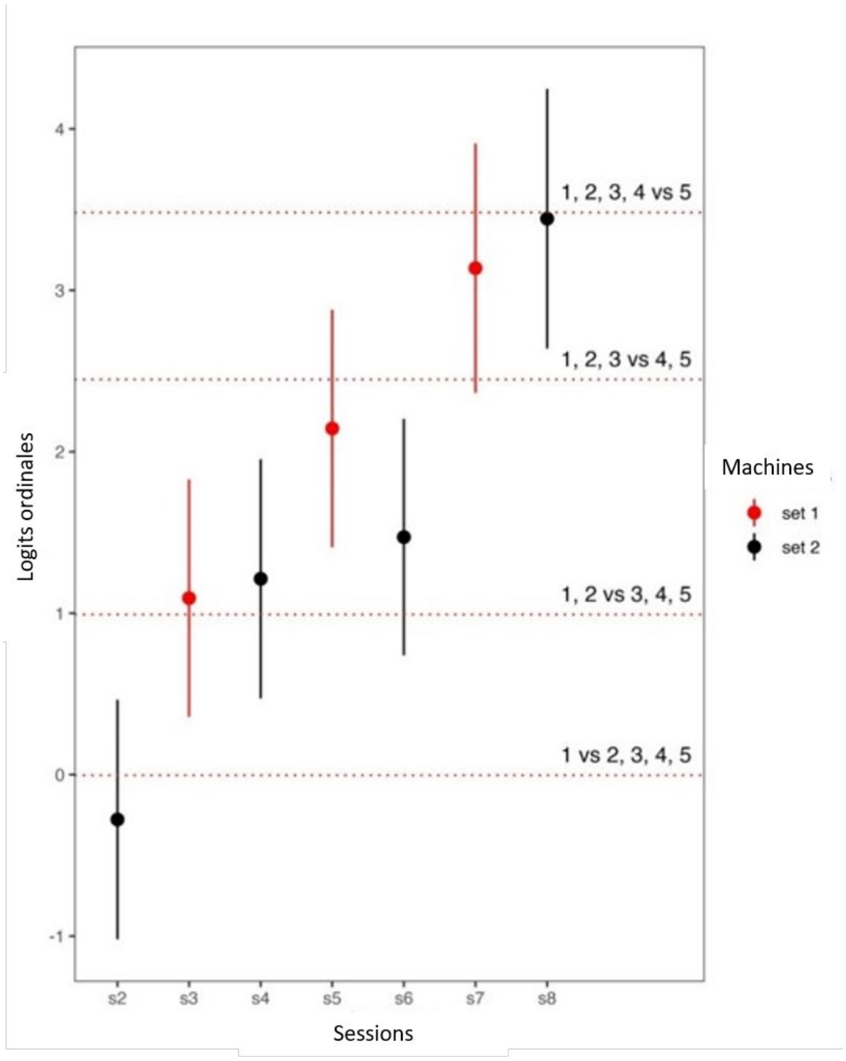
3.3 Model 2 Set Effect

Set 2 presents a slight difference compared to the results of set 1, which is because it may be slightly more challenging to solve. However, the difference is not significant, as it is very close to the results of set 1.

3.4 Partial Effects of Each Session - Model 3

Figure 4 suggests that the results obtained with set 1 are similar to those obtained with set 2, which is why the session pairs 1-2, 3-4, 5-6, and 7-8 follow the same upward diagonal pattern. Figure 3 is a graphical representation of what occurs with model 3, and it shows which sessions present higher scientific reasoning, that is, the relative session-by-session changes in student performance. The results of the sessions, when considered together, suggest that it is unreasonable to think that the results are different between the sets. This is also consistent with the result of model 2, which indicates that the difference between both sets is small.

Figure 4. Item-person map generated with model 3.



The logits observed in Figure 3 show linear growth as the sessions progress. The logits are shown in pairs of sessions corresponding to sets 1 and 2 (1 and 2; 3 and 4; 5 and 6; 7 and 8). Therefore, it can be said that in the first pair of sessions, students have a low result. However, as the sessions progress, performance improves session by session. Furthermore, it can be seen that the differences in the level of participants' scientific reasoning complexity between the pairs of sessions do not suggest differences between the physical mechanisms of the sets. The previous results allow this study's hypotheses to be confirmed.

