

Part 4 / Strand 4

Digital Resources For Science Teaching And Learning

Co-editors: *Nathalia Helena Azevedo & Leticia Garcia-Romano*

Part 4 / Strand 4 Digital Resources For Science Teaching And Learning

Design, evaluation and characterization of innovative resources, technological applications, digital tools, software, and environments for teaching/learning science, e.g., ICT and TEL in science education. Includes online learning environments, simulation, modelling tools, virtual and/or remote laboratories, game-based tools; self-regulation, reflection, and collaboration in digital learning environments; and implementation of machine learning, augmented reality, and artificial intelligence into teaching/learning science.

Sub-themes:

- 1) Integration of Augmented, Virtual, and Extended Reality in Science Education
- 2) Artificial Intelligence Applications in Science Education
- 3) Machine Learning Applications in Science Education
- 4) Development and Evaluation of Simulations and Digital Learning Environments
- 5) Game-based Learning Tools for Science Education

Contents

Strand 4: Digital Resources For Science Teaching And Learning.....	189
Examining The Generalisability Of Trained Computer Vision Models For Instructional Support In Science Experiments	192
Partners In Science Lesson Planning: GenAI-Driven Design With Human-In-The-Loop Validation	197
Game-Based Learning In Computer Science Lectures – A Study In Higher Education.....	205
Activating Epistemic Emotions To Enhance Student Engagement In Socio-Scientific Issues Through Digital Curation	213
A New Paradigm For Science Lesson Planning: Adaptive Expertise And Generative AI Integration.....	222
Conception Of Inclusive Learning Environments Related To Biodiversity And Geodiversity: An Innovative Project For Pre-Service Teachers	227
Bridging Digital And Physical Worlds: Leveraging AR And AI For Personalized Science Education In Geography And Astronomy	233
Talking To Or Talking Through Teachers - An Analysis Of A Digital Teacher's Guide	238
The Relationship Between Critical Thinking And Generative Artificial Intelligence: A Two-Way Street For Science Education.....	244
Exploring The Use Of ChatGPT In Science Lesson Planning	253
Artificial Intelligence In Teacher Education: Attitudes, Usage, And The Road To AI Literacy	263
Augmented Reality For Chemistry Education: Visualizing The Invisible In The Nitrogen Biogeochemical Cycle – A Pilot Study	274
Beware The Slider’s Range: A Design Lesson From Physics Simulation Development	282
What Factors Influence Pre-Service Science Teachers’ Ethical Use Of ChatGPT?	292
Generative AI Training For Pre-Service Teachers: Impacts And Adoption Patterns	297
Design-Based Research To Create A Collaborative Game To Foster Spatial Skills: The Role Of Teachers.....	303
Designing Digital STEM Teaching-Learning Sequences By Primary Pre-Service Teachers...	308
Digital Experimental Protocols And Literacy: Interventions For More Learner Support In Science Education	316
Potential Of ChatGPT As A Pedagogical Tool In Biology Education: Enhancing Student Engagement And Critical Thinking.....	322
How To Teach Science To 3-6-Year-Old Pupils: Pre-Service Teachers’ Perceptions Of Immersive Training With 360° Classroom Videos	329
Learning Through Comics And Students’ Stress In Physics: An EDA Approach.....	338
Building Bridges: AI Custom Chatbots As Mediators Between Mathematics And Physics	344

Strand 4: Digital Resources For Science Teaching And Learning

Nathália Helena Azevedo¹ and Leticia Garcia-Romano²

¹University of São Paulo, Brazil

²National University of Córdoba and CONICET, Argentina

Technologies In Dialogue Across Sections

This collection brings together 22 articles that explore the growing presence of digital technologies in science education research and practice. Although the studies differ in terms of educational level, disciplinary focus, methodological approach, and technological tools, they collectively reflect an ongoing effort to understand how digital technologies are shaping learning experiences, student engagement, and teacher education in science-related fields. Across the collection, technologies such as generative artificial intelligence, virtual and augmented reality, digital games, simulations, multimodal resources, and online instructional platforms appear as elements that reorganise classroom interactions, pedagogical practices, and forms of participation in science learning.

The recurring presence of artificial intelligence immediately stands out. In just a few years, AI-based tools have moved from technological curiosity to a dominant topic in conferences and journals in the field. Several papers here follow this movement by investigating the use of customised chatbots, ChatGPT, and generative AI for lesson planning and the preparation of future teachers. Still, perhaps the most interesting aspect of the collection is that the papers do not adopt a homogeneous stance toward these technologies. Rather than announcing radical transformations or replacing pedagogical discussions with technological enthusiasm, the studies seek to observe how teachers and students use these tools in teaching situations. In this sense, the papers help compose a picture of the field of experimentation, seeking to understand both the possibilities and the limitations of these new digital mediations. The collection can therefore be read through three broad thematic axes that cut across the different contributions.

Artificial Intelligence, Teacher Education, And Pedagogical Mediation

The first group brings together studies focused on the use of generative artificial intelligence in teaching and teacher education contexts. This includes investigations into customised chatbots in Mathematics and Physics education, inquiry-based activities mediated by ChatGPT in Biology classes, lesson planning with generative AI, teacher preparation, and discussions surrounding AI literacy. Despite the diversity of contexts, the papers converge in several ways. First, AI appears as a tool to support, reorganise, or expand certain pedagogical practices. The most interesting results emerge when there is some form of intentional teacher mediation, creating opportunities for students to move beyond rapid, instrumental uses of these tools toward more elaborated, exploratory forms of questioning once they receive guidance on how to interact with the technology.

Another recurring aspect concerns teacher education amid the rapid spread of these tools. Some studies suggest that future teachers express interest in and openness to AI use, while also expressing uncertainty about ethical issues, authorship, information reliability, and institutional boundaries surrounding its use. The papers reveal how AI reopens longstanding educational

discussions related to teacher autonomy, the role of pedagogical mediation, and the broader meanings of school learning.

Immersion, Engagement, And Affective Dimensions Of Learning

A second axis groups papers that explore immersive technologies and multimodal languages to enhance student engagement. This includes studies on 360° videos for teacher education, virtual reality, augmented reality, hybrid games, and digital curation in socio-scientific issues. What these studies share is a focus on students' experiences during the learning process. Their interest extends beyond conceptual gains to encompass motivation, stress, participation, immersion, and emotional involvement. The study on comics, for example, suggests that narrative and visual formats may help create less stressful learning environments for students. Meanwhile, studies on virtual reality and hybrid games investigate how more interactive, embodied experiences can bring students and teachers closer to inquiry-based STEM practices.

At the same time, the papers show that incorporating these technologies also entails concrete limitations and tensions. Participants report physical discomfort during immersive experiences, while other studies highlight the need for careful planning to prevent technology from becoming merely a source of distraction or spectacle. What emerges from this group of papers is an attempt to understand the conditions under which these tools may contribute to more meaningful learning experiences.

Digital Materials Design And The Organisation Of Teaching Practices

The third axis brings together studies focused on the design of digital materials, activities, and environments for science education. These include research on digital laboratory reports, STEM teaching sequences developed on digital platforms, physics simulations, digital learning objects, and digital teacher guides. These studies are connected by a shared concern with the organisation of pedagogical practices and the ways in which different digital formats influence learning. Some investigate issues related to cognitive load and visual design, while others analyse how pre-service teachers structure inquiry activities within digital environments.

Interestingly, several of these studies point to a recurring tendency: even in technologically innovative contexts, highly guided, linear approaches still predominate. Convergent activities, structured sequences, and more controlled forms of inquiry appear frequently in participants' materials. This does not diminish the relevance of these proposals, but it suggests that technological changes are not always accompanied by equivalent shifts in pedagogical conceptions.

Looking Ahead With Open Questions For Technology And Science Education

The papers in this collection portray a field undergoing rapid transformation while still marked by important continuities. Questions related to student engagement, teacher education, the organisation of school scientific inquiry, and pedagogical mediation remain central, even as new technologies increasingly occupy a visible place in research. Perhaps for this reason, the main contribution of this collection lies in documenting how researchers and teachers are attempting to negotiate the possibilities and limitations of these technologies in real teaching contexts.

At the same time, the collection also makes visible a series of open questions that may guide future research. Many studies demonstrate how digital technologies can support engagement, participation, immersion, or instructional design, but less attention is given to the broader institutional, political, and economic conditions shaping the adoption of these technologies in schools and teacher education programs. As digital platforms, generative AI systems, and data-

driven educational tools become increasingly embedded in educational systems worldwide, questions about dependence on private technological infrastructure, data governance, platformization, and commercial interests become difficult to ignore.

A more explicitly critical perspective on technology and science education still has significant room to develop within the field (Garcia-Romano et al., 2021). Future studies may increasingly need to ask whose interests are served by particular technological agendas, which forms of knowledge and participation become privileged, and which educational practices risk becoming marginalised. These questions become especially relevant in the context of the ongoing neoliberalization of education across different parts of the world, where efficiency, personalisation, scalability, and performativity often shape educational innovation agendas.

As schools and universities continue to incorporate AI systems, immersive environments, and digital learning platforms into everyday practices, science education research may benefit from approaches that also examine how technology reshapes relationships between teachers, students, institutions, knowledge production, and educational policy (Azevedo and Santos, 2025). In this sense, the studies gathered here may also be read as part of a larger conversation about the future of science education itself: a future in which technological change is likely to remain central, but where pedagogical, ethical, and political questions will continue to demand careful and critical attention.

References

- Azevedo, N. H., & Santos, P. G. F. D. (2025). The dominance of utilitarianism and the weakening of criticism: generative artificial intelligence and the values at stake in science. *Ensaio Pesquisa em Educação em Ciências (Belo Horizonte)*, 27, e59484. <https://doi.org/10.1590/1983-2117-59484T>
- Garcia Romano, L., Occelli, M., & Adúriz-Bravo, A. (2021). School Scientific Argumentation Enriched by Digital Technologies: Results With Pre- and in-Service Science Teachers. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(7), em1982. <https://doi.org/10.29333/ejmste/10990>

Examining The Generalisability Of Trained Computer Vision Models For Instructional Support In Science Experiments

Edwin Chng¹ and Bertrand Schneider²

¹National Institute of Education, Nanyang Technological University, Singapore

²Harvard Graduate School of Education, Harvard University,
United States of America

When science teachers facilitate conduct of student experiments, they face challenges in monitoring students' procedural steps due to presence of multiple students and limited cognitive bandwidth. Yet, without adequate teacher supervision, students' procedural missteps, incorrect experimental techniques, and safety violations become missed teacher observations that can result in the formation of undesirable student practices. To address this, computer vision can be deployed in science laboratories to continuously monitor students' procedural steps and highlight relevant student information for teachers to intervene. However, the generalisability of trained computer vision models for instructional support in science experiments is untested. To this end, we examined the performance of a trained computer vision model (SlowFast) on YouTube videos. The SlowFast model was trained on videos collected from 140 middle school students (which is of a specific population and context), and the gathered YouTube videos consists of 77 individuals from a wide variety of populations and contexts. Results indicate that model generalisability depends on choice of pretrained model and size of training dataset, together with recording conditions and data consistency of the new context. More specifically, precaution ought to be taken to ensure that collected training data contains at least hundreds of training instances, and that camera setup in all contexts are of similar viewpoints and zoom factors. Furthermore, variations in training instances should imitate expected variations in different contexts. Overall, this work attained preliminary evidence of the general feasibility of deploying trained computer vision models for instructional support in science experiments.

Keywords: Artificial Intelligence, Laboratory Instruction, Science Education

Introduction

When science teachers facilitate the conduct of student science experiments, they face challenges in monitoring students' procedural steps. This is because in a typical science laboratory setting, science teachers are tasked with attending to the learning needs of several students, and as humans, they experience natural cognitive limits in being able to pay sufficient attention to every action taken by each student (Sherin & Star, 2011). Yet, awareness of students' procedural steps is critical for science teachers to be able to provide students with suitable formative feedback on their experimental work (Stiggins, 2006). Without adequate teacher awareness, students' procedural missteps, incorrect experimental techniques, and safety violations become missed teacher observations that can result in the formation of undesirable student practices.

In this case, there exists a teacher observation gap that can be fulfilled with the use of artificial intelligence (AI). Specifically speaking, computer vision models can be deployed in science laboratories to continuously monitor students' procedural steps and highlight relevant student information for teachers to intervene (Chng, 2024). However, training of computer vision models is sensitive to data inputs (Gudivada et al., 2017) and the generalisability of trained computer vision models for instructional support in science experiments is untested. In particular, it is unknown if a computer vision model that is trained with a specific set of student data can perform well in the observation of another population in another context. With this in mind, this research

work seeks to examine the generalisability of trained computer vision models so as to establish the general feasibility of deploying AI for instructional support in science experiments.

Methods

Our research question asks, “What is the generalisability of a trained computer vision model in conducting student observations from another population and context?” As detecting students’ procedural steps is the primary concern for teachers during science experiments, we focus on the generalisability of action recognition models in this research.

Our chosen action recognition model is that of SlowFast (Feichtenhofer et al., 2019). In terms of model performance, SlowFast belongs to the latest generation of models that have demonstrated state-of-the-art capabilities for action recognition (Fan et al., 2021). In terms of model architecture, SlowFast is “partially inspired by biological studies on the retinal ganglion cells in the primate visual system” (Feichtenhofer et al., 2019, p. 6203) and mimics the way in which visual information is processed by humans. With these advantages in mind, we target SlowFast as the intended computer vision model for our generalisability test. A SlowFast model that was pretrained with the Kinetics400 dataset (Kay et al., 2017) was downloaded from the PyTorchVideo library (Fan et al., 2021). The motivation for using a pretrained model is due to the fact that such models contain primitive understandings of lower-level action features (Sharif Razavian et al., 2014), which renders them easier to train and adapt for our purposes (Tajbakhsh et al., 2016).

Table 1. Statistics of video clips in our video datasets.

Action ID	Student Action	Number of Student Video Clips	Number of YouTube Video Clips	Action Category
1	close air hole	117	55	Light Bunsen burner
2	close gas tap	166	74	
3	ignite Bunsen burner	367	84	
4	open air hole	253	60	
5	open gas tap	220	72	

Our target set of student actions is related to the lighting of a Bunsen burner (see Table 1 for detailed list of actions). This set of student actions was chosen because lighting Bunsen burner is a common procedure in science experiments. To prepare our downloaded SlowFast model for test of generalisability, we first train it with videos collected from 140 middle school students as they light a Bunsen burner. With the trained model, we then evaluate its performance on YouTube videos of 77 individuals lighting a Bunsen burner. YouTube was selected as a source for our generalisability test since it contains videos of individuals from a wide variety of populations and contexts. In other words, we can address our research question by examining the YouTube performance of a SlowFast model that was trained on data collected from a middle school (which is of a specific population and context, as compared with YouTube videos which is of a wide variety of populations and contexts). Findings from this investigation will ultimately apprise us

of the need to retrain action recognition models for use in other science experiments and provide us with an appreciation of the general feasibility of deploying AI for instructional support in science experiments.

Table 1 shows the statistics of video clips in each of our video datasets (i.e., student dataset and YouTube dataset). Each video clip represents an uninterrupted instance of a particular action taken by the student/individual within the video. Since each action can possibly be taken by the student/individual more than once or none at all, the number of video clips obtained for each action deviates from the total number of videos collected in each dataset.

Results

Table 2 shows the class prediction accuracies of our trained SlowFast model on YouTube videos. Surveying Table 2, our model’s class prediction accuracies are consistently above 70% and has an overall accuracy of 76.81%. Compared with Slowfast’s reported state-of-the-art accuracy of 76.94% (Fan et al., 2021), our model’s performance on the unseen YouTube dataset with different populations and contexts is respectable. This is possibly because our trained SlowFast model reaped the performance benefits of being pretrained with the Kinetics400 dataset to learn lower-level action features and being further trained with a sizeable student dataset to learn our target set of student actions. It is noteworthy here that the reported state-of-the-art accuracy was attained with a dataset that contains at least 400 video clips for each action class (Fan et al., 2021). By contrast, our smallest action class (i.e., “ID 1 close air hole”) only possessed 117 video clips and our largest action class (i.e., “ID 3 ignite Bunsen burner”) contains no more than 400 video clips. Yet, our trained model is able to perform on unseen data with a class prediction accuracy of 80.00% for our smallest action class and achieve an overall model accuracy of 76.81%. This signals the importance of transfer learning for AI models and hints at the general feasibility of deploying trained AI models for use in other science experiments. On balance, our test on data from the wild (i.e., YouTube videos) demonstrates the generalisability of trained computer vision models as long as the choice of pretrained model is appropriate and the size of training dataset is sufficient.