Considering the results obtained in the present study and the first hypothesis posed that the level of complexity of students' scientific reasoning will improve in subsequent sessions, this hypothesis is interpreted as confirmed. Models 2 and 3 show a significant increase in scientific reasoning scores as the sessions progress ($\gamma = 0.54$, $p < 0.001$). This growth is also evident in the distribution of the highest scores (level 5) across the eight sessions, as shown in Figure 2. The probability of students producing more complex responses progressively increased, being highest in the later sessions. These results suggest that continued participation in problem-solving activities such as the Science Machine favors the gradual development of scientific engagement skills. The sustained improvement in causal explanations, variable identification, and the use of physical and mathematical concepts reflects a construction of scientific engagement as a process over time.

Regarding the second hypothesis, growth (change in scientific reasoning among students over time) in scientific reasoning will occur, and this growth will be heterogeneous across students, the results confirm this hypothesis. The results of the multivariate growth model indicate a heterogeneous pattern of change across students, evidenced by the significant variance of the growth parameters (Variance of θ_j = 0.33 in Model 2 and 0.34 in Model 3). This means that, although the group as a whole shows a positive trend in scientific reasoning, this variability can be attributed to intraindividual factors (such as differences in reasoning strategies or prior knowledge) and interindividual factors (including interactions, collaboration, and teacher mediation). The heterogeneity of growth supports the idea that scientific reasoning does not develop uniformly, but is built from personal trajectories influenced by practice, reflection, and the learning context. It would be interesting to explore the observed variability in a second phase of this study or in future research.

4 Discussion

This study explored the scientific reasoning complexity that challenging questions evoked in students when solving a STEAM problem involving physical mechanisms in an interactive simulation. When faced with these challenging questions about physical mechanisms, the student sample predominantly demonstrated cognitive demands related to descriptive reasoning in sessions 1, 2, 3, and 4. In contrast, from session 5 onwards, their responses showed a shift towards complex causal reasoning and global mechanistic reasoning. In fact, it was in the fifth session where complex causal reasoning appeared in a greater proportion, and from that point on, more complex reasoning began to be more evident in the students' responses. The assessment carried out in the initial moments accounts for an initial competence, while in the later ones, the complexity of the students' scientific reasoning becomes more robust. This aligns with Alonzo and Loughland's (2022) proposal that learning unfolds across sessions. Feedback from each session informs how students approach subsequent problem situations, reflecting a dynamic assessment process that evolves continuously and should be considered holistically, not just at specific points, such as session 1 or 8. Furthermore, the results demonstrate the increasing sophistication of students' reasoning, which was captured through their responses to challenging, software-generated questions within the problem situation. These results are similar to those of Saysal (2019), who found that teachers' questions in science classes can be intentionally designed to promote cognitive demands related to science.

Since the Scientific Machine and challenging questions were consistently present throughout the sessions, the observed increase in reasoning complexity as sessions advanced likely indicates the students' growing mastery of physical concepts. This progress between sessions can be attributed to their ongoing school physics classes, which occurred concurrently with the sessions. Methodologically, the session-by-session progress seen in this study aligns with Van Vondel et al. (2017). They suggested that students' scientific understanding develops optimally when school interventions use stimulating assessments and challenging questions extended over a period of time.

There may also be a practice effect generated by the problem situation, which might have made it possible for reasoning to be expressed in more complex ways as the cognitive demand prompted by the question was reiterated. In fact, without this type of challenging question integrated into the Scientific Machine, it is possible that students would have progressed in solving the problem, but their understanding of why they were progressing would have remained implicit. Thus, the challenging question designed for this situation began to encourage the expression of more complex reasoning starting from the 5th session for complex causal reasoning, and from the seventh and eighth sessions for global mechanistic reasoning (Moreira et al., 2019). Scientific reasoning is not achieved immediately upon completing the learning process with the Scientific Machine in the final session; rather, it is a process that unfolds through different stages and degrees of difficulty (Alonzo and Loughland, 2022).

In this specific case, the cognitive demand of the question, which involved scientific explanations, aligned with the reasoning demonstrated by the majority of students in the fifth session. This highlights that student reasoning is a process that takes time, organization, and reorganization. These findings resonate with the study by Chin (2007), who found that questions served as scaffolding for thought and helped students build scientific knowledge. Furthermore, they are a call to consider assessment not only at the beginning or end of learning but throughout the entire process of interaction with the simulation when learning sciences.

When interpreting the previous results, it is essential to consider that experience gained in the early sessions allows students to approach successive tasks more systematically. This improvement occurs because previous attempts provide feedback or prompt the development of new and better explanations for solving the problem, which allows the student to achieve an increasingly more profound and better understanding of the problem situation (Efremova et al., 2020). The reasoning that the student shows in solving the task was identified in this study as a catalyst for a process of understanding the physical mechanisms involved in a process constructed throughout the sessions. The student gradually adopted an active position throughout the sessions, creating explanations of greater complexity and planning the application of knowledge, thereby improving their problem-solving skills, as noted by Tsankov (2018).

The results of the present study demonstrate how, throughout the sessions, the complexity of students' reasoning improves. This improvement is related to the possibility of educational immersion offered by the Scientific Machine. The virtual situation allows the student to put their knowledge into practice and experiment by making several attempts to solve it (Wu et al., 2021). The participant controls and analyzes variables involved in the physical functioning of the mechanisms to improve their performance and activate the bell as they approach the goal. Additionally, the microgenetic design favors STEAM educational immersion over eight sessions (Piboon et al., 2024). The process experienced throughout the sessions makes it possible to approach scientific knowledge of physics, experiential learning, critical thinking, and problem-solving (Beck et al., 2024).

From the results, it is evident that the question in this study demanded specific cognitive and practical requirements when solving the situation. Additionally, the question was challenging as it was framed within the contextualized problem-solving situation that involved linking physical mechanisms to activate the target element. All the foregoing characteristics, according to Márquez and Roca (2006), signal that the questions used are conceived as challenging and, possibly, promote physics knowledge in students.

The results obtained and the experience of working with the Scientific Machine enabled the formulation of educational implications. Firstly, this type of challenging question, although it may seem difficult at first, can constitute a tool to assess and foster scientific reasoning and desired competencies for students' academic and professional future. Secondly, the assessment of scientific reasoning must be dynamic and extended, as otherwise, the results may be partial and not account for the complexity or variability with which reasoning fluctuates during the student's training period. The results of the present study contribute to the current need to promote science teaching, as expressed in the research conducted by Castañeda Zapata et al. (2024).

The substantive question of the study was to verify whether it could be assumed that learning growth was linear with the number of sessions, or whether there were some sessions where growth was more rapid than others. However, as a hypothesis, we believe that students' participation in the Scientific Machine and the challenging questions posed imply an implicit pedagogical link (for the purpose of this study) that also favors participants' scientific reasoning. Specifically, the Scientific Machine and the challenging questions generate potential scaffolds that contribute to the construction of participants' scientific reasoning throughout the sessions. This could be explained from a sociocultural perspective, based on Vygotsky (Utomo & Santoso, 2021).

As an alternative in digital pedagogy, the Scientific Machine enables students to tackle practical challenges, thereby improving their STEAM skills and knowledge (Dhitasarifa & Wusqo, 2024). Furthermore, its autonomous application simplifies how teachers can pose and record questions. The Scientific Machine and its questions represent a departure from traditional teaching methods (Castro, 2022). For instance, conventional physics classes might address simple machines with a typical textbook problem like: A 3 kg body slides on an inclined plane at a 20-degree angle with the horizontal. The coefficient of friction is

0.3. If released from rest, what is its acceleration and velocity after 4 seconds? This traditional problem is closed-ended and seeks a short, memory-based answer.

In contrast, the challenging questions used in this study's Scientific Machine aim for knowledge centered on causal explanations. Both these questions and the Scientific Machine's virtual problem-solving environment were designed to foster significant connections between scientific reasoning, physics concepts, and students' practical problem-solving skills (Alanazi et al., 2024). The observed increase in the complexity of students' scientific reasoning across the eight sessions suggests student improvement and highlights the benefits of a scenario promoting STEAM educational immersion (Wu et al., 2021).