Table 2: Class prediction accuracies of our trained SlowFast model on YouTube videos.

Action ID	Student Action	Class prediction accuracies
1	close air hole	80.00%
2	close gas tap	78.38%
3	ignite Bunsen burner	83.33%
4	open air hole	70.00%
5	open gas tap	70.83%
-	Overall	76.81%

In an effort to better understand model generalisability, I further investigated examples of incorrect model predictions qualitatively. In this respect, two inherent sources of model confusion were discovered. First, even though the camera viewpoint and zoom factor of our collected student videos is fixed, those of YouTube videos can be changing. For instance, some YouTube videos provide a shifting perspective as the Bunsen burner is being lit (e.g., from front to side). Otherwise, some YouTube videos zoomed in on the gas tap when it is being opened or closed and later zoomed out to reobtain a full view. By contrast, all of our student videos adopted an all-

encompassing frontal viewpoint with unvarying zoom. Second, the apparatus used in school science laboratories may differ from that of YouTube videos. For example, gas lighters appeared in all our student videos, but certain YouTube videos featured spark lighters. Altogether, these differences in recording conditions and apparatus used resulted in more incorrect model predictions when contexts are switched, which serves to remind us that model generalisability is dependent on recording conditions and data consistency as well.

Discussion

In contemplating the widespread deployment of trained AI models, we think about the implications of our investigation findings and deliberate on the integration of computer vision models for instructional support in science experiments. From our model investigation, it became evident that model generalisability depends on the choice of pretrained model and the size of training dataset, together with the recording conditions and data consistency of the new context. Consequently, we should take precaution to ensure that collected training data contains at least hundreds of training instances for each action class, and that camera setup in all contexts are of similar viewpoints and zoom factors. Furthermore, as with common machine learning approaches, variations in training instances should imitate expected variations in different contexts. For example, if both spark and gas lighters are expected to be found within deployment contexts, then gathered training instances ought to contain ignition of Bunsen burner with both spark and gas lighters. In this way, widespread deployment of AI models in general science experiments can be achieved without concerns for reduced performance in new contexts.

As for the integration of computer vision models for instructional support in science experiments, we note that teachers face severe time constraints in schools and would not possess sufficient bandwidth to go through model outputs for every student (Chng, 2024). Under these circumstances, the delivery of instructional support should include added processing steps to consolidate pertinent information for teachers. More specifically, we can apply auxiliary algorithms to extract needed student information from detected student actions. For instance, if teachers wish to track the order of procedural steps taken by students for formative feedback, then we can rely on sequence analysis (Mannila et al., 1997) to ascertain order of student actions and highlight to teachers whenever students deviate from expected sequences. In essence, the integration of computer vision models for instructional support should take into consideration the realities of teacher circumstances and pedagogical concerns so as to promote widespread deployment of AI models in general science experiments.

Finally, although we have attained preliminary evidence of model generalisability in this investigation, more testing is required to establish the parameters and boundaries of general applicability. Specifically, more research work is needed to uncover the nuanced performance of trained AI models across different experiments, for diverse students, and in varied environments.

References

- Chng, E. (2024). Designing Computer Vision Support for Science Practical Work: A Qualitative Investigation into the Noticing Practices and Support Preferences of Science Teachers. *Journal of Science Education and Technology*, 33(5), 718-728. <https://doi.org/10.1007/s10956-024-10116-w>.
- Fan, H., Murrell, T., Wang, H., Alwala, K. V., Li, Y., Li, Y., ... & Feichtenhofer, C. (2021). PyTorchVideo: A deep learning library for video understanding. In *Proceedings of the 29th ACM international conference on multimedia* (pp. 3783-3786).
- Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6202-6211).
- Gudivada, V., Apon, A., & Ding, J. (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software*, 10(1), 1-20.

- Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., ... & Zisserman, A. (2017). The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*.
- Mannila, H., Toivonen, H., and Verkamo, A. I. (1997). Discovery of frequent episodes in event sequences. *Data mining and knowledge discovery*, 1(3):259–289.
- Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 806-813).
- Sherin, B., & Star, J. R. (2011). Reflections on the study of teacher noticing. *Mathematics teacher noticing: Seeing through teachers' eyes*, 66.
- Stiggins, R. (2006). Assessment for learning: A key to motivation and achievement. *EDGE: the latest information for the education practitioner*, 2(2), 1-19.
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning?. *IEEE transactions on medical imaging*, 35(5), 1299-1312.

Partners In Science Lesson Planning: GenAI-Driven Design With Human-In-The-Loop Validation

*Kalliopi Meli*¹, *Yaron Lehavi*², *Ioanna Taouki*³ and *Dimitrios Pantazatos*⁴

¹University of Patras, Greece

²David Yellin Academic College of Education, Israel
3100mentors, Greece

²National Technological University of Athens, Greece

Generative Artificial Intelligence (GenAI) has emerged as a disruptive technology with vast potential in education. A workshop that took place during the ESERA 2025 conference explored the intersection of GenAI and science education, focusing on how educators utilize GenAI tools to develop lesson plans that consider students' common alternative frameworks. We investigated the awareness and adoption levels of GenAI among science educators, explored how researchers and practitioners use GenAI to construct science lessons, and assessed how collaboration with other experts in science education and research can validate and enhance instructional design developed with GenAI. During the workshop, participants engaged in hands-on activities using a customized GenAI tool, allowing them to design instructional material tailored to their specific educational needs. This session emphasized the importance of integrating scientific accuracy and pedagogical principles in exercise design, and discussions centred on best practices for leveraging GenAI in educational contexts while addressing potential concerns and limitations. The workshop equipped participants with a comprehensive understanding of GenAI's potential in science education and served as a stepping stone towards developing more effective, personalized, and engaging teaching environments.

Keywords: Generative Artificial Intelligence; Human-in-the-loop; Students' alternative frameworks

Introduction

Generative Artificial Intelligence (GenAI) has emerged as a disruptive technology with vast applications across various domains, including education. GenAI offers opportunities to create more engaging, personalized, and effective learning environments (Meli et al., 2024). From the educator's perspective, integrating GenAI streamlines instructional material development, reduces instructor workload, and enhances material quality (Dickey & Bejarano, 2024).

GenAI can potentially assist educators in generating instructional materials and supplementary resources tailored to the curriculum and individual student needs (Bonde, 2024). While previous research has explored various applications of GenAI in educational settings, there remains a need for further investigation into its potential for supporting instructional design in a structured and efficient manner (Karpouzis et al., 2024). The integration of GenAI into science education may present a significant opportunity to transform lesson planning, not only by enhancing its feasibility and robustness for educators but also by increasing educators' awareness of students' learning requirements within the context of the scientific content they must acquire (Almasri, 2024).

While educators may acknowledge that GenAI potentially facilitates and enhances their work, concerns persist regarding the limitations of this emerging technology, particularly in disciplines where the model's training data are insufficient. To mitigate these risks, it is crucial to maintain a "human-in-the-loop" approach, which generally refers to the need for human interaction, intervention, and judgment to control or change the outcome of a process (Meng, 2023, p.1). This

ensures that critical decisions are not left solely to automated systems and allows for the detection and correction of potential mistakes or unintended consequences and the improvement of the overall outcome quality (Chen et al., 2023).

Addressing these multifaceted opportunities and challenges, the workshop held during the ESERA 2025 conference aimed to transcend individual trials and provide structured guidance and support to assist educators in enhancing their awareness of GenAI capabilities (Bonde, 2024). Specifically, it sought to facilitate the utilization of GenAI for context-specific needs, particularly for developing lesson plans that incorporate relevant pedagogical aspects. Additionally, the workshop intended to expose participants to the "human-in-the-loop" methodology, where the human is an expert in teaching a specific scientific topic, to validate and enhance the results they co-created with GenAI.

Theoretical framework

Addressing students' alternative frameworks is essential for designing effective science education. Educators are expected to engage students in activities that challenge their ideas and promote conceptual change toward a scientifically accurate understanding of the world (Vosniadou, 2011). This aligns with the emphasis on integrating science practices and developing sophisticated teaching methods to support student learning (Arias et al., 2016).

However, traditional textbooks and guidelines often focus more on scientific content, rather than on the alternative frameworks students may employ in their interpretations (Zourmpakis et al., 2023). Consequently, embracing more progressive pedagogical approaches challenges science educators at an individual level as they frequently encounter diverse content and varying levels of complexity that are often related to a broad spectrum of alternative frameworks (Johsua & Dupin, 1991).

In this respect, GenAI shows the potential to accelerate educators' lifelong learning progression and effectively support them in implementing their day-to-day educational duties (Pantazatos et al., 2024). GenAI tools introduce new methodologies for effective teaching and learning, addressing traditional educational challenges. This technology can empower educators to develop dynamic teaching materials, fostering innovation in educational practices and facilitating ongoing learning and skill development. Adopting GenAI in education can transform instructional design by analysing and expanding subject-specific content and creating learning pathways informed by educational research (Meli et al., 2024).

Integrating GenAI into science education presents a promising avenue for enhancing teaching methodologies and addressing students' alternative frameworks. GenAI can simulate complex physical phenomena and provide visual and interactive learning experiences, which are often challenging to replicate in traditional classroom settings. By leveraging GenAI, educators can personalize their teaching by streamlining their workflow, creating more efficient lesson plans, generating diverse problem sets, and developing tailored explanations for complex concepts. This technology can assist in identifying and targeting common areas of alternative conceptions among students, allowing for more effective instruction (Cooper, 2023). In teaching planning, GenAI can direct educators to consider established knowledge in the discipline they teach and in the discipline of teaching it. This can enhance their awareness of using their own PCK in planning and conducting their teaching.

Concurrently, implementing GenAI in educational contexts presents several significant challenges (Alier et al., 2024). The efficacy of the models underlying GenAI depends heavily on the quality of user-provided prompts. The caliber of responses can exhibit considerable variability

based on the application domain, potentially resulting in inconsistent or inadequate information in certain areas. The underlying models can generate content that appears authoritative but may be entirely fabricated or irrelevant to the query, which can be misleading in educational settings. The responses may perpetuate biases and reinforce stereotypes present in the training data. Finally, from a personal development aspect, there exists a risk of diminished creativity and critical thinking among users due to excessive reliance on the outputs that GenAI can readily produce. While the respective tools can assist, they cannot supplant the value of human interaction, crucial for users' development.

Mitigating the risks of GenAI in education requires a multifaceted approach that addresses ethical, pedagogical, and technological concerns. Developing critical GenAI literacy among educators is crucial for responsible innovation and informed use of GenAI (Rudolph et al., 2024). This involves fostering an environment where stakeholders can understand and critically evaluate GenAI-generated content. Maintaining human oversight is essential to address potential privacy, bias, and accuracy issues in GenAI tools (Noroozi et al., 2024). Human-in-the-loop approaches, informed by experts' input, can solve problems related to data scarcity, feedback accuracy, annotation costs, and decision-making reliability in educational material development. Combining human expertise with artificial intelligence capabilities offers promising solutions for creating more effective and efficient educational tools and experiences (Chen et al., 2023).

To complement these theoretical considerations, further inquiry is needed into how researchers or practitioners and institutions can critically and productively integrate GenAI into science instruction. Several open questions emerge as particularly relevant to ongoing research and practice:

1. What would effective integration of GenAI tools into teaching planning and delivery look like, and what improvements might it bring to instructional quality and educator workload?
2. How can GenAI be leveraged to identify and address students' alternative frameworks, or other considerations embedded in science education, in supporting deeper conceptual understanding?
3. What strategies can be employed to combine human expertise and GenAI to ensure that instructional materials are both pedagogically sound and scientifically accurate?
4. How might additional human input—through peer collaboration, student feedback, or expert review—be used to maintain the quality and integrity of GenAI-generated content in educational contexts?

These questions not only reflect the pedagogical tensions and opportunities surrounding GenAI but also help frame future empirical studies and practical design considerations in science education.

Method

To explore and exemplify a means to address the questions above, we organized a workshop as part of the ESERA 2025 conference. The workshop engaged the participants with GenAI tools for lesson planning, while emphasizing educator agency, peer collaboration, and reflection. In essence, the workshop served as a pilot implementation of an educator-centred GenAI approach in science education.

Participants And Context

The workshop brought together higher education science researchers and practitioners ($N=25$) from diverse backgrounds, most of whom were curious about using GenAI but had limited prior experience. An initial poll confirmed that less than half had experimented with tools like ChatGPT, roughly mirroring the split found in larger surveys (e.g., Pantazatos et al., 2024) on GenAI adoption. This mix provided a rich environment for peer learning. The focus on science education (as opposed to general teaching) allowed us to tailor scenarios to common science topics and students' alternative frameworks, making the experience directly relevant to participants' real-life contexts.

Workshop Design

We structured the 45-minute session into four interactive phases aligned with the human-in-the-loop design: orienting and framing GenAI use, hands-on experimentation, collaborative expert review, and reflective synthesis.

Orientation And Critical Framing

The workshop began with a brief introduction to GenAI's capabilities and limitations, using current research and classroom examples to provoke reflection. We demonstrated GenAI's potential to simulate scientific phenomena and its tendency to state confidently scientific errors. Framing was informed by literature highlighting GenAI's epistemic risks (e.g., Cooper, 2023; Rudolph et al., 2024), setting a tone of curiosity paired with critical literacy. Participants shared their hopes and concerns in an open dialogue: many were enthusiastic about reducing lesson-planning load, while others voiced concerns about delegating too much conceptual work to GenAI or inadvertently circulating misconceptions. This ideas exchange helped establish the workshop's central tension: how to meaningfully integrate GenAI without surrendering pedagogical judgment.

Hands-On GenAI Tool Use

Participants then logged into a fine-tuned GenAI tool interface developed by our team. The tool was inspired by prompt-engineering systems like that of Karpouzis et al. (2024), and invited users to input topics, student level, known alternative frameworks, and teaching preferences. Based on this, the tool generated a structured lesson plan, typically including: (1) learning objectives, (2) an introductory warm-up targeting a common alternative framework, (3) an interactive activity, and (4) an assessment item. To scaffold engagement, we pre-loaded one editable prompt that served as a model:

Design a 60-minute lesson plan for first-year university students focusing on the First Law of Thermodynamics, particularly energy change and energy conservation.

Address common alternative frameworks such as the idea that energy is 'used up' or disappears during a process. The lesson should support students in developing a correct understanding of energy conservation through both conceptual and quantitative reasoning.

Include: (1) a context-rich example involving heat and work (e.g., gas expansion, heating a metal block), (2) a guided problem-solving activity, and (3) a checkpoint for discussion where students compare their initial ideas with formal thermodynamic reasoning.

Conclude with a summary that reinforces the conservation principle across system boundaries.

Participants worked in pairs to simulate collaborative use, actively revised outputs, applying their pedagogical knowledge to fill in gaps or correct inaccuracies, demonstrating the central role of

educator judgment in human-in-the-loop design. As they interacted with the tool, facilitators supported prompt refinement (e.g., adding context, specifying alternative frameworks). Participants were reminded that GenAI outputs were first drafts to be evaluated critically. Educators noted several emergent patterns: some plans were praised for creative analogies (e.g., comparing chlorophyll to a “solar panel”), while others lacked experimental rigor (e.g., omitting control variables in suggested investigations).

Human-in-the-Loop Validation Exercise

Following initial design work, groups submitted a selected lesson plan for live review by a science education researcher (co-facilitator), modelling expert oversight. The reviewer projected excerpts from a participant-generated plan and engaged in a public “think-aloud,” identifying both strengths and areas of concern. For instance, a proposed analogy about current flow was deemed accessible but conceptually imprecise; the expert revised the phrasing on the spot. Positive design features were also highlighted, such as an GenAI-suggested student role-play where learners acted as parts of a circuit. This segment gave participants a concrete example of what human-GenAI co-design might look like in situ, reinforcing the necessity of content validation. Feedback from attendees suggested this was one of the most impactful workshop moments, providing a mental model for how educators might vet GenAI-generated materials in their own practice.

Outcomes

This workshop provided participants with a structured, immersive experience at the intersection of GenAI and science education. By engaging in hands-on activities, collaborative design, and guided expert review, participants explored how GenAI can support lesson planning—particularly in addressing students’ alternative conceptions. The implementation foregrounded human agency and critical engagement with GenAI, aiming not merely to showcase technological capabilities but to situate them within authentic pedagogical challenges (Zourmpakis et al., 2023; Chen et al., 2023).

By the end of the session, participants reported feeling more confident in experimenting with GenAI tools for educational purposes. Many articulated a shift in mindset, from skepticism or apprehension to a more empowered and strategic view of GenAI use. As one participant remarked, “I was worried AI would take over, but now I see it’s up to me to direct it. It’s like a super-smart apprentice – helpful, but it still needs supervision.” This comment reflects a broader theme that emerged across reflections: GenAI, when used thoughtfully, can extend educator capacity without displacing professional judgment.

Participants left the workshop with concrete lesson plan drafts (customized to their classroom context and structured using the GenAI tool) that could be further refined for actual use. Several noted that the tool had helped them generate ideas they might not have considered otherwise, affirming literature on GenAI’s role in ideation support and workload reduction (e.g., Pantazatos et al., 2024). Yet this creative potential was matched by a shared recognition of GenAI’s limitations. All participants identified at least one factual error, omission, or pedagogical mismatch in the GenAI output, reinforcing the necessity of educator oversight. This outcome was viewed positively: the experience functioned as a form of inoculation against uncritical use of GenAI, strengthening participants’ critical GenAI literacy.

The workshop also illuminated the dynamics of educator agency and collaboration. Educators who initially expressed low confidence in using GenAI tools showed greater enthusiasm and initiative when paired with peers. This suggests that co-use scenarios not only improve outcomes but also reduce intimidation, highlighting peer collaboration as a key factor in building collective

confidence and competence. Furthermore, the visible role of the invited science education researcher (as a constructive and approachable “critical friend”) helped normalize the idea that integrating GenAI into practice is a shared, iterative process. Participants responded positively to this modelling of expert feedback, indicating openness to seeking similar validation mechanisms in their own contexts.

In sum, the workshop outcomes support the viability of a teacher-centred, human-in-the-loop approach to GenAI in science education. Participants valued the time-saving and creative affordances of GenAI, while also deepening their appreciation for the irreplaceable role of human judgment. These insights not only confirm existing hypotheses in the literature but also offer direction for future implementations, particularly the need to scaffold peer collaboration, embed expert review, and foster critical literacy as core components of any GenAI-supported instructional design process.

Discussion

The findings of this workshop-based study offer valuable insights into the evolving role of GenAI in science education, particularly when implemented through a human-in-the-loop, teacher-centred design. As the results illustrate, educators engaged with GenAI not merely as consumers of pre-generated content but as critical co-designers, actively shaping and refining outputs in response to pedagogical needs. This aligns with existing calls for educators to retain epistemic authority in the classroom, even in the presence of increasingly sophisticated AI tools (Cooper, 2023).

The varied responses to GenAI-generated lesson plans underscore a central theme in the literature: that while GenAI offers substantial affordances (such as idea generation, time efficiency, and personalized content creation) it also demands rigorous human oversight to mitigate the risks of inaccuracy, superficiality, and bias (Alier et al., 2024; Noroozi et al., 2024). Participants’ recognition of these risks and their capacity to critique and improve the GenAI outputs reflect the growing importance of critical GenAI literacy in educator professional development (Rudolph et al., 2024). Rather than viewing GenAI as a threat to educator autonomy, many participants reframed it as a tool requiring pedagogical steering—consistent with the “super-smart apprentice” metaphor voiced during the session.

Notably, the workshop demonstrated how GenAI might be harnessed to better address students’ alternative frameworks in science. By prompting the GenAI to design warm-ups and activities that target known alternative frameworks, participants practiced a generative form of lesson planning aligned with constructivist goals and conceptual change theory (Vosniadou, 2011; Arias et al., 2016). This suggests that when GenAI is guided by well-crafted prompts and supported by expert or peer knowledge, it can serve not only to expedite instructional design but also to embed more responsive and diagnostic pedagogical strategies.

In this regard, the findings also validate the theoretical potential of GenAI as a tool for accelerating educators’ professional growth (Pantazatos et al., 2024; Meli et al., 2024). Participants’ experiences echo research suggesting that GenAI-powered platforms can scaffold educators’ reflection and creativity, provided they are used in contexts that value human interpretation over automation. The design of the workshop—emphasizing peer collaboration and expert validation—resonates with broader literature advocating for multi-layered human input in educational GenAI systems (Chen et al., 2023). The observed benefits of pair work among less confident participants further emphasize the need to structure GenAI training within socially

supported collaborative environments that promote collective educator agency.

At the same time, the workshop exposed important design tensions. For instance, the variability of GenAI outputs highlighted the ongoing challenge of prompt precision (Alier et al., 2024), and participants' mixed trust in GenAI-generated facts mirrored broader concerns about misinformation and fabricated content. This reinforces the notion that GenAI's value lies not in its autonomy but in its ability to extend human thinking, provided researchers and practitioners are equipped to filter, contextualize, and critique its suggestions.

Ultimately, this study contributes to a growing body of evidence indicating that successful integration of GenAI in education requires more than access to tools—it requires frameworks for thoughtful, ethical, and collaborative use. As such, it affirms that teacher-centred, critically reflective approaches are not only compatible with GenAI but essential for its responsible and pedagogically sound implementation.

Conclusions

This study contributes to the emerging landscape of GenAI in science education by demonstrating how teacher-centred, human-in-the-loop approaches can enable thoughtful integration of AI into lesson planning. The workshop highlighted that while GenAI offers clear benefits, such as efficiency, idea generation, and flexible adaptation, it must be embedded within pedagogical frameworks that prioritize human judgment, contextual relevance, and critical oversight.

Participants' experiences affirmed that educators are eager to explore GenAI, but they require structured opportunities to build confidence, develop prompt literacy, and collaboratively review outputs. These findings support a vision of GenAI not as a replacement for researcher or practitioner expertise, but as a partner in pedagogical design; one that, when properly guided, can enhance science instruction and help address persistent challenges such as student alternative frameworks.

Looking ahead, future work should explore longitudinal implementations of teacher-GenAI co-design, particularly across varied school contexts and subject areas. Developing professional development models that promote peer collaboration, expert feedback, and iterative refinement of GenAI-generated content will be key. Additionally, more research is needed into how GenAI can support formative assessment, differentiation, and inclusive pedagogy. As GenAI tools continue to evolve, the education community must remain committed to evaluating their impacts, not only on efficiency and design quality, but also on educator identity, student understanding, and the broader ethics of AI-enhanced teaching.

Acknowledgement

The authors would like to thank all the participants of the ESERA 2025 workshop for their enthusiastic engagement and thoughtful contributions.

References

- Alier, M., García-Peñalvo, F.-J., & Camba, J. D. (2024). Generative Artificial Intelligence in Education: From Deceptive to Disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 5. <https://doi.org/10.9781/ijimai.2024.02.011>
- Almasri, F. (2024). Exploring the Impact of Artificial Intelligence in Teaching and Learning of Science: A Systematic Review of Empirical Research. *Research in Science Education*, 54(5), 977–997. <https://doi.org/10.1007/s11165-024-10176-3>
- Arias, A. M., Davis, E. A., Marino, J.-C., Kademian, S. M., & Palincsar, A. S. (2016). Teachers' use of educative curriculum materials to engage students in science practices. *International Journal of Science Education*, 38(9), 1504–1526. <https://doi.org/10.1080/09500693.2016.1198059>

- Dickey, E., & Bejarano, A. (2024). GAIDE: A framework for using generative AI to assist in course content development. In 2024 IEEE Frontiers in Education Conference (FIE) (pp. 1-9). IEEE. <https://doi.org/10.1109/FIE61694.2024.10893132>
- Bonde, L. (2024). A Conceptual Design of a Generative Artificial Intelligence System for Education. *International Journal of Research and Innovation in Applied Science*, 9(4), 457–469. <https://doi.org/10.51584/ijrias.2024.904034>
- Chen, S., Liu, Q., & He, B. (2023). A generative AI-based teaching material system using a human-in-the-loop model. 2023 International Conference on Intelligent Education and Intelligent Research (IEIR), 1-8. <https://doi.org/10.1109/IEIR59294.2023.10391244>
- Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, 32(3), 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- Johsua, S., & Dupin, J. (1991). In physics class, exercises can also cause problems . . . *International Journal of Science Education*, 13(3), 291–301. <https://doi.org/10.1080/0950069910130307>
- Karpouzis, K., Pantazatos, D., Taouki, J., & Meli, K. (2024). Tailoring education with GenAI: A new horizon in lesson planning. 2024 IEEE Global Engineering Education Conference (EDUCON), 1-10. <https://doi.org/10.1109/EDUCON60312.2024.10578690>
- Meli, K., Taouki, J., & Pantazatos, D. (2024). EMPOWERING EDUCATORS WITH GENERATIVE AI: THE GENAI EDUCATION FRONTIER INITIATIVE. 1, 4289–4299. <https://doi.org/10.21125/edulearn.2024.1077>
- Meng, X.-L. (2023). Data Science and Engineering With Human in the Loop, Behind the Loop, and Above the Loop. *Harvard Data Science Review*, 5(2). <https://doi.org/10.1162/99608f92.68a012eb>
- Noroozi, O., Farrokhnia, M., Banihashem, S. K., & Soleimani, S. (2024). Generative AI in Education: Pedagogical, Theoretical, and Methodological Perspectives. *International Journal of Technology in Education*, 7(3), 373–385. <https://doi.org/10.46328/ijte.845>
- Pantazatos, D., Taouki, J., & Meli, K. (2024). GENERATIVE AI THROUGH THE TEACHER’S LENS: PERSPECTIVES ON ADOPTION IN SCHOOL EDUCATION. 1, 4173–4181. <https://doi.org/10.21125/edulearn.2024.1051>
- Rudolph, J., Mohamed Ismail, F. M., & Popenici, S. (2024). Higher Education’s Generative Artificial Intelligence Paradox: The Meaning of Chatbot Mania. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/54fs5e77>
- Vosniadou, S. (2011). Reframing the Classical Approach to Conceptual Change: Preconceptions, Misconceptions and Synthetic Models (pp. 119–130). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9041-7_10
- Zourmpakis, A.-I., Kalogiannakis, M., & Papadakis, S. (2023). A Review of the Literature for Designing and Developing a Framework for Adaptive Gamification in Physics Education (pp. 5–26). Aip Publishing Llc melville New York. https://doi.org/10.1063/9780735425712_005

Game-Based Learning In Computer Science Lectures – A Study In Higher Education

Melissa Gruber

University of Education Karlsruhe, Germany

In higher education, incentivizing approaches are shifting from traditional teaching to constructivist, self-determined learning methods. The latter allow students to study at their own pace and focus on individual needs. One such approach is game-based learning, which is understood as the use of games for educational purposes. Based on predetermined learning objectives, such serious games are specifically designed to foster learning by playing educational games. Considering digital transformation of learning materials, serious games are often realized as digital games. These are considered particularly suitable for students with prevalent engagement with computers and high understanding of digital tools, such as computer science students. Within the scope of an elective subject at the faculty for Business Informatics at the University of Applied Sciences Karlsruhe, Germany, digital serious games were introduced as an integral part of the lectures. Over the course of 3 semesters, these educational games were evaluated by students. In total, n=87 higher education computer science students participated in the study. Results shed light on students generally expressing enthusiasm as well as appreciation for educational games. Students respond positively to the adoption of innovative game-based teaching methodologies. There is a notable appreciation for the replacement of traditional unidirectional teaching methods. In addition to providing positive feedback, students offer suggestions for the enhancement of digital serious games regarding their utilization, incorporation into lectures, and game design. Students indicate that digital serious games are enjoyable and anticipate their incorporation into future curricula. This article delves into the survey results and provides recommendations for further application of game-based learning in higher education.

Keywords: Constructivism, Digital Learning, Student Engagement

Introduction

In the contemporary era, a decline in the attention span of students has been observed. This leads to educators seeking innovative methods to gain students' attention and interest in lectures. The integration of incentivizing approaches into higher education lectures has gained significant traction as to enhance student engagement, motivation, and attention (Gruber, 2025). Hence, game-based learning has evolved as a popular method for interactive learning (Bencsik et al., 2021).

Game-based learning is an established approach in higher education, especially in science, technology, engineering and mathematics (STEM) disciplines (Semmler et al., 2024). The problem-solving approach that is taught and applied in STEM disciplines can be aligned with solving problems in educational games (Ishak et al., 2021; Manzano-León, Camacho-Lazarraga, Guerrero, et al., 2021). Platforms like genially and Kahoot! have emerged as popular tools for educational games, offering interactive experiences that can transform traditional educational paradigms (Martinez et al., 2023).

Even though there are many studies on the effect of digital serious games on student motivation (Bröker et al., 2021; Ehrlich et al., 2020; O'Brien & Pitera, 2019), computer science students' perception of quiz-based digital games have yet to be investigated. Therefore, this study showcases computer science students' perception of digital educational games based on the

aforementioned platforms in higher education settings, drawing insights from survey-based studies and evidence-based research.

Background

Constructivist learning theory focusses on learners actively constructing their own understanding and knowledge of the world through experiences and reflection (Kergel & Heidkamp-Kergel, 2020). Rooted in the work of theorists like Jean Piaget and Lev Vygotsky, this theory posits that learning is a process of meaning-making rather than passive absorption of information. In other words, constructivist learning theory favours active learning and refrains from the concept of teaching. Constructivism advocates for learner-centred approaches, where students engage in problem-solving, exploration, and critical thinking, often through collaborative activities. This approach highlights the importance of prior knowledge, social interaction, and contextual learning in shaping new understandings (Vygotsky et al., 1980). Teachers act as facilitators, guiding learners to build connections between new and existing knowledge, fostering deeper comprehension and skill development (Franco & DeLuca, 2019).

Following the idea of students constructing their own knowledge and taking an active part in the learning process, constructivist learning environments focus on presenting information. The information is based on the learning objective and part of the learning content. Students process the information provided individually and reconstruct their own knowledge. In the next step, students show a certain behaviour applying their knowledge. Teachers take a supporting role in diagnosing competences as well as potential weaknesses. Monitoring students' behaviour and success, teachers modify the information and learning support offered. At this stage, the cycle of constructivist learning environment closes at the stage of information presentation (Kerres, 2024).

In accordance with the theory of constructivism, constructivist learning environments enable students to self-regulate their learning process. Working on complex and lifelike problems with minimal intervention by teachers is key (Gräsel & Gniewosz, 2011). This approach is particularly well suited to digital learning methods. As they offer students the advantage of being able to work through and access digital learning resources at their own individual pace, they support constructivist learning (van Ackeren et al., 2017). Among many digital learning methods, digital learning games are evolving as popular constructivist learning approach.

Digital game-based learning refers to the use of digital games to facilitate learning. This approach makes use of entire games, specifically designed for educational purposes addressing learning objectives. These games are referred to as serious games (Weselek, 2022). In contrast to game-based learning and serious games, gamification is defined as the use of single game elements (Deterding et al., 2011). Gamification elements and serious games can be used in a physical and in a digital environment (Gruber & Faßbender, 2025).

Regarding the conceptualization of serious games, it is of importance to consider the target group. In higher education, students often face difficulties concentrating on lecture materials (Bencsik et al., 2021). This necessitates a comprehensive understanding of the students' existing knowledge and skills. Especially the learning process needs to be considered. As digital serious games operate based on different stimuli, the internal learning process and attention span can differ from learning in traditional settings (Frey & Uemminghaus, 2021). Selective perception can cause the differentiation of how students perceive digital settings (Gagné, 1980). Therefore, it is important to communicate expectations and the purpose of digital serious games to the students.

In higher education contexts, digital serious games aim to create immersive and interactive environments that promote active learning, addressing individual learning paths by means of constructivist learning environments (Frey & Uemminghaus, 2021). The constructivist framework is particularly relevant in digital serious games, as interactive games create helpful environments for experimentation, feedback, and active learning (Kerres, 2024). Digital serious games are particularly promising for students with a high affinity for gaming in their leisure time. It is evident that this is the case for STEM programs (Fleuren, 2016; Ishak et al., 2021; Manzano-León, Camacho-Lazarraga, Guerrero, et al., 2021). Indications in the literature suggest that students' perception of and their satisfaction with the lecture can be affected by digital serious games (Priesto et al., 2021). In general, previous studies show that STEM audiences are suitable for the use of digital serious games (Buchner et al., 2022; Makri et al., 2021; Taraldsen et al., 2022).