The observed increase in the complexity of students' scientific reasoning over the eight sessions demonstrates the relevance of studies using repeated measurements and microgenetic analysis. The results show sustained growth in scientific reasoning, reflecting an evolutionary process that unfolds over time. However, this change does not follow a single trajectory but can manifest itself at different rates and directions (Guevara et al., 2016; Dejonckheere et al., 2009). The session-by-session progress indicates a steady improvement in scientific reasoning skills, with each new learning process building upon prior knowledge. This demonstrates that scientific reasoning emerges moment by moment, articulating intraindividual and interindividual processes. Additionally, the results invite reflection on learning assessment criteria. The difference between the performances observed in Session 1 and Session 8 highlights the importance of considering temporality in the assessment of knowledge, recognizing that learning involves both short- and long-term processes of change.

On the other hand, the combination of virtual gaming and challenging questions served as a key pedagogical mediator, encouraging students to formulate causal explanations for their actions during the game. These explanations included formulating hypotheses, identifying variables that affect the mechanisms of the virtual environment, understanding how simple machines (such as levers and pulleys) work, using problem-solving tools, and applying physical concepts, such as speed and gravity, through mathematical operations. The results show a progressive improvement in the quality of scientific explanations, which contributes to narrowing the gap identified by the PISA test, where Colombian students typically perform below the OECD average.

Additionally, the findings of this study align with the goals of STEAM education, which aims to strengthen scientific reasoning skills by integrating the knowledge domains of science, technology, engineering, art, and mathematics (Guzey, 2015). A relevant aspect of the Scientific Machine is that it provides an opportunity to access digital resources, which are still limited in some educational contexts, thereby facilitating the development of digital competences in students (Madelson and Seifert, 2017). Bearing in mind the above and the results obtained in the present study, some practical recommendations can be made for teachers to consider: 1) Integrate interactive virtual environments such as virtual scientific games that favor the analysis of variables, the formulation of hypotheses, and problem solving. 2) Use open-ended questions that invite students to justify, predict, and argue their answers. 3) Explore with students the metacognitive processes that emerge during the completion of STEAM activities, considering, for example, the following approaches: why you think that, how you would solve the problem, and what strategies you could employ.

In conclusion, challenging questions foster students' scientific reasoning when they are part of robust instructional designs within problem-solving contexts and contain structured characteristics such as those proposed by Castro (2022). The result of the present study is also in line with Bargiela et al. (2022), who found that questions in science class activate the explanation, inference, and analysis skills of the student sample. The findings also highlight the academic engagement of the participants and an improvement in their scientific reasoning, as reported by Álvarez-Pérez et al. (2021).

This study analyzed the cognitive demands elicited in 10th-grade students through challenging questions posed within the 'Scientific Machine' problem-solving environment. Although the characteristics of challenging questions make them specific and allow them to scaffold learning, care must be taken when generalizing the results. The results of the present study align with the theoretical and methodological views that support microgenetic analysis as appropriate for fostering cognitive development in participants (Sánchez et al., 2022; Bargiela et al., 2022). As limitations, we believe that increasing the sample size, including another coder, and implementing the study in more classrooms would allow for continued research and enriched results.

It can be concluded from the present study that the participants' scientific reasoning demonstrated progressive complexification throughout the sessions, moving from descriptive approaches towards increasingly sophisticated modes, such as mechanistic reasoning. This progression suggests that challenging questions played a triggering or catalyzing role in improving student reasoning throughout the sessions. Notably, an alignment between the cognitive demands of the tasks and the evidence of students' reasoning became particularly evident from the fifth session onwards. The variability in the results implies the need for dynamic and extensive assessment of scientific competencies to guarantee a complex analysis of how scientific reasoning develops over time.

Research Ethics

Author contributions

K.L.: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing
M.G.: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing
V.C.: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing

J.M.: conceptualization, investigation, methodology, writing—original draft preparation, writing—review and editing.

All authors have read and agreed to the published version of the manuscript.

Artificial intelligence

No artificial intelligence was used in the research or writing of this article.

Funding

This work was supported by ANID – MILENIO – NC2021_014 and FONDECYT project 1221716.