Prominent examples for digital serious games in the context of digital game-based learning are escape games, simulations and quizzes (Gruber & Faßbender, 2025). Quizzes are tests including numerous questions as well as related pre-formulated answer sets. In a quiz, questions can be phrased either single or multiple choice (Georg, 2022). Tools for the development of quizzes utilized in higher education are *genially* and *Kahoot!*:

- *genially* is a web-based tool offering features for creative content creation. The portfolio includes presentations, infographics and game-based learning like quizzes (Genially Web, 2023). Moreover, the platform allows for the incorporation of multimedia elements and interactive scenarios, making learning more engaging and dynamic. *Genially* has been evaluated for its pedagogical quality and is noted for increasing student engagement through gamified quizzes (Martinez et al., 2023; Sanchez & Plumettaz-Sieber, 2019).
- *Kahoot!* is a game-based learning web platform that allows educators to create quizzes and interactive lectures (Kahoot!, 2024). It is evident that the use of *Kahoot!* for educational purposes enhances constructivist learning, as the digital environment encourages learners to actively engage and construct knowledge and concepts. This allows for self-determined learning and developing students' problem-solving competences (Allegra et al., 2019).

In summary, *genially* and *Kahoot!* can contribute positively to constructivist learning in higher education. Considering the aptitude of digital serious games for STEM students, it is imperative to examine quizzes on these two platforms within the context of this study. The following section describes the research interest in more detail.

Objectives

The main objective of this project is to conduct a study on the use of digital quizzes as serious games in STEM higher education lectures. In addressing the issue of poor student engagement in recent lectures, this research is directed towards student motivation and engagement. The subordinate objectives of this project are summarized as follows:

- Creating digital quizzes based on predefined learning objectives for investigation.
- Evaluating engagement and motivation to measure the impact of these platforms in the learning process.

- Identifying challenges and limitations associated with implementing digital serious games in STEM higher education settings.
- Developing best practices to establish recommendations for effectively integrating these platforms into higher education teaching.

Based on these objectives, the following research question arises: *How do computer science students perceive digital serious games in STEM higher education?* The underlying research methodology selected for this project is outlined in the following section.

Methodology

Based on the aforementioned objectives, the study examines digital quizzes in STEM higher education lectures. For this purpose, 8 digital quizzes in *genially* as well as 2 *Kahoot!* quizzes were created based on lecture materials for computer science students at the University of Applied Sciences Karlsruhe, Germany. The *genially* quizzes are designed as single player games, whereas the *Kahoot!* quizzes are to be played in pairs against each other in plenary sessions. These 10 games were incorporated into the curriculum over the course of 3 semesters. On average, one to two games were used per lecture.

To accompany the use of the games in teaching, an anonymous evaluation was carried out at the end of each semester. This approach is chosen to evaluate the digital quizzes in the computer science lectures in STEM higher education, allowing students to share their opinions and give feedback. The anonymous questionnaire allows students to reply honestly and openly in the feedback process without fearing judgement on their opinions. Therefore, questionnaires are suitable to gather data on perceptions, experiences, motivation, and attitudes towards these platforms (Manzano-León, Camacho-Lazarraga, Guerrero, et al., 2021). The use of surveys allows the participants to work through the questions at their own pace. Digital surveys underscore the advantage of self-determination and anonymity and is therefore often applied to educational settings (Reinders et al., 2022).

The questionnaire was designed based on the standard lecture evaluation at the University of Applied Sciences Karlsruhe, Germany. The survey included questions on students perception of digital serious games, based on relevant literature on game-based learning (Manzano-León, Camacho-Lazarraga, Guerrero-Puerta, et al., 2021). Ethical considerations, including informed consent and anonymity, were addressed to protect participants' rights and privacy. As the questionnaire was anonymous, no conclusions can be drawn about individual participants. A total of n=87 higher education computer science students participated in the study.

The survey comprised open-text fields, and the respective entries in the open-text fields were assessed employed qualitative content analysis according to Philipp Mayring (2015). Qualitative content analysis is applied to explore meanings and patterns in qualitative data (Mayring, 2015). A systematic approach with predefined rules ensures that the analysis is both reproducible and transparent, addressing scientific quality criteria. The main objective is the development of categories, following deduction or induction: Deduction is the application of pre-defined, theory-based category to the material, whereas induction is defined as observing patterns in the data without pre-defined categories (Reinders et al., 2022). Regardless of the chosen approach, qualitative content analysis can be conducted in three different ways: (1) summarizing the qualitative data by shortening the text; (2) explicating the data by enriching the text with more

data sources; or (3) structuring the material by units of meaning (Mayring, 2015).

In this study, the structured qualitative content analysis was employed. Following this approach, the participants' entries were analysed using thematic coding, providing a comprehensive understanding of the research subject. The findings are outlined in the following section.

Findings

The qualitative content analysis results in four main categories: *Pleasant atmosphere*, *Positive perception of digital serious games*, *Interactivity and engagement*, and *Diversification of lectures by using digital serious games*. The categories and the respective absolute as well as relative frequencies are displayed in the following Table 1.

Table 1. Qualitative Content Analysis Categories.

Category	Absolute frequencies	Relative frequencies
Pleasant atmosphere	3	3,6%
Positive perception of digital serious games	48	57,8%
Interactivity and engagement	25	30,2%
Diversification of lectures by using digital serious games	7	8,4%

Table 1 shows that most of the student responses can be categorized as positive properties of digital serious games. The majority of all entries can be classified under Positive perception of digital serious games. Student responses show that students generally favour digital game-based learning. Enthusiasm for educational games and positive responses to the adoption of novel teaching methodologies was captured. There is a notable appreciation for the replacement of conventional teaching methods.

In particular, students appreciate being incentivized to actively participate in lectures. This is represented by the category *Interactivity and engagement*, which accounts for 30% of the survey responses. The students value their active participation in the quizzes. Students emphasized that studying with digital serious games like quizzes helps them testing their knowledge. The digital serious games are appreciated for their student-centred design and intuitive usability.

The platforms *genially* and *Kahoot!* were assessed suitable for digital game-based learning by the students. Students enjoyed the digital learning games, especially the quiz characteristics were mentioned to be helpful for consolidating lecture content. The *Diversification of lectures by using digital games* comprises 8,4% of all survey entries. Students appreciate the incorporation of digital serious games, which is partly attributable to their computer science affinity. Consequently, it is not surprising that students did not have any issues with the game usability or navigation. Following, 3,4% of entries show that participants value the *Pleasant atmosphere* created by using digital serious games. Participants value the competitiveness of the quizzes in *Kahoot!* and mentioned positive effects on group dynamics when working in pairs. The integration of both *genially* and *Kahoot!* incentivized students to participate and actively engage in the learning activity. Traditional unidirectional teaching by the lecturer in which students are

passive shifted to the students actively by playing the games, adopting an active role in the learning process.

In addition to providing feedback, the participants suggested possible enhancements of digital serious games regarding their utilization, frequency, and design. The following recommendations for higher education lecturers are derived from this study:

- The education of digital educational games into higher education is recommended once or twice per lecture to consolidate previously acquired knowledge. The incorporation of gamified elements is imperative in the creation of competitive and motivating characteristics within the game.
- The challenge of digital games lies in the necessity to learn how to utilize the tools and invest time in creating and maintaining them. However, it should be noted that the functionality of these devices may be constrained when employed within the specific context of a given didactic application.
- Digital games have the capacity to stimulate learners' motivation by achieving an equilibrium between playful elements, competition, and enjoyment. Group competitions, in which all learners competed against each other, were found to be particularly appealing.
- It is important that these digital learning games are aligned with the learning objectives. These should be communicated transparently in advance, allowing students to allocate the digital serious game in their learning process.

Evaluating these findings, it can be stated that the game design needs to be student-centred and intuitive to minimize barriers and maximize student engagement. In addition, the content created on *genially* and the quizzes designed on Kahoot! is to be closely aligned with the learning objectives to minimize barriers for students to engage with the material, maximizing the educational impact.

Discussion

This study sheds light on how computer science students perceive digital serious games in a quiz-based format during lectures. Based on the underlying research question *How do computer science students perceive digital serious games in STEM higher education?*, this study shows that the participants are generally in favour of digital serious games. In brief, the participating students have indicated that the learning games were enjoyable and anticipate their incorporation into future curricula. Digital game-based learning platforms like *genially* and *Kahoot!* foster student engagement by addressing students' interest in playing games. Student motivation is addressed by the playfulness of the game, having a positive impact on student engagement and activation.

In addition to the positive experience with digital learning games, minor technical issues were observed within the scope of this study, such as network instability, which affected students experience with *Kahoot!*. These technical issues had a disruptive effect on the learning experience. Ensuring reliable internet connectivity is crucial for the seamless implementation of these platforms.

Moreover, it is worth mentioning that the time required to create interactive content on *genially*

and design quizzes on *Kahoot!* can be a challenge for educators, especially those with limited technological proficiency. Designing quiz-based digital serious games requires more resources than designing traditional unidirectional teaching materials. This aspect needs to be considered when deciding for game-based learning.

The qualitative analysis of participant feedback allowed to deeply examine the data. Nevertheless, it is recommended to further assess students' perception of digital serious games. It may be of interest to the research community to investigate which game design elements are particularly beneficial and which are rather harmful to learning environments. Targeting the challenge of students being easily distracted while learning, immersive game design elements could incentivize students to be more involved in digital serious games, therefore leading to less distraction. Therefore, immersive game design elements are recommended to be investigated in future research.

Acknowledgement

The author declares that financial support was received for the research, authorship, and/or publication of this article. This project is funded in full as part of a PhD program "Wissensmedien" initiated by the Ministry of Science, Research and Arts Baden-Württemberg, Germany. The aim is to investigate innovative teaching and learning methods in higher education.

References

- Allegra, M., Gentile, M., & Söbke, H. (Eds.) (2019). *Games and Learning Alliance*. : Vol. 11385. Springer International Publishing.
- Bencsik, A., Mezeiova, A., & Samu, B. O. (2021). Gamification in Higher Education (Case Study on a Management Subject). *International Journal of Learning, Teaching and Educational Research*, 20(5), 211–231. <https://doi.org/10.26803/ijlter.20.5.12>
- Bröker, T., Voit, T., & Zinger, B. (2021). Gaming the System: Neue Perspektiven auf das Lernen. In Hochschulforum Digitalisierung (Ed.), *Digitalisierung in Studium und Lehre gemeinsam gestalten* (pp. 497–513). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-32849-8_28
- Buchner, J. T., Rüter, M., & Kerres, M. (2022). Learning with a digital escape room game: Before or after instruction? *Research and Practice in Technology Enhanced Learning*, 17(1), 10. <https://doi.org/10.1186/s41039-022-00187-x>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). ACM. <https://doi.org/10.1145/2181037.2181040>
- Ehrlich, I., Filipenko, M., Kranawetleitner, T., Krebs, H., Loew, R., Pistoll, P., & Thurner-Irmeler, J. (Eds.) (2020). *Escaping the Everyday Chaos: Assessing the Needs for Internal Knowledge Transfer in SMEs via an Escape Room*.
- Fleuren, D. (2016). *Open MINT Labs : Mit virtuellen Laboren zu höherem Lernerfolg*. Waxmann. <https://doi.org/10.25656/01:16821>
- Franco, P. F., & DeLuca, D. A. (2019). Learning Through Action: Creating and Implementing a Strategy Game to Foster Innovative Thinking in Higher Education. *Simulation & Gaming*, 50(1), 23–43. <https://doi.org/10.1177/1046878118820892>
- Frey, D., & Uemminghaus, M. (Eds.). (2021). *Innovative Lehre an der Hochschule*. Springer Berlin Heidelberg.
- Gagné, R. M. (1980). *Die Bedingungen des menschlichen Lernens* (5th ed.). *Beiträge zu einer neuen Didaktik. Reihe A*. Schroedel.
- Genially Web, S. L. (2023). *genially*. <https://app.genially.com/>
- Georg, S. (2022). *Möglichkeiten zur E-Learning-gestützten Lehre: Anwendung am Beispiel des Fachs Kostenrechnung. essentials*. Springer Gabler. <http://www.springer.com/https://doi.org/10.1007/978-3-658-36821-0>
- Gräsel, C., & Gniewosz, B. (2011). Überblick. In H. Reinders, H. Ditton, C. Gräsel, & B. Gniewosz (Eds.), *Lehrbuch. Empirische Bildungsforschung: Gegenstandsbereiche* (pp. 15–20). VS Verlag für Sozialwissenschaften; Springer Fachmedien Wiesbaden.

- Gruber, M. (2025). Unlocking Knowledge: Digitale educational Escape Games im MINT-Studium. In J. Blank, E. Niederhafner, & A. Bleicher (Eds.), *Transformation in Bildung und Hochschule* (pp. 111–122). Verlag Barbara Budrich.
- Gruber, M., & Faßbender, S. (2025). Digital educational escape game design for STEM higher education. *Frontiers in Education*, *10*, Article 1497291, 1–10. <https://doi.org/10.3389/feduc.2025.1497291>
- Ishak, S. A., Din, R., & Hasran, U. A. (2021). Defining Digital Game-Based Learning for Science, Technology, Engineering, and Mathematics: A New Perspective on Design and Developmental Research. *Journal of Medical Internet Research*, *23*(2), 1-14. <https://doi.org/10.2196/20537>
- Kahoot! (2024, December 27). Kahoot! | Learning games | Make learning awesome! <https://kahoot.com/>
- Kergel, D., & Heidkamp-Kergel, B. (2020). *E-Learning, E-Didaktik und digitales Lernen*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-28277-6>
- Kerres, M. (2024). *Mediendidaktik: Lernen in der digitalen Welt* (6th ed.). De Gruyter Studium. De Gruyter Oldenbourg.
- Makri, A., Vlachopoulos, D., & Martina, R. A. (2021). Digital Escape Rooms as Innovative Pedagogical Tools in Education: A Systematic Literature Review. *Sustainability*, *13*(8), 1–29. <https://doi.org/10.3390/su13084587>
- Manzano-León, A., Camacho-Lazarraga, P., Guerrero, M., Guerrero-Puerta, L., Aguilar-Parra, J. M., Trigueros, R., & Alias, A. (2021). Between Level Up and Game Over: A Systematic Literature Review of Gamification in Education. *Sustainability*, *13*(4), 1–14. <https://doi.org/10.3390/su13042247>
- Manzano-León, A., Camacho-Lazarraga, P., Guerrero-Puerta, M. A., Guerrero-Puerta, L., Alias, A., Aguilar-Parra, J. M., & Trigueros, R. (2021). Development and Validation of a Questionnaire on Motivation for Cooperative Playful Learning Strategies. *International Journal of Environmental Research and Public Health*, *18*(3). <https://doi.org/10.3390/ijerph18030960>
- Martinez, E., Álvarez Sánchez, J., Carbonell Padrino, M., Florez, M., & Montoya, E. (2023). Escaping the Lab: Increasing Motivation in the Physics Classroom through educational Escape Rooms. *Romanian Reports in Physics*(75), Article 901.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12th ed.). Beltz Pädagogik. Beltz.
- O'Brien, K., & Pitera, J. (2019). Gamifying Instruction and Engaging Students With Breakout EDU. *Journal of Educational Technology Systems*, *48*(2), 192–212. <https://doi.org/10.1177/0047239519877165>
- Prieto, F. Y., Su Jong, J., & González-Gómez, D. (2021). Virtual escape room and STEM content: Effects on the affective domain on teacher trainees. *Journal of Technology and Science Education*(2), Article 11, 1–7. <https://doi.org/10.3926/jotse.247>
- Reinders, H., Bergs-Winkels, D., Prochnow, A., & Post, I. (2022). *Empirische Bildungsforschung*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-27277-7>
- Sanchez, E., & Plumettaz-Sieber, M. (2019). Teaching and Learning with Escape Games from Debriefing to Institutionalization of Knowledge. In M. Allegra, M. Gentile, & H. Söbke (Eds.), *Games and Learning Alliance* (pp. 242–253). Springer International Publishing.
- Semmler, L., Belova, N., Engstler, V., Gruber, M., Strippel, C. G., & Groß, K. (2024). Escape Rooms in Chemie: Einsatz- und Forschungsszenarien in der Lehre. In H. van Vorst (Chair), *Jahrestagung in Bochum 2024*. Symposium conducted at the meeting of Gesellschaft für Didaktik der Chemie und Physik, Bochum.
- Taraldsen, L. H., Haara, F. O., Lysne, M. S., Jensen, P. R., & Jenssen, E. S. (2022). A review on use of escape rooms in education – touching the void. *Education Inquiry*, *13*(2), 169–184. <https://doi.org/10.1080/20004508.2020.1860284>
- van Ackeren, I., Kerres, M., & Heinrich, S. (2017). *Flexibles Lernen mit digitalen Medien. Strategische Verankerung und Handlungsfelder an der Universität Duisburg-Essen*. Waxmann. <https://doi.org/10.25656/01:15385>
- Vygotsky, L., Cole, M., Jolm-Steiner, V., Scribner, S., & Souberman, E. (1980). *Mind in Society*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Weselek, J. (2022). *Digitale Bildung Für Nachhaltige Entwicklung: Herausforderungen und Perspektiven Für Die Hochschulbildung*. Springer Berlin / Heidelberg. <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=7102060>