Institutional review board statement

The study was ethically endorsed by the Ethics Committee of the Universidad del Valle Cali, Colombia, as recorded in the APPROVAL RECORD No. (025-021).

Informed consent statement

Informed consent was obtained from all research participants.

Data availability statement

The data are stored on the researchers' computers and hard drives. They are duly labeled and stored. They may be accessed with prior authorization. For questions regarding data availability, please contact jose.julian.kenji@gmail.com.

Acknowledgements

We would like to thank the Faculty of Education at Pontificia Universidad Católica de Chile for their support translating this article, and Professor Diego Carrasco, of Pontificia Universidad Católica de Chile, for his collaboration in reviewing the statistical analysis of this study.

Conflicts of interest

The authors declare no conflicts of interest.

References

Ahmad, A. R., Chew, F. P., Zulnaidi, H., Sobri, K. M., & Alfitri. (2019). Influence of School Culture and Classroom Environment in Improving Soft Skills amongst Secondary Schoolers. *International Journal of Instruction, 12*(2), 259–274. https://doi.org/10.29333/iji.2019.12217a

Alanazi, A., Osman, K., & Halim, L. (2024). Enhancing physics problem-solving skills through guided discovery and scaffolding strategies: Evidence from Saudi technical colleges. *LUMAT: International Journal on Math, Science and Technology Education, 12*(4), Article 4. https://doi.org/10.31129/LUMAT.12.4.2329

Alderete, M. V., Di Meglio, G., & Formichella, M. M. (2017). Acceso a las TIC y rendimiento educativo: ¿una relación potenciada por su uso? Un análisis para España. *Revista de Educación*, 377. https://doi.org/10.4438/1988-592X-RE-2017-377-353

Asale, T. S. (2017). Teachers' Perception and Practices Towards Continuous Assessment of Mathematics Classes: The Case of Secondary School in Wolaita Zone, Snnpr Region.

Barbera, E. (2004). Quality in virtual education environments. *British Journal of Educational Technology, 35*(1), 13–20. https://doi.org/10.1111/j.1467-8535.2004.00364.x

Bargallo, C. M., & Tort, M. R. (2006). Plantear preguntas: Un punto de partida para aprender ciencias. *Revista Educación y Pedagogía, 18(*45), Article 45.

Bargiela, I. M., Anaya, P. B., & Puig, B. (2022). Las preguntas para la indagación y activación de pensamiento crítico en educación infantil. Enseñanza de las Ciencias. *Revista de investigación y experiencias didácticas, 40*(3), Article 3. https://doi.org/10.5565/rev/ensciencias.5470

Beck, D., Morgado, L., & O'Shea, P. (2024). Educational Practices and Strategies with Immersive Learning Environments: Mapping of Reviews for Using the Metaverse. In IEEE Transactions on Learning Technologies, vol. 17, pp. 319–341, 2024, https://doi.org/10.1109/TLT.2023.3243946

Belbase, S., Mainali, B. R., Kasemsukpipat, W., Tairab, H., Gochoo, M., & Jarrah, A. (2021). At the dawn of science, technology, engineering, arts, and mathematics (STEAM) education: prospects, priorities, processes, and problems. *International Journal of Mathematical Education in Science and Technology, 53*(11), 2919–2955. https://doi.org/10.1080/0020739X.2021.1922943

Berezi, I. U. (2025). Virtual Learning Environment: Redefining Higher Educational Delivery for Efficiency and Accessibility. *International Journal of Educational Management, Rivers State University., 1*(1), 451–467.

Bertrand, M. G., & Namukasa, I. K. (2020). STEAM education: student learning and transferable skills. *Journal of Research in Innovative Teaching & Learning 24 June 2020, 13*(1): 43–56. https://doi.org/10.1108/JRIT-01-2020-0003

Blythe, C., & and Harré, N. (2020). Encouraging transformation and action competence: A Theory of Change evaluation of a sustainability leadership program for high school students. *The Journal of Environmental Education, 51*(1), 83–96. https://doi.org/10.1080/00958964.2019.1629381

Boytchev, P., & Boytcheva, S. (2020). Gamified Evaluation in STEAM for Higher Education: A Case Study. *Information, 11*(6), 316. https://doi.org/10.3390/info11060316

Bråten, I., & Braasch, J. L. G. (2017). Key Issues in Research on Students' Critical Reading and Learning in the 21st Century Information Society. In C. Ng & B. Bartlett (Eds.), Improving Reading and Reading Engagement in the 21st Century: International Research and Innovation (pp. 77–98). Springer. https://doi.org/10.1007/978-981-10-4331-4_4

Brown, G. T. L. (2022). The past, present and future of educational assessment: A transdisciplinary perspective. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.1060633

Bucarey, C. J., Pérez, S. M., Marín, M. G., & Castillo, L. A. (2023). Brechas en percepción de contribución de competencias genéricas entre estudiantes en Chile. *Revista de ciencias sociales, 29*(1), 386–403.

Cabello, V. M., Martinez, M. L., Armijo, S., & Maldonado, L. (2021). Promoting STEAM learning in the early years: "Pequeños Científicos" Program. *LUMAT: International Journal on Math, Science and Technology Education, 9*(2), 33–62. https://doi.org/10.31129/LUMAT.9.2.1401

Caspari, I., Weinrich, M. L., Sevian, H., & Graulich, N. (2018). This mechanistic step is " productive ": Organic chemistry students' backward-oriented reasoning. *Chemistry Education Research and Practice, 19*(1), 42–59. https://doi.org/10.1039/C7RP00124J

Castañeda Zapata, E. A., López Ríos, S. Y., & Osorio Vélez, J. A. (2024). Naturaleza de la ciencia para la enseñanza de la relatividad general: Una revisión sistemática. *Aula abierta, 53*(2), 149–158.

Castro, P. (2022). Reflexiones sobre la educación STEAM, alternativa para el siglo XXI. *Praxis, 18*(1), 158–175.

Cheng, S.-C., She, H.-C., & Huang, L.-Y. (2017). The Impact of Problem-Solving Instruction on Middle School Students' Physical Science Learning: Interplays of Knowledge, Reasoning, and Problem Solving. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(3). https://doi.org/10.12973/ejmste/80902

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching, 44*(6), 815–843. https://doi.org/10.1002/tea.20171

Conejera, A. R., Joglar, C., & Jara, R. (2020). Promoviendo la Formulación de Buenas Preguntas en la Clase de Biología en Secundaria: Una propuesta didáctica a partir de situaciones problema. Ciência & Educação (Bauru), 26, e20034. https://doi.org/10.1590/1516-731320200034

Dejonckheere, P. J. N., Van De Keere ,Kristof, & and Mestdagh, N. (2009). Training the Scientific Thinking Circle in Pre- and Primary School Children. *The Journal of Educational Research, 103(*1), 1–16. https://doi.org/10.1080/00220670903228595

Dengel, A., & Mägdefrau, J. (2018). Immersive Learning Explored: Subjective and Objective Factors Influencing Learning Outcomes in Immersive Educational Virtual Environments. 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 608–615. https://doi.org/10.1109/TALE.2018.8615281

Dhitasarifa, I., & Wusqo, I. U. (2024). The effect of STEAM approach digital teaching materials on increasing creative problem-solving skills. *Turkish Online Journal of Distance Education, 25*(3), 18–27. https://doi.org/10.17718/tojde.1302079

Domenici, V. (2022). STEAM Project-Based Learning Activities at the Science Museum as an Effective Training for Future Chemistry Teachers. *Education Sciences, 12*(1), 30. https://doi.org/10.3390/educsci12010030

Efremova, N., Shapovalova, O., & Huseynova, A. (2020). Innovative technologies for the formation and assessment of competencies and skills in the XXI century. E3S Web of Conferences, 210, 18021. https://doi.org/10.1051/e3sconf/202021018021

Eshach, H., Dor-Ziderman, Y., & Yefroimsky, Y. (2014). Question Asking in the Science Classroom: Teacher Attitudes and Practices. *Journal of Science Education and Technology, 23*(1), 67–81. https://doi.org/10.1007/s10956-013-9451-y

Feldman, B. J., & Rabe-Hesketh, S. (2012). Modeling Achievement Trajectories When Attrition Is Informative. *Journal of Educational and Behavioral Statistics, 37*(6), 703–736. https://doi.org/10.3102/1076998612458701