Activating Epistemic Emotions To Enhance Student Engagement In Socio-Scientific Issues Through Digital Curation

Remah Haj and Dina Tsybulsky
Technion Israel Institute of Technology, Israel

Learning through socio-scientific issues (SSI) confronts students with uncertainty, competing knowledge claims, and the need to justify positions in the absence of clear-cut answers. Such epistemically demanding learning situations are likely to evoke a range of emotions that are closely tied to students' processes of sense-making and participation in learning, yet these emotional dynamics remain underexplored, particularly in digitally mediated SSI contexts. Addressing this gap, the present study examined the potential of Digital Curation (DC) as a learning pathway for integrating SSI learning with intentional activation of epistemic emotions. The study employed a qualitative bounded case design in a tenth-grade biology classroom in Israel (n=42). Students participated in four structured units on animal welfare (food choices, environmental conservation, animal experimentation, and legislation). Data were collected through reflective written reports, semi-structured interviews, and video recordings of classroom discussions. A combined deductive–inductive content analysis was employed to investigate the interplay between epistemic emotions and dimensions of student engagement. The findings revealed that the SSI-DC learning pathway initially elicited negative epistemic emotions such as confusion, frustration, and anxiety, which gradually transformed into curiosity, enjoyment, and surprise as students progressed through the learning process. This emotional trajectory supported sustained engagement across cognitive, behavioural, emotional, social, and agentic dimensions. These findings underscore the importance of designing emotionally responsive digital learning environments that intentionally leverage epistemic emotions to enhance student engagement and foster a sense of ownership in learning.

Keywords: Socio-scientific issues, digital curation, epistemic emotions, student engagement

Introduction

In recent years, science education has shifted towards actively involving students in socio-scientific issues (SSIs) that directly connect to their everyday lives, rather than solely imparting scientific knowledge. This shift supports students' understanding of scientific concepts while encouraging critical examination, informed decision-making, and the construction of scientific arguments in response to ill-structured social dilemmas (Sadler et al., 2016). However, sustaining student engagement with SSIs remains challenging, particularly in digitally mediated learning environments (Klosterman et al., 2012; Solli, 2021).

Integrating Digital Curation (DC) into SSI teaching has recently attracted considerable attention due to its potential to enhance student engagement (Dayan & Tsybulsky, 2024). DC involves selecting, organizing, and preserving digital content to create coherent collections of relevant and reliable information (Tsybulsky, 2020) and has been shown to foster critical thinking and media literacy practices needed to navigate complex online information (Gadot & Tsybulsky, 2023; 2025; Forkosh et al., 2021).

Beyond its documented contribution to student engagement, integrating DC into SSI learning also places students in epistemically demanding situations that require them to search for information, evaluate the credibility of sources, negotiate conflicting claims, and justify knowledge-based positions. These features position DC not only as a pedagogical tool, but as a learning

environment that foregrounds processes of sensemaking under conditions of uncertainty.

Building on this perspective, our research group proposed the SSI-DC instructional pathway as a learning pathway within the SSI teaching and learning framework, integrating the social and scientific dimensions of SSI learning through digital curation (Dayan & Tsybulsky, 2024). In our prior SSI-DC work (Haj & Tsybulsky, under review), students frequently expressed confusion, curiosity, frustration, and surprise, suggesting that epistemic emotions - knowledge-focused emotions arising during information processing and sensemaking - may function as mechanisms that shape engagement (Brun et al., 2008; Pekrun & Stephens, 2012; Muis et al., 2018). Yet, research has rarely examined how epistemic emotions are activated in digitally mediated SSI contexts and how they contribute to multidimensional engagement (Fredricks et al., 2004; Sinatra et al., 2015). Accordingly, this study demonstrates how a designed SSI-DC pathway can transform negative epistemic emotions (confusion, frustration, anxiety) into positive ones (curiosity, enjoyment, surprise) and thereby strengthen cognitive, behavioural, emotional, social, and agentic engagement. The guiding research question was: How are epistemic emotions activated during the SSI-DC learning pathway, and how do they shape students' multidimensional engagement (cognitive, behavioural, emotional, social, and agentic)?

Method

Research Setting And Participants

This qualitative bounded case study was conducted in a secondary school in Israel and focused on a single tenth-grade biology classroom ($n = 42$). The classroom was selected as a bounded case (Yin, 2003), enabling an in-depth exploration of how the SSI-DC pathway influenced students' epistemic emotions and multidimensional engagement within their authentic instructional environment. The students were enrolled in an advanced science track and had prior exposure to SSI. Throughout the learning pathway, students participated in four structured SSI-DC units related to animal welfare (food choices, environmental conservation, animal experimentation, and legislation). Each unit spanned approximately 8 hours and required students to collaboratively research, curate, and reflect on digital content using Wakelet. Each unit began with an open-ended, provocative question, followed by whole-class discussion and peer dialogue in self-selected groups, and concluded with group presentations and whole-class evaluation of competing positions. All activities were conducted during regular classroom hours; units were required but ungraded. Ethical approvals were obtained from the Ministry of Education and the Institutional Review Board (IRB). Informed consent was obtained from all participating students and their parents, including permission for classroom video recordings. Confidentiality was ensured, and pseudonyms were used in all transcripts and reports.

Data Collection

Data were collected from three sources: (A) guided reflective written reports completed after each animal welfare lesson, focusing on students' emotional responses, engagement, and participation in the SSI-DC process; (B) face-to-face semi-structured student interviews conducted after completing the SSI-DC unit, including 14 open-ended questions examining students' emotions and perceptions of learning and engagement. Each interview lasted about 50 minutes, was recorded and transcribed, and remained flexible to follow participants' responses while prompting students to reconstruct the steps of creating their digital collections; and (C) video documentation of selected classroom sessions capturing group discussions, affective expressions, and collaborative reasoning, enabling analysis of verbal and non-verbal interaction and how epistemic emotions emerged and evolved in real time within the SSI-DC environment.

Data Analysis

We analysed the data using a combined deductive–inductive qualitative content analysis (Schreier, 2012; Kuckartz, 2016). Deductive coding was guided by established frameworks of student engagement (Fredricks et al., 2018; Sinatra et al., 2015) and epistemic emotions (Pekrun et al., 2016), while inductive coding enabled the identification of emergent patterns not captured by the predefined categories (Elo & Kyngäs, 2008). First, reflective reports, interview transcripts, and selected classroom video recordings were reviewed to familiarize ourselves with the data. Epistemic emotions were then coded using a theory-informed codebook developed based on Pekrun’s conceptualization of epistemic emotions (Pekrun et al., 2016) and related empirical work on knowledge-related emotions in learning contexts. This codebook included definitions of each epistemic emotion (e.g., confusion, frustration, anxiety, curiosity, enjoyment, and surprise), their key features, and the cognitive conditions and triggers through which they typically emerge (e.g., contradictory information, uncertainty, difficulty verifying sources). In parallel, student engagement was coded using a multidimensional codebook grounded in established engagement models (Fredricks et al., 2018; Sinatra et al., 2015). This engagement codebook operationalized five dimensions—cognitive, behavioural, emotional, social, and agentic engagement—by specifying definitional characteristics and indicative manifestations in students’ written reflections, interview responses, and classroom discourse. The data were subsequently revisited to examine how epistemic emotions co-occurred with, and shaped, students’ engagement throughout the SSI-DC pathway, and themes were refined through iterative research team discussions.

Trustworthiness

To ensure trustworthiness, we applied triangulation across data sources—student reflections, interviews, and classroom observations—to enhance credibility (Lincoln & Guba, 1985). Multiple researchers independently coded the data and resolved discrepancies through discussion. The analytic process was systematically documented, and key interpretations were supported with direct excerpts from the data. Rich contextual information about the setting and participants was included to support transferability, while transparency in coding and decision-making strengthened confirmability (Shenton, 2004).

Findings

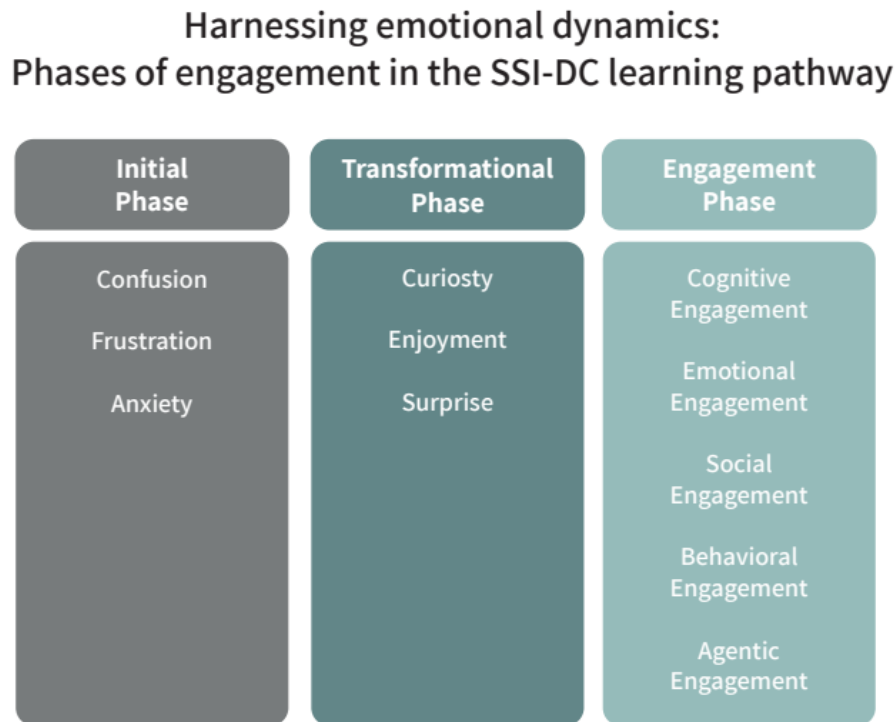
The findings presented in this section are organized into the Initial Phase, Transformational Phase, and Engagement Phase (see Figure 1). Each phase represents a different stage in the learning process and shows the dynamic interplay between the dimensions of engagement and the evolution of epistemic emotions in the SSI-DC learning pathway, thereby helping account for how engagement and epistemic emotions interact throughout the learning experience.

Initial Phase: Pre-SSI-DC Learning Epistemic Emotions

The initial phase reflects students’ baseline epistemic emotions prior to engaging in SSI-DC. This stage was dominated by confusion, frustration, and anxiety, stemming from uncertainty about the unfamiliar learning method and the complexity of SSI tasks. Although these emotions were strongest at the outset, they occasionally reappeared later, especially when students confronted contradictory evidence or were required to justify their claims publicly. Positive emotions were limited during this phase and became more salient only in later stages.

A. Confusion: Introducing SSI-DC initially generated confusion, mainly due to unfamiliarity with the digital curation process and uncertainty about how to approach complex dilemmas.

Figure 1. The three phases highlight the dynamic interplay between the engagement dimensions and the evolving epistemic emotions in the SSI-DC learning pathway.



B. Frustration: Emerged when students struggled with information overload, unreliable sources, and disagreement within groups. In many cases, frustration also pushed students toward deeper searching and verification.

C. Anxiety: Was common as students worried about succeeding in an unfamiliar task and meeting expectations, especially under time pressure or peer comparison.

Table 1. Students' quotes illustrating epistemic emotions in the Initial Phase.

Emotion	Students quotes
Confusion	"At the start, I couldn't settle on a clear position, so I kept going back and forth and hesitated to speak." (DF10, interview) "I didn't understand how to approach the topic at first, and that uncertainty made me feel lost." (NF10, interview)
Frustration	"It frustrated me when others' views didn't align with mine, and it pushed me to search more intensely for supporting information." (DF10, interview) "There were so many sources, and many didn't seem reliable, which made the task feel overwhelming." (DF10, reflective report)
Anxiety	"At the beginning, I was worried and hesitant because the process was new and I wasn't sure I could handle it." (FF10, interview) "I felt anxious about not finding strong materials and not meeting expectations." (LF10, reflective report)

Transformational Phase: Transformation Of Emotions During SSI-DC

As students progressed through SSI-DC, the early negative emotions gradually decreased and were replaced by curiosity, enjoyment, and surprise. Although confusion and frustration occasionally resurfaced, they became less central. Positive epistemic emotions increasingly supported deeper inquiry, stronger collaboration, and more meaningful engagement.

A. Curiosity: Became a key driver of exploration and persistence, encouraging students to investigate multiple perspectives and verify information.

B. Enjoyment: Increased once students became comfortable with the platform and experienced successful collaboration and progress.

C. Surprise: Often appeared when students encountered unexpected insights or new arguments, prompting them to reconsider ideas and deepen reasoning.

Table 2. Students' quotes illustrating epistemic emotions in the Transformational Phase.

Emotion	Students quotes
Curiosity	<p>“Exploring different sources and perspectives kept me engaged and made me want to continue investigating.” (AF10, interview)</p> <p>“When I found evidence that challenged what I believed, I became more curious and checked additional sources to understand better.” (JF10, classroom discussion)</p>
Enjoyment	<p>“Once I started working with different materials like videos, images, and articles, I felt increasingly excited.” (FF10, reflective report)</p> <p>“I enjoyed building the digital collection and discussing it with the class—it made the learning feel engaging.” (NF10, interview)</p>
Surprise	<p>“I was surprised by how much I learned and by how engaging the whole process turned out to be.” (DF10, reflective report)</p> <p>“When another group presented evidence we hadn't considered, it surprised me and made us rethink our position.” (HF10, classroom discussion)</p>

Table 3. Students' quotes illustrating the five dimensions of engagement.

Engagement	Students quotes
Cognitive engagement	<p>“Following the structured steps and building the group collection helped me understand the topic in a deeper way.” (FF10, interview)</p> <p>“When classmates challenged my source, I rechecked it and justified it, which pushed me to analyse more carefully.” (JF10, classroom discussion)</p>
Emotional engagement	<p>“As I learned more, the process became more enjoyable and broke the routine—it felt energizing.” (KF10, interview)</p> <p>“Presenting our collection made me feel proud, especially because my contribution really mattered.” (TF10, classroom discussion)</p>
Social engagement	<p>“Class discussions exposed me to ideas and information I hadn't encountered before.” (DF10, reflective report)</p> <p>“Working together on the collection helped me understand more and made the learning experience more enjoyable.” (KF10, interview)</p>
Behavioural engagement	<p>“I stayed highly involved and organized, completed all the steps, and even went beyond them because I wanted to keep researching.” (DF10, interview)</p> <p>“Even when the group disagreed, I kept searching for evidence because I didn't want to give up.” (LF10, classroom discussion)</p>
Agentic engagement	<p>“This approach made me feel like an independent researcher, responsible for building my understanding.” (DF10, interview)</p> <p>“I suggested reorganizing our collection so it would be clearer, and once the group agreed, I felt my voice had real impact.” (JF10, classroom discussion)</p>

Engagement Phase: During SSI-DC Learning

In the engagement phase, curiosity, enjoyment, and surprise translated into sustained engagement across all five dimensions. Negative emotions appeared rarely and no longer shaped the overall

experience. Instead, positive epistemic emotions were the main drivers of students' effort, persistence, collaboration, and agency.

A. Cognitive Engagement: Students demonstrated deeper processing, synthesis, and critical evaluation of information as they built and defended their curated arguments.

B. Emotional Engagement: Was reflected in interest, enthusiasm, pride, and a growing sense of belonging during collaborative work and presentations.

C. Social Engagement: Peer dialogue and teamwork supported richer reasoning, exposure to new perspectives, and stronger learning relationships.

D. Behavioural Engagement: Students showed persistence, high effort, and increased time investment, even when facing disagreements or difficulties.

E. Agentic Engagement: Students increasingly shaped their learning by initiating ideas, personalizing tasks, and influencing group decisions.

Discussion and Conclusion

This study examined how epistemic emotions function as a mechanism for enhancing student engagement within the SSI-DC instructional pathway. Across the three phases (Initial, Transformational, Engagement), students shifted from negative epistemic emotions (confusion, frustration, anxiety) to positive ones (curiosity, enjoyment, surprise), and this trajectory supported sustained engagement across cognitive, behavioural, emotional, social, and agentic dimensions. These patterns align with research positioning SSI learning as engagement with ill-structured dilemmas that require learners to evaluate evidence, negotiate values, and construct arguments through discourse (Sadler, 2004; Zeidler, 2014; Sadler et al., 2017). They also reinforce the view that student engagement is multidimensional and dynamically shaped by instructional context (Fredricks et al., 2004; Sinatra et al., 2015). These phase-based patterns were identified through the theory-informed codebooks used in the analysis, which operationalized epistemic emotions and the five engagement dimensions across students' reflections, interviews, and classroom discourse (Pekrun et al., 2016; Fredricks et al., 2018; Sinatra et al., 2015).

The observed emotional transitions are consistent with theoretical accounts suggesting that epistemic emotions emerge when learners encounter uncertainty, contradiction, or cognitive incongruity in knowledge-rich tasks (Pekrun & Stephens, 2012; Muis et al., 2018; D'Mello & Graesser, 2012). Specifically, confusion and frustration appeared when students faced competing claims or struggled to verify sources, reflecting findings that epistemic emotions can be triggered by unresolved conflict and information overload in complex learning situations (Di Leo et al., 2019; D'Mello et al., 2014). Over time, these emotions shifted toward curiosity, enjoyment, and surprise—emotions that have been linked to sustained information exploration, deeper processing, and more productive engagement with knowledge-building tasks (Litman, 2008; Loewenstein, 1994; Brun et al., 2008). This progression supports Pekrun's control-value theory, which highlights how both negative and positive emotions can shape learning outcomes, but positive activating emotions are especially relevant for maintaining attention, persistence, and meaningful participation (Pekrun, 2006, 2016).

From a design perspective, three features of SSI-DC appear central to this transformation: (1) posing open-ended, controversial questions that create productive uncertainty and invite position-taking; (2) guiding students through iterative digital curation practices (locating, evaluating, organizing, and synthesizing resources) that surface contradictions and require justification; and (3) structuring collaborative dialogue and public sharing that prompts peer critique and refinement of claims. These design features resonate with research on technology-supported SSI

learning emphasizing digital practices such as evaluating sources, media literacy, and participation in discourse as core components of contemporary SSI pedagogy (Vlachopoulos & Makri, 2017; Solli, 2021; Soomro et al., 2020). They also align with digital curation literature positioning curation as an active knowledge-construction process that supports critical thinking, authenticity, and learner agency (Deschaine & Sharma, 2005; Pasek & Hargittai, 2015; Antonio & Tuffley, 2015).

The findings further showed that cognitive and emotional engagement were especially prominent, supporting prior work highlighting the interdependence of engagement dimensions and the central role of emotion in sustaining meaningful learning (Fredricks et al., 2016; Wang & Sui, 2020). Emotional engagement through curiosity and enjoyment appeared to stabilize students' persistence and investment in complex tasks, consistent with evidence linking positive affective experiences to academic engagement and success (Ucar & Sungur, 2017; Fredricks et al., 2016). At the same time, the collaborative structure supported social and agentic engagement, reflecting research emphasizing that peer dialogue and shared reasoning contribute to deeper engagement in SSI contexts (Sadler et al., 2017) and that agentic contributions are strengthened when students are positioned as active participants who shape their learning environment (Sinatra et al., 2015; Fredricks et al., 2018). Together, the phased model proposed here extends prior SSI-DC work by foregrounding epistemic emotions not only as outcomes of learning but as mechanisms that can be intentionally leveraged to strengthen multidimensional engagement in digitally mediated SSI instruction.

Acknowledgement

We thank the participating students and their science teacher for their engagement and commitment throughout the SSI-DC units. We also acknowledge the contributions of the research team for their support in data collection, analysis, and ongoing discussion throughout the study. We are grateful to the ESERA Proceedings Editors for their careful editorial work and guidance, and to the ESERA Conference and Doctoral School organizers for providing a supportive and inspiring scholarly community.

References

- Antonio, A., & Tuffley, D. (2015). The role of digital curation in knowledge management. *Journal of Digital Curation*, 10(1), 57-72.
- Dayan, E., & Tsybulsky, D. (2024). Designing socio-scientific issues' teaching: models for implementing digital curation in science education. *Journal of Science Education and Technology*, 1-20.
- Deschaine, M., & Sharma, S. A. (2015). The five Cs of digital curation: Supporting twenty-first-century teaching and learning. *InSight: A Journal of Scholarly Teaching*, 10, 19–24. <https://doi.org/10.46504/10201501de>
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109. <https://doi.org/10.3102/00346543074001059>
- Fredricks, J. A. (2011). Engagement in school and out-of-school contexts: A multidimensional view of engagement. *Theory Into Practice*, 50(4), 327-335. <https://doi.org/10.1080/00405841.2011.607401>
- Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction*, 43, 1-4.
- Fredricks, J. A., Wang, M., Linn, J. S., Hofkens, T., Sung, H. C., Parr, A., & Allerton, J. (2016). Using qualitative methods to develop a survey measure of math and science engagement. *Learning and Instruction*, 43, 5–15. <https://doi.org/10.1016/j.learninstruc.2016.01.009>
- Fredricks, J. A., Hofkens, T., Wang, M., Mortenson, E., & Scott, P. J. (2018). Supporting girls' and boys' engagement in math and science learning: A mixed methods study. *Journal of Research in Science*

- Teaching, 55(2), 271–298. <https://doi.org/10.1002/tea.21419>
- Gadot, R., Tsybulsky, D. Digital Curation as a Pedagogical Approach to Promote Critical Thinking. *J Sci Educ Technol* 32, 814–823 (2023). <https://doi.org/10.1007/s10956-022-10016-x>
- Gadot, R., Tsybulsky, D. Taxonomy of digital curation activities that promote critical thinking. *Smart Learn. Environ.* 12, 17 (2025). <https://doi.org/10.1186/s40561-025-00365-6>
- Haj, R., Tsybulsky, D. (Under review). From confusion to curiosity: epistemic emotions in socio-scientific issue learning through digital curation. *Journal of Disciplinary and Interdisciplinary Science Education Research*.
- Klosterman, M. L., Sadler, T. D., & Brown, J. (2012). Science teachers' use of mass media to address socio-scientific and sustainability issues. *Research in Science Education*, 42(1), 51–74. <https://doi.org/10.1007/s11165-011-9256-z>
- Kuckartz, U. (2016). Qualitative text analysis: A systematic approach. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 341–356). SAGE.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Pasek, J., & Hargittai, E. (2015). Digital literacy in an age of information overload. *Journal of Information Science*, 41(6), 723-731.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259–282). Springer. https://doi.org/10.1007/978-1-4614-2018-7_12
- Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2016). Measuring emotions during epistemic activities: The epistemically related emotion scales. *Cognition and Emotion*, 31(6), 1268–1276. <https://doi.org/10.1080/02699931.2016.1204989>
- Pekrun, R., & Stephens, E. J. (2012). *Academic emotions*.
- Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, 41(5), 513–536. <https://doi.org/10.1002/tea.20009>
- Sadler, T. D. (2009). Situated learning in science education: Socio-scientific issues as contexts for practice. *Studies in Science Education*, 45(1), 1–42. <https://doi.org/10.1080/03057260802681839>
- Sadler, T. D. (2011). *Socio-scientific issues in the classroom: Teaching, learning, and research*. Springer.
- Sadler, T. D., Romine, W. L., & Topçu, M. S. (2016). Learning science content through socio-scientific issues-based instruction: A multi-level assessment study. *International Journal of Science Education*, 38(10), 1622-1635. <https://doi.org/10.1080/09500693.2016.1204481>
- Sadler, T. D., Foulk, J. A., & Friedrichsen, P. J. (2017). Evolution of a model for socio-scientific issue teaching and learning. *International Journal of Education in Mathematics, Science and Technology*, 5(2), 75-87.
- Schreier, M. (2012). *Qualitative content analysis in practice*. SAGE.
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63-75. <https://doi.org/10.3233/EFI-2004-22201>
- Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, 50(1), 1–13. <https://doi.org/10.1080/00461520.2014.1002924>
- Solli, A., Hillman, T., & Mäkitalo, Å. (2019). Navigating the complexity of socio-scientific controversies: How students make multiple voices present in discourse. *Research in Science Education*, 49, 1595-1623. <https://doi.org/10.1007/s11165-017-9668-5>
- Solli, A. (2021). Appeals to science: Recirculation of online claims in socio-scientific reasoning. *Research in Science Education*, 51(S2), 983–1013. <https://doi.org/10.1007/s11165-019-09878-w>
- Soomro, K. A., Kale, U., Curtis, R., Akcaoglu, M., & Bernstein, M. (2020). Digital divide among higher education faculty. *International Journal of Educational Technology in Higher Education*, 17, Article 5. <https://doi.org/10.1186/s41239-020-00191-5>
- Tsybulsky, D. (2020). Digital curation for promoting personalized learning: A study of secondary-school science students' learning experiences. *Journal of Research on Technology in Education*, 52(3), 429–440. <https://doi.org/10.1080/15391523.2020.1728447>
- Vlachopoulos, D., & Makri, A. (2017). The effect of games and simulations on higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, 14, Article 22. <https://doi.org/10.1186/s41239-017-0062-1>
- Wang, L., & Sui, J. (2020). Correlation analysis between emotional engagement and achievement of university students in online learning based on an elective course. *Transactions on Edutainment*, 16, 13-24.

- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). SAGE.
- Zeidler, D. L. (2014). Socio-scientific issues as a curriculum emphasis: Theory, research, and practice. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research in science education* (pp. 711-740). Routledge.

A New Paradigm For Science Lesson Planning: Adaptive Expertise And Generative AI Integration

*YeJun Bae*¹ and *Ercin Sahin*²

¹Appalachian State University, USA

²Ministry of National Education, Turkey

This study examines pre-service teachers' reflections on using ChatGPT for lesson planning. As AI-integrated learning environments gain prominence in schools, preparing teachers for their future roles in science education is increasingly important. Building on Davis et al.'s (2024) emphasis on integrating AI tools in teacher preparation, we argue that using such tools is a critical practice for developing pre-service teachers' adaptive expertise. The data comprises 116 quotes from the reflexive writing of elementary pre-service teachers enrolled in science teaching methods courses during two semesters. Participants used ChatGPT as part of classroom activities to design 5E lesson plans. A coding scheme was developed based on two key dimensions of adaptive expertise—efficiency and innovation—and an additional category, collaboration, identified through open coding. Findings revealed that ChatGPT enhanced pre-service teachers' efficiency by streamlining the brainstorming process and generating lesson plan ideas quickly. It also supported their innovation by encouraging critical reflection and adaptation of ChatGPT's outputs to align with classroom contexts and student learning needs. Additionally, integrity emerged as an essential component of adaptive expertise in AI-integrated learning environments, highlighting the need for ethical awareness and critical evaluation of ChatGPT's relevance and accuracy. The study underscores the importance of repeated practice in using ChatGPT effectively and fostering critical thinking to ensure the appropriate adoption of AI tools in lesson planning. These findings offer valuable insights for practitioners and researchers aiming to prepare teachers for the challenges of 21st-century classrooms.

Keywords: AI literacy, Preservice teacher education, Qualitative research

Background

The use of AI tools, particularly ChatGPT, has grown substantially in education, with notable applications in teacher education programs. Goodman et al. (2024) highlighted the critical need to enhance pre-service science teachers' competence in effectively integrating ChatGPT to improve teaching practices and streamline lesson preparation. Furthermore, there is a growing demand for establishing ethical guidelines to ensure the responsible use of generative AI tools in teacher education, addressing concerns such as plagiarism and biases (University of South California, 2024, October 14). Despite these calls, clear strategies for supporting pre-service teachers in developing the competencies needed for lesson planning with generative AI tools remain lacking. Adaptive expertise, as emphasized by Hess et al. (2023), offers a comprehensive framework to address these challenges, balancing timesaving and creating a new idea. This framework not only helps preservice teachers to well prepare for the use of generative AI tools but also ensures that they critically evaluate AI outputs and adapt their practices to diverse and evolving educational contexts.

Adaptive expertise is the ability to flexibly navigate between routine efficiency and creative modification to solve problems (Opre, 2015; Schwartz et al., 2005). This competency is essential for managing the complexities of routine teaching tasks (Hess et al., 2023). Davis et al. (2024) highlight adaptive expertise as a foundational component in establishing teacher competencies for integrating generative AI tools into instructional development. Two key dimensions of adaptive expertise—efficiency (streamlining routine processes) and innovation (generating novel ideas)—are particularly relevant for effectively utilizing generative AI tools to create inclusive

science learning environments. For example, a 5th-grade science teacher might use ChatGPT to quickly generate ideas for inclusive activities related to a life science topic. The teacher can then critically reflect on these suggestions, assessing their understanding of inclusivity and adapting them to better meet students' needs. This iterative process of leveraging AI tools while engaging in reflective practice exemplifies the dynamic interplay between efficiency and innovation.

Building on the vital role of adaptive expertise in teaching, generative AI tools like ChatGPT significantly support teachers in exercising this competency, particularly in lesson planning. These tools enhance efficiency by enabling teachers to quickly access familiar information, such as the components of the 5E learning cycle (Engage, Explore, Explain, Elaborate, Evaluate), through simple prompts. In addition to this ability, teachers also need to be able to apply judgment and adapt AI-generated content to student needs, ensuring ethical use and addressing biases.

Theoretical Framework

Our framework emphasizes the interplay between adaptive expertise and the integration of AI tools in preservice teacher education, specifically within the context of science lesson planning. We identify three critical components essential for effectively leveraging adaptive expertise when utilizing generative AI tools in this context: efficiency, innovation, and integrity. Efficiency and Innovation represent the core dimensions of adaptive expertise, streamlining routine processes and fostering novel ideas. Building on this foundation, we introduce Integrity as an additional essential component. Integrity encompasses the ethical and reflexive application of AI tools, ensuring that their use aligns with educational goals and values while fostering trust in the learning process.

Efficiency refers to the ability to apply knowledge or generate solutions based on established routines quickly. According to De Arment et al. (2013), experienced teachers often exhibit high levels of efficiency in routine practices such as classroom management and lesson documentation. However, preservice teachers typically require additional support to develop efficiency, as they lack extensive experience in classroom teaching and interacting with students who have diverse needs.

Innovation, the second key component, is closely linked to flexibility—the ability to adjust context-specific decisions based on various students' learning needs. Unlike efficiency, which is rooted in routine and time-saving behaviours, innovation requires teachers to draw on their knowledge and skills to adapt prior experiences and make decisions that optimize student learning (De Arment et al., 2020). It involves grappling with complexity and finding ways to navigate and understand it within diverse classroom contexts.

The third component, Integrity, is particularly relevant in the context of generative AI tools. While efficiency and innovation enable effective decision-making, the integration of AI tools introduces the need for critical reflection on digital resources. Davis et al. (2024) emphasize that evaluating AI-generated content's reliability, bias, and ethical implications is a key competency in an AI-integrated educational landscape.

Balancing these three components, efficiency, innovation, and Integrity, is vital for leveraging generative AI tools effectively. Despite their importance, few studies have explored how adaptive expertise intersects with the use of generative AI tools in lesson planning. Our study aims to fill this gap by examining how these aspects contribute to teacher competency in AI-integrated environments.

Purpose of Study

The purpose of this study is to investigate how preservice teachers potentially demonstrate three key competencies—efficiency, innovation, and integrity—when using ChatGPT in science lesson

planning. Specifically, our research seeks to address the following questions:

1. To what extent do preservice teachers exhibit the three key competencies, efficiency, innovation, and integrity, when integrating ChatGPT into science lesson planning?
2. What common patterns emerge from preservice teachers' lived experiences with ChatGPT in the context of science lesson planning?