Gim, N. (2021). Development of Life Skills Program for Primary School Students: Focus on Entry Programming. *Computers, 10(*5), Article 5. https://doi.org/10.3390/computers10050056

Guevara, M., van Dijk, M., & van Geert, P. (2016). Microdevelopment of peer interactions and scientific reasoning in young children / Microdesarrollo de la interacción entre pares y el razonamiento científico en niños pequeños. *Journal for the Study of Education and Development, 39*(4), 727–771. https://doi.org/10.1080/02103702.2016.1215083

Guzey, S. (2015). The Need for a STEM Road Map Tamara J. Moore, Carla C. Johnson, Erin E. Peters-Burton, and S. STEM Road Map: A Framework for Integrated STEM Education.

Hidri, S. (2019). Static vs. dynamic assessment of students' writing exams: A comparison of two assessment modes. *International Multilingual Research Journal, 13*(4), 239–256. https://doi.org/10.1080/19313152.2019.1606875

ICFES (2024). Programa para la Evaluación Internacional de Alumnos (PISA). Informa nacional de resultados para Colombia 2022. Ministerio de Educación Nacional. https://www.mineducacion.gov.co/1780/articles-421217_recurso_03.pdf

Kaur, D. P., Kumar, A., Dutta, R., & Malhotra, S. (2022). The role of interactive and immersive technologies in higher education: A survey. J*ournal of Engineering Education Transformations, 36*(2), 79-86. https://doi.org/10.16920/jeet/2025/v38is2/25001

Kawalkar, A., & Vijapurkar, J. (2011). Scaffolding Science Talk: The role of teachers' questions in the inquiry classroom. *International Journal of Science Education, 35*(12), 2004–2027. https://doi.org/10.1080/09500693.2011.604684

Krippendorff, K. (2011). Agreement and Information in the Reliability of Coding. *Communication Methods and Measures, 5*(2), 93–112. https://doi.org/10.1080/19312458.2011.568376

Kumar, B., & Deák, C., (2023). Evolving Minds: A Literature-Driven and Empirical Explorationof STEAM Skill Development and Learning Approaches. *Journal of Innovation Management, 11(*4), 71–96. https://doi.org/10.24840/2183-0606_011.004_00041

Lau, A. M. S. (2016). 'Formative good, summative bad?' – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education, 40*(4), 509–525. https://doi.org/10.1080/0309877X.2014.984600

Lecturer, School of Education, University of New South Wales Sydney, Australia, d.alonzo@unsw.edu.au, Alonzo, D., Loughland, T., & Assoc. Prof., School of Education, University of New South Wales Sydney, Australia, tony.loughland@unsw.edu.au. (2022). Variability of Students' Responses to Assessment Activities: The Influence of Achievement Levels. *International Journal of Instruction, 15*(4), 1071–1090. https://doi.org/10.29333/iji.2022.15457a

Lee, Y., & Kinzie, M. B. (2012). Teacher question and student response with regard to cognition and language use. *Instructional Science, 40*(6), 857–874. https://doi.org/10.1007/s11251-011-9193-2

López-Aguilar, D., Álvarez-Pérez, P. R., & Garcés-Delgado, Y. (2021). El engagement académico y su incidencia en el rendimiento del alumnado de grado de la Universidad de La Laguna. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa, 27(*1), Article 1. https://doi.org/10.30827/relieve.v27i1.21169

Manassero-Mas, M., & Vázquez-Alonso, Á. (2020). Pensamiento científico y pensamiento crítico: competencias transversales para aprender. Indagatio Didactica, 12(4), 401-419. https://doi.org/10.34624/id.v12i4.21808

Maraza-Quispe, B., Caytuiro-Silva, N., & Arizaca-Machaca, E. (2022). Sistema de Enseñanza-Aprendizaje inteligente basado en el Razonamiento Basado en Casos. 1(2).

Martínez Villalba, J. A., & Sánchez Muñoz, S. (2018). Generación de Competencias con Base en la Gestión de Conocimiento Científico. REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación. https://doi.org/10.15366/reice2018.16.2.004

McPhail, G. (2018). Curriculum integration in the senior secondary school: A case study in a national assessment context. *Journal of Curriculum Studies, 50*(1), 56–76. https://doi.org/10.1080/00220272.2017.1386234

Michael, B., & Michael, R. (2019). Show me and I'll remember: Association between instructional modality and memory for learning. *Journal of International Education in Business, 12*(1), 95–110.