Method

Data was collected over two semesters, including the Spring and Fall semesters of 2024. Participants consisted of preservice elementary education students enrolled in elementary science teaching methods courses. As part of their coursework, students were required to design a 5E lesson plan to implement during their internships. To support this process, students were encouraged to use ChatGPT for brainstorming ideas, enhancing their ability to support diverse learners, and creating more inclusive learning environments.

Following the classroom activity, students submitted reflections as part of an exit task. These reflections were collected via a Google Form that included questions about the frequency of their ChatGPT usage, their preferences for using ChatGPT, and open-ended prompts about their overall experiences with the tool in lesson planning.

The data were analysed using a deductive qualitative analysis approach (Pandey, 2019). A coding scheme was developed based on the three key competencies for integrating generative AI tools in lesson planning, as outlined in the theoretical framework section (Table 1). Within these three broad categories, we developed specific codes by reviewing the data through several rounds of review.

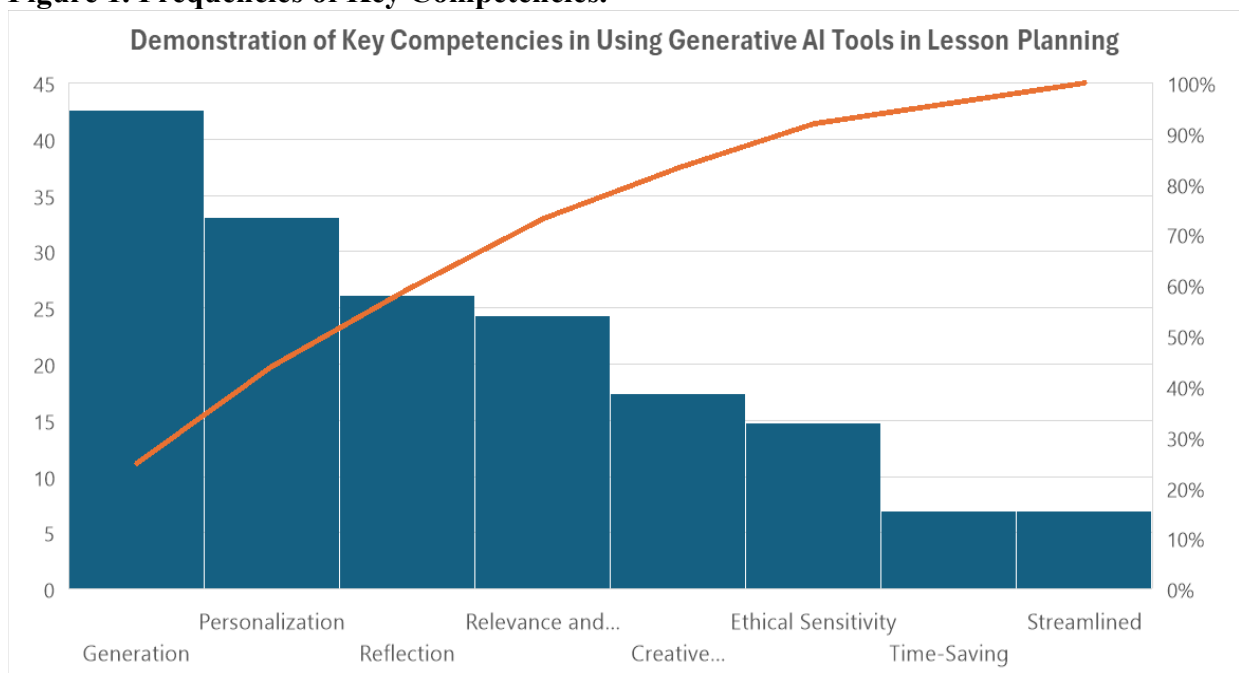
Table 1. Coding Scheme.

Category	Code	Explanation
Efficiency	Timesaving	AI reduces the time needed for lesson planning
	Streamlined Process	Preservice teachers easily adapt AI ideas to specific classroom needs.
	Facilitated Reflection	AI prompts deeper familiar tasks connecting new ideas with prior knowledge.
Innovation	Idea Generation	AI sparks strategies and ideas for science teaching as a brainstorming tool
	Creative Adaptation	Preservice teachers use AI outputs to develop various novel lesson plan ideas that suit their classrooms.
	Personalization for Equity	Preservice teachers recognize the necessity of tailoring AI suggestions to better meet diverse student needs and learning contexts.
Integrity	Relevance and Accuracy	Preservice teachers ensure AI outputs are relevant, accurate, and culturally sensitive.
	Ethical Sensitivity	Preservice teachers remain mindful of ethical concerns, including bias, and over-reliance on AI.

Findings

The data analysis allowed us to find answers to two research questions. First, we identified the competencies that participating preservice teachers demonstrated most frequently. Based on the analysis, participants showed a strong tendency toward idea generation, a component of the innovation category (Figure 1). Specifically, participants appreciated using generative AI tools to generate initial ideas for lesson design. This finding underscores a critical area for supporting preservice teachers in science lesson planning. De Arment et al. (2013) emphasized that unlike experienced in-service teachers, preservice teachers often require additional scaffolding to bridge their theoretical understanding from teacher education courses to practical applications in real classroom environments. Our study highlights the importance of fostering innovation through generative AI tools, demonstrating their potential to enhance preservice teacher education.

Figure 1. Frequencies of Key Competencies.



Additionally, the findings revealed that participants exhibited a high sensitivity to the integrity aspect. While they found generative AI tools helpful for creative brainstorming and starting lesson plans, they also expressed awareness of ethical considerations and potential biases inherent in using such tools. This aligns with Davis et al. (2024), who stress that integrity is a critical competency in the emerging era of generative AI in education.

Additionally, the findings revealed that participants exhibited a high sensitivity to the integrity aspect. While they found generative AI tools helpful for creative brainstorming and starting lesson plans, they also expressed awareness of ethical considerations and potential biases inherent in using such tools. This aligns with Davis et al. (2024), who stress that integrity is a critical competency in the emerging era of generative AI in education.

Our findings are consistent with existing literature that emphasizes the importance of integrating AI tools into science teacher education. However, this study advances the conversation by focusing on three critical competencies, efficiency, innovation, and integrity, necessary for the effective and responsible use of generative AI tools. These findings have important implications for science teacher educators, suggesting the need for the intentional incorporation of generative AI tools into teacher preparation programs to foster adaptive expertise while addressing both opportunities and challenges.

Conclusion

This study demonstrates how ChatGPT into science lesson planning is not merely about AI proficiency but about fostering adaptive expertise. Our findings highlight a critical shift in teacher preparation: while ChatGPT significantly enhances efficiency by streamlining the brainstorming process for science lesson planning, its true value lies in its role as a catalyst for innovation and integrity. By engaging in interactive prompting, pre-service teachers moved beyond passive adoption of AI generated content. Instead, they practiced critical reflection, adjusting AI suggestions to better cater to diverse learning needs and promote effective science learning environments. This process of innovative planning with AI underscores the necessity of high-level teacher agency. While AI helps brainstorm, the pre-service teachers develop their lesson plan ideas based on the context of classroom learning environments.

References

- Davis, C., Bush, T., & Wood, S. (2024). Artificial intelligence in education: Enhancing learning experiences through personalized adaptation. *International Journal of Cyber and IT Service Management*, 4(1), 26-32.
- De Arment, S. T., Reed, E., & Wetzel, A. P. (2013). Promoting adaptive expertise: A conceptual framework for special educator preparation. *Teacher Education and Special Education*, 36(3), 217-230.
- Goodman, J., Handa, V., Wilson, R. E., & Bradbury, L. U. (2024). Promises and pitfalls: Using an AI chatbot as a tool in 5E lesson planning. *Innovations in Science Teacher Education*, 9(1), 1-13.
- Hess, B. J., Cupido, N., Ross, S., & Kvern, B. (2024). Becoming adaptive experts in an era of rapid advances in generative artificial intelligence. *Medical Teacher*, 46(3), 300-303.
- Opre, D. (2015). Adaptive expertise: Efficiency and innovation. *Cognitie, Creier, Comportament/Cognition, Brain, Behaviour*, 19(2).
- Pandey, J. (2019). *Deductive approach to content analysis*. In *Qualitative techniques for workplace data analysis* (pp. 145-169). IGI Global.
- Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. *Transfer of learning from a modern multidisciplinary perspective*, 3(1), 1-51.
- University of South California (2024, October 14). *Is using AI plagiarism? MPA Online professor breaks down generative AI in education*. <https://priceonline.usc.edu/blog/generative-ai-in-education/>

Conception Of Inclusive Learning Environments Related To Biodiversity And Geodiversity: An Innovative Project For Pre-Service Teachers

António Almeida¹, Adriana Cardoso², Filipa Lopes³ and Clarisse Nunes⁴

¹Instituto Politécnico de Lisboa / EDUNOVA.ISPA – Centro de Investigação Interdisciplinar em Educação, ISPA – Instituto Universitário / CICS.NOVA, Universidade Nova de Lisboa, Portugal

²Instituto Politécnico de Lisboa, Portugal

³Instituto Politécnico de Lisboa, Portugal

⁴Instituto Politécnico de Lisboa, Portugal / CI&DEI polo do IPL / CIED

Young people's contact with nature has been decreasing, leading to negative consequences for their physical and psychological well-being. Additionally, education faces the challenge of including all students in the learning process, regardless of their functional, social, linguistic, or cultural characteristics. This project aimed to address these issues and involved 36 primary pre-service teachers in the development of inclusive educational resources focused on areas rich in biodiversity and geodiversity, while also integrating mother tongue content. The participants received scientific and didactic training to design more inclusive digital resources. The evaluation of these resources was based on indicators aligned with recommended principles. On a positive note, the pre-service teachers managed to explore biodiversity, geodiversity, and mother tongue learning in an integrated manner. However, they faced challenges in formulating coherent questions—both closed and open—and in simplifying scientific concepts. The resources were deemed adequate in aspects such as font type, size, and spacing, but the choice of colours often failed to follow inclusive criteria, being more driven by aesthetic preferences. Although the resources included audio texts, the diction was sometimes unsuitable for students with Portuguese as a non-native language. As a result, characteristics such as flexibility, simplicity, and equitable use were only partially achieved. The most promising resources will be tested with state school students later this school year, allowing for further adjustments before being made openly available on the website of the institution responsible for the project.

Keywords: Inclusive Learning, Digital Resources, Biodiversity and Geodiversity.

Introduction

The contact of young people with nature has been declining over recent decades (Freeman & Quigg, 2009; Woolley & Griffin, 2015; Almeida, Rato, & Dabaja, 2023). The consequences of this decline are significant, as such contact contributes to children's mental and physical health, stimulates the development of their cognitive, social, and emotional skills (Frost, Wortham, & Reifel, 2012). Furthermore, this lack of contact appears to reduce knowledge about native biodiversity, affecting not only children but also teachers in training (Almeida, García Fernández, & Strecht-Ribeiro, 2019). Louv (2010) even states that we are experiencing a deficit in nature contact and that the widespread use of the internet and electronic devices has also contributed to this situation.

Nevertheless, technologies are increasingly present in people's daily lives. Their utility and necessity in various contexts have grown, making it impossible for schools to ignore them. Technologies offer an ever-expanding range of applications, and many educational resources are in digital format. As Almenara and Ruiz-Palmero (2018) quote, Information and Communication Technologies (ICT) are one of the key variables in the knowledge society. Whether we are for or against it, the world in which we operate is increasingly technological and will continue to be so,

making their inclusion in schools, and consequently in teaching and learning processes, inevitable.

Education is also currently faced with the challenge of including all students in the learning process, regardless of their functional, social, linguistic, or cultural characteristics. In Portugal, the legal framework for inclusive education (Decreto-Lei n.º 54/2018) assumes that all students have the right to fully participate in a common educational project and to benefit from support that meets their needs, potential, and expectations. So, students from diverse social backgrounds, and a growing number of children from countries where the Portuguese language is not the official language, highlighting the multicultural reality of regular schools nowadays. The decree embodies a vision of inclusive education aligned with Universal Design for Learning (UDL) principles, which provide a framework that can be used to proactively design inclusive, flexible, and welcoming learning environments, considering the variability of all students (Rose & Meyer, 2006). The UDL stands at the forefront of contemporary efforts to create universal access to educational curricula for all students, including those with disabilities. Thus, the need to design more inclusive resources using digital tools aligns with UNESCO's (2009) recommendation of "education for all," as well-designed digital resources can better reach students with diverse characteristics.

The European Agency for Special Needs and Inclusive Education (2015) also emphasizes the fundamental role of inclusive practices in promoting equitable educational opportunities. Similarly, the Sustainable Development Goals (SDGs), specifically Goal 4 (Quality Education) and Goal 10 (Reduced Inequalities), underline the importance of these practices. However, such practices still lack the necessary prominence (Stinken-Rösner et al., 2020), with these authors pointing to a significant gap in research and practice in this critical area.

In this context, the present project aims to contribute to addressing the identified gap and fostering inclusive educational practices.

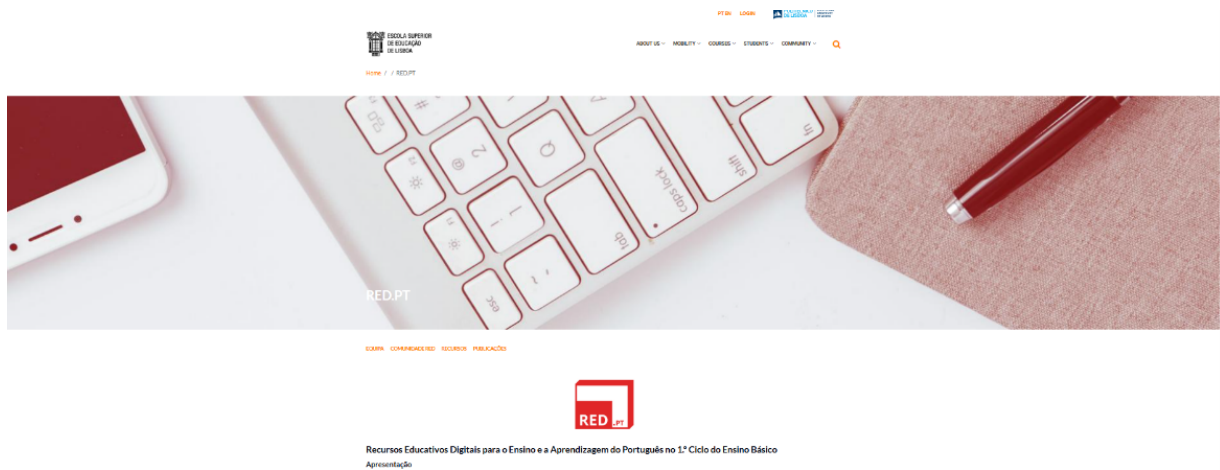
The Project

The present project, with the acronym GreenLink, aimed to address the various challenges outlined earlier. On one hand, it seeks to promote contact with biodiversity and geodiversity, thereby increasing knowledge and interest in nature. On the other hand, it aims to foster this contact through inclusive digital resources, supporting the inclusion of students from diverse backgrounds.

The project is focused on enhancing the training of pre-service teachers by deepening their scientific knowledge of biodiversity and geodiversity, as well as their pedagogical knowledge for designing inclusive digital resources for primary education. This involved applying knowledge acquired through outdoor activities.

At this stage of the project, during which the resources were designed, the goal is to highlight the challenges faced by the pre-service teachers in creating these resources after their training. Based on this evaluation, the resources have since been improved, and their testing with primary school children is scheduled to take place during the current academic year. The goal is to make these resources available on the institution's website, which offers a range of open-access digital resources for teachers.

Figure 1. The webpage of Lisbon School of Education where the resources will be available.



Methodology

The project involved 36 (35 females) primary pre-service teachers enrolled in two master's courses, with a mean age of 23 (SD=2.013), and four higher education professors. The project was developed across two curricular units: one scientific, Earth Science, focused on biodiversity and geodiversity knowledge, which included outdoor activities in nature; and another didactic unit called Integration Projects: Mother Tongue-Science. The latter explored integrative themes between language and science, emphasizing different text types, metaphorical and analogical language, argumentation on socioscientific topics, and children's literature. Additionally, it highlighted principles to follow in constructing assessment items and designing more inclusive digital resources, elements that were evaluated in the resources created. Pre-service teachers were tasked with designing an interdisciplinary inclusive resource for children, focused on the educational exploration of natural sites visited during the Earth Science curricular unit.