Moreira, P., Marzabal, A., & Talanquer, V. (2019). Investigating the effect of teacher mediation on student expressed reasoning. *Chemistry Education Research and Practice, 20*(3), 606–617. https://doi.org/10.1039/C9RP00075E

Nadelson, L. S., & Seifert, A. L. (2017). Integrated STEM defined: Contexts, challenges, and the future. *The Journal of Educational Research, 110*(3), 221–223. https://doi.org/10.1080/00220671.2017.1289775

Piboon, K., Khlaisang, J., & Koraneekij, P. (2024). Virtual Studio Learning Environment Based on STEAM Education Concept Integrated With Socio-Scientific Issues to Enhance Learner's Scientific Creativity. *Journal of Information and Learning [JIL], 35*(2), Article 2.

Prince Torres, Á. C. (2022). El aprendizaje inmersivo como alternativa educativa en contextos de emergencia. Podium, 42, 19–38. https://doi.org/10.31095/podium.2022.42.2

Sánchez, Á., Haye, A., & Sebastián, C. (2022). Affective learning as a higher psychological process: A microgenetic analysis of evaluative conditioning from a sociocultural approach (Aprendizaje afectivo como proceso psicológico superior: análisis microgenético del condicionamiento evaluativo desde un enfoque sociocultural). *Studies in Psychology, 43*(2), 287–310. https://doi.org/10.1080/02109395.2021.1991132

Sánchez, N. M., Lorenzo, M. M. G., & Valdivia, Z. Z. G. (2009). Modelo para diseñar sistemas de enseñanza-aprendizaje inteligentes utilizando el razonamiento basado en casos. *Avances en Sistemas e Informática, 6*(3), 67–78.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46(*6), 606–620. https://doi.org/10.1037/0003-066X.46.6.606

Sommer, M. y Cabello, V. M. (2020) Andamios de retiro gradual. Parte 2: Apoyos a la construcción de explicaciones en ciencia primaria. *Estudios Pedagógicos, 46*(1), 269–284. http://dx.doi.org/10.4067/S0718-07052020000100269

Soroko,N. (2024). Features of organizing steam educational projects using immersive technologies. *Physical and Mathematical Education, 39*(2), 51–59. https://doi.org/10.31110/fmo2024.v39i2-07

Soto, C., & Natalia, M. (2023). Los ODS como elemento transversal y clave para el cambio hacia una educación de calidad. 1–184.

Soysal, Y. (2020). Investigating discursive functions and potential cognitive demands of teacher questioning in the science classroom. *Learning: Research and Practice, 6*(2), 167–194. https://doi.org/10.1080/23735082.2019.1575458

Tort, M. R. (2005). Las preguntas en el proceso de enseñanza- aprendizaje de las ciencias.

Tsankov, N. (2018). The transversal competence for problem-solving in cognitive learning. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE), 6*(3), 67–82.

Utomo D. P., & Santoso T. (2021). Zone of proximal development and scaffolding required by junior high school students in solving mathematical problems. *The Education and Science Journal, 23*(9): 186–202. https://doi.org/10.17853/1994-5639-2021-9-186-202

van Vondel, S., Steenbeek, H., van Dijk, M., & van Geert, P. (2017). Ask, don't tell; A complex dynamic systems approach to improving science education by focusing on the co-construction of scientific understanding. *Teaching and Teacher Education, 63*, 243–253. https://doi.org/10.1016/j.tate.2016.12.012

Wu, C. H., Tang, Y. M., Tsang, Y. P., & Chau, K. Y. (2021). Immersive Learning Design for Technology Education: A Soft Systems Methodology. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.745295

Zeegers, Y., & and Elliott, K. (2019). Who's asking the questions in classrooms? Exploring teacher practice and student engagement in generating engaging and intellectually challenging questions. *Pedagogies: An International Journal, 14*(1), 17–32. https://doi.org/10.1080/1554480X.2018.1537186