In preparing the pre-service teachers, emphasis was placed on the seven principles defined by Santos and Fernandes (2017) for designing inclusive resources: (i) equitable use, which refers to the resource being usable by everyone, regardless of their skills; (ii) design flexibility, which ensures accommodation of diverse user preferences and individual capabilities; (iii) simple and intuitive use, which highlights the resource's functional aspects; (iv) perceptible information, adapted to users' sensory needs, whether visual, auditory, or linguistic; (v) error tolerance, addressing potential negative consequences resulting from unintentional actions during use; (vi) low physical effort, ensuring the resource does not cause user fatigue; (vii) size and space for approach and use, designed to meet accessibility and manipulation requirements, considering factors such as posture, height, or mobility.

The pre-service teachers' training also incorporated principles outlined by the Centre on Inclusive Software for Learning (CISL, n.d.), such as line and letter spacing, font size, and type. It is widely agreed that greater spacing facilitates reading, especially for children and individuals with dyslexia. Additionally, larger font sizes improve comprehension and reading fluency for individuals with dyslexia and benefit those with visual impairments. Optimal font sizes vary across studies, ranging between 18 and 26, while simpler fonts like Arial or Verdana are considered more effective.

The choice of colours for resources also requires careful consideration to combine background and text/image colours that are complementary. These contrast patterns result in more efficient and comfortable visual performance, particularly for individuals with visual impairments or dyslexia.

Eight resources were developed in groups using Genially software. Group working was supported through tutorials, which allowed for iterative improvements to the resources. After their conclusion, each project was graded as part of the assessment of the curricular unit with didactic content. The marks given were based on an analysis grid, which included indicators reflecting the principles of inclusive digital resource design. Both positive aspects and persistent shortcomings were noted, even after the tutorials. Testing with primary school students, planned for the current academic year, will further evaluate and refine each resource.

Results

The resources were focused on several of the sites visited during the outdoor activities, with three being the most frequently chosen. This preference may be attributed to their potential for linking geodiversity and biodiversity and the presence of legends associated with two of them. These legends were seen by the pre-service teachers as a great possibility to integrate mother tongue and science, leading them to explore the natural features of these sites alongside the legends they inspired.

Other language-related contents incorporated into the resources included various text types (informative, narrative), word families, and derivation through prefixation and suffixation, integrated with multiple scientific concepts.

The construction of resource items, despite improvements during the tutorial sessions, posed challenges on several levels. Due to their lack of professional experience, pre-service teachers struggled to use simple language when defining scientific concepts. Difficulties were also identified in the formulation of open questions, which were often overly complex. In other resources, common errors in multiple-choice items included options with significantly unequal lengths and inconsistent grammatical constructions.

The selection of images was a positive aspect, as pre-service teachers had taken many photos and created short videos during the outdoor activities. Additionally, the use of artificial intelligence-generated imagery as a complement to their photos was a common practice among most groups and proved to be successful.

Regarding the seven principles to be considered in the development of digital resources, equitable use, flexibility of use, and simple and intuitive use were partially achieved. In fact, linguistic complexity, the failure to include audio instructions in several cases, poor diction in reading some texts, and the lack of instructions on how to approach certain tasks were evident in almost all resources, although in most cases these problems were found in only parts of the resources. Consequently, the information was not always easily presented for those with auditory or linguistic difficulties.

On a more positive note, the resources ensured that only a few negative consequences occurred to the users when moving forward in a resource, and the instructions, when present, did not require extra effort from the user.

Among other positive aspects, the font type and size chosen in most cases were effective. However, the selection of colours was somewhat neglected in terms of ensuring sufficient contrast, which is crucial for inclusivity. The pre-service teachers seemed to prioritize aesthetic preferences over the needs of all users.

Discussion and Conclusions

From the analysis conducted, it can be concluded that the design of inclusive digital resources is not an easy task and that the possibility of reaching all students is a difficult goal to achieve.

Nevertheless, the training provided allowed pre-service teachers to consider aspects that made the resources more inclusive, although others were overlooked. The attempt to reach students whose first language is not Portuguese proved to be partially successful, as better diction in reading contents revealed a need for improvement. The tendency to approach scientific concepts in a more complex manner also resulted, in part, from the scientific preparation that the pre-service teachers had been undergoing and their difficulty in finding simpler ways to explore these same concepts now for a younger audience. The exploration of biodiversity and geodiversity in an articulated manner proved to be more successful, perhaps because the course they had been attending sought this same articulation. The connection between science content and mother tongue content was also generally successful, with pre-service teachers showing some creativity in the selected articulations. As mentioned, the resources will only be freely accessible after being tested with students from state schools with diverse characteristics. However, the option to access the resources via provided links or QR codes already allows for an initial assessment of their potential and limitations.

Moreover, the process highlighted the importance of ongoing reflection and feedback when designing inclusive resources, emphasizing that initial attempts often need substantial refinement. Observing how different students interact with the materials will provide critical insights into the adaptability and accessibility of the resources. Future revisions are likely to benefit from closer collaboration with experts in language acquisition and special education, helping to address any remaining gaps more effectively. Ultimately, this experience underscores that inclusivity in educational resource design is an iterative process, requiring patience, continuous learning, and a willingness to adjust strategies based on real-world classroom experiences.

Acknowledgement

The present study is part of the project IPL/IDI&CA2024/GreenLink_ESELx, funded by the Instituto Politécnico de Lisboa.



References

- Almeida, A., Rato, V., & Dabaja, Z. F. (2023). Outdoor activities and contact with nature in the Portuguese context: a comparative study between children's and their parents' experiences. *Children's Geographies*, 21(1), 108–122.
- Almeida, A., García Fernández, B., & Strecht-Ribeiro, O. (2019). Primary school children and pre-service teachers' knowledge of Iberian native and African savannah mammals. *Journal of Baltic Science Education*, 18 (6), 833-847.
- Almenara, J. C. & Ruiz-Palmero, J. (2018). Las Tecnologías de la Información y Comunicación para la inclusión: reformulando la brecha digital. *International Journal of Educational Research and Innovation*, 9, 16-30.
- CISL - Centre on Inclusive Software for Learning at CAST (n.d.). *Research on Features*. <https://cisl.cast.org/research/features>
- Decreto-Lei n.º 54/2018, de 6 de julho (Portugal). Diário da República, 1.ª série, n.º 129.
- European Agency for Special Needs and Inclusive Education (2015). *Empowering Teachers to Promote Inclusive Education. Literature Review*. UNESCO.
- Freeman, C., & Quigg, R. (2009). Children's Lives: A Dunedin Study. *Children's Issues: Journal of the Children's Issues Centre*, 13 (1), 35-42.
- Frost, J., Wortham, S., & Reifel, S. (2012). *Play and Child Development*. Pearson.
- Louv, R. (2010). *Last child in the woods. Saving Our Children from Nature - Deficit Disorder*. Atlantic Books.
- Rose, D. H., & Meyer, A. (2006). *A practical reader in universal design for learning*. Harvard Press.

- Santos, C., & Fernandes, S. (2017). Um ambiente virtual interativo de aprendizagem para usuários com/sem limitações sensoriais [comunicação oral]. Anais do V Congresso Internacional da Pró-Inclusão, 'Educação, Inclusão e Inovação, São Paulo, Brasil.
- Stinken-Rösner, L., Rott, L., Hundertmark, S., Baumann, T., Menthe, J., Hoffmann, T., Nehring, A, & Abels, S. (2020). Thinking inclusive science education from two perspectives: Inclusive pedagogy and science education. *Research in Subject-matter Teaching and Learning (RISTAL)*, 3 (1), 30-45.
- UNESCO (2009). *Policy guidelines on inclusion in education*. UNESCO.
- Woolley, H., & Griffin, E. (2015). Decreasing experiences of home range, outdoor spaces, activities and companions: changes across three generations in Sheffield in north England. *Children's Geographies*, 13 (6), 677-691.

Bridging Digital And Physical Worlds: Leveraging AR And AI For Personalized Science Education In Geography And Astronomy

Anastasia Georgiou, Athina Karatza and Apostolia Galani
National and Kapodistrian University of Athens, Greece

This article presents a Preliminary Design Report (PDR) of the prototype of the educational game AcademIa of ARistarchus, which integrates Augmented Reality (AR) and Artificial Intelligence (AI) to revolutionize science education. By addressing persistent challenges such as student misconceptions, the game focuses on enhancing the understanding of complex concepts in geography and astronomy. In an era of rapid technological advancement and increasing environmental concerns, this game aims to prepare students for future challenges. AR allows students to interact with and visualize abstract scientific phenomena, including the greenhouse effect and planetary atmospheres, while AI provides personalized learning experiences by offering real-time feedback. Guided by established learning theories, the game intends to promote critical thinking, problem-solving, and collaboration among students. By providing a technology-enhanced model for science education, AcademIa of ARistarchus lays the foundation for future innovations in educational technology.

Keywords: Augmented Reality, Artificial Intelligence, Science Education

Introduction

The integration of technology into education is vital in today's rapidly evolving digital landscape. Emerging technologies such as AR and AI have shown immense potential for transforming how students engage with complex scientific concepts. In response to the conference theme of transitions and transformations in education, this research explores the design of *AcademIa of ARistarchus*, a prototype that aims to enhance student comprehension of geography and astronomy through innovative digital tools.

A detailed literature review of AR and AI-based educational applications revealed a gap in integrated solutions for these disciplines (Georgiou et al., 2024; Georgiou et al., 2025). While AR enhances understanding by visualizing abstract concepts (Yen et al., 2013), AI provides personalized learning pathways through adaptive feedback (Ayeni et al., 2024). Despite their potential, few initiatives effectively combine these technologies to create immersive educational tools. Addressing this gap is essential for making science education accessible and engaging.

The design of the prototype *AcademIa of ARistarchus* aims to fill this gap by engaging students in immersive scenarios that explore the greenhouse effect on Earth, Mars and Venus, the planets' atmospheres and the role of human activities in Earth's greenhouse effect. The final purpose is to develop an educational game to address research questions, such as whether it enhances students' knowledge about geography and astronomy topics such as the greenhouse effect and planetary atmospheres.

Prototype Design Framework

Core Idea And Purpose

The core idea of the *AcademIa of ARistarchus* is to develop an engaging and educational game centred on planetary atmospheres and the greenhouse effect. The game will integrate modern technological tools, such as AR and AI, to enhance the learning experience. Designed for students

aged 10 to 12, it aims to be both educational and enjoyable. Subsequently, the educational aspects of the game as well as the gameplay features are illustrated.

Learning Objectives And Learning Theories Used In The Design

Learning Objectives

The prototype's learning objectives are designed in alignment with Bloom's Taxonomy (Krathwohl, 2002). Additionally, relevant goals from the Greek Geography curriculum have been incorporated. The knowledge objectives of the game, are organized into three thematic units, each with corresponding learning outcomes, as described following:

Thematic_Unit_1: The atmosphere composition of Earth, Mars, and Venus. Students will be able to:

- Define the atmosphere.
- Compare the atmospheres of Earth, Mars, and Venus.

Thematic_Unit_2: The greenhouse effect on these planets and its role in regulating planetary temperatures. Students will be able to:

- Define the greenhouse effect.
- Justify why Venus has higher temperatures compared to Earth and Mars.
- Explain the greenhouse effect and the role of greenhouse gases in increasing planetary temperatures.
- Compare the temperatures of Earth, Mars, and Venus.

Thematic_Unit_3: The impact of human activity on the greenhouse effect. Students will be able to:

- Describe how human activities influence the greenhouse effect.
- Categorize the human activities that intensify the greenhouse effect.
- Propose solutions to mitigate the impact of human activities on the environment.
- Evaluate whether human activities are the primary cause of the increase in Earth's temperature.
- Critically assess which solutions are most effective in reducing CO₂ emissions.

The prototype also aims to develop collaboration and communication skills through teamwork and problem-solving.

Learning Theories

In AcademIa of ARistarchus, several learning theories were integrated into the prototype design to enhance student engagement, described as following:

- **Constructivism:** Players build knowledge by solving interactive problems, experimenting with scientific scenarios, and utilizing AR/AI features to enhance their understanding of key scientific concepts.
- **Behaviourism:** Players receive rewards for their actions, promoting motivation and reinforcing positive behaviour in learning.
- **Active Learning:** The prototype fosters hands-on engagement through problem-solving, requiring players to make decisions.

- **Social Learning:** Group tasks and teamwork allow players to share strategies and learn collaboratively, enhancing group dynamics.
- **Multidimensional Learning:** The use of AR/AI offers diverse perspectives and learning styles, enriching the educational experience.

Gameplay Features

Prototype Scenario

The story of AcademIa of ARistarchus takes place in the future, where the effects of rising temperatures are still undeniable. The players take on the role of aspiring scientists tasked with developing strategies to protect the planet's future. To accomplish this, they have to explore the greenhouse effects and planetary atmospheres of Mars and Venus and participate in activities to propose solutions to mitigate the impact of human activities on Earth's atmosphere. The players' objective is to answer questions accurately and successfully complete activities, earning stamps for their pAsspoRt and collecting tokens as they progress. The winner is the player who has collected all the stamps in the pAsspoRt and has gathered the most tokens.

Gameplay Mechanics

Drawing from principles of immersive learning and serious gaming (Bopp, 2008; Wang & Huang, 2021), the design of the prototype combines clear educational objectives, real-world challenges, compelling storytelling, interactive gameplay, and effective reward systems to create a meaningful and engaging experience for players. More specifically, the gameplay mechanics implemented are the following:

- **Starting Point:** Players embark on their journey from the island of Samos, Greece, equipped with a grant of tokens from the Cosmos Exploration Institute. They are supported by Aristarchus, an AI assistant tutor, which will help them overcome challenges along the way. Movement across the game board is determined by a micro:bit dice roll, allowing players to advance 0 to 3 steps. This adds an element of excitement to the journey.
- **Interactive Challenges:**
 - (a) The Cosmos Library offers three types of cards: Action_Cards, Exploration_Cards, and AR_Cards. As players progress, they draw cards, each presenting unique, time-bound challenges designed to test their knowledge and critical thinking skills. Action_Cards present collaborative challenges where players have to team up to solve critical questions and activities. Exploration_Cards focus on key scientific topics like the greenhouse effect and human impact on the environment, offering both knowledge-based and thought-provoking challenges. AR_Cards introduce interactive experiences through QR codes. For example, players will have to adjust greenhouse gas levels and observe how these changes alter global temperatures and the greenhouse effect. These cards offer an immersive and hands-on learning experience.
 - (b) Along the way, players must also overcome obstacles like solar storms, which adds complexity to their journey by disrupting their progress.
- **Motivational Rewards:** These tokens (RFID cards) are crucial for progressing and unlocking additional challenges. There are several parts in the game where players gain tokens. Firstly, they earn tokens from the Cosmos Exploration Institute for successfully completing missions. In addition, Action_Cards reward players when collaborating effectively in solving critical questions and activities. Moreover, the Exploration_Cards'

varying difficulty levels provide corresponding tokens. The token-based reward system motivates players to engage deeply with the content, encouraging both learning and progress throughout the game.

- Checkpoints: Key locations on Earth, such as the Arctic, which are highly connected to impacts of human activities, serve as checkpoints where players interact with AR_cards to explore the atmosphere, the greenhouse effect, and the impacts of human activities on these phenomena. Each checkpoint visit is marked with a stamp in the player's pAsspoRt.
- Equipment for Space Missions: layers who invest in higher-tier equipment will gain a strategic advantage in completing challenges and activities, allowing them to solve questions and tasks more thoroughly.
- Winner: Upon completing all tasks, players return back to Samos. The player who successfully finishes the journey and collects every stamp in the pAsspoRt will be ranked on the leaderboard according to the number of tokens earned. Subsequently, the top player will be crowned the winner.

Design Model Of The Prototype

- The game board features an illustrated Global Map of the Earth at its centre with Mars and Venus positioned on opposite sides, symbolizing interplanetary exploration. Players traverse paths across Earth's diverse regions, leading to the other planets for space missions.

The design of the prototype follows a hybrid approach, combining the waterfall model (Ruparelia, 2010) for the initial requirements specification and prototype design, with the agile methodology, as outlined by Al-Azawi et al. (2014), for iterative development and continuous refinement. The initial phase involved specifying the requirements and designing the prototype to ensure alignment with the learning objectives described in the previous section. Subsequently, iterative development cycles were applied, incorporating expert review and theoretical refinement at each stage. This approach, consistent with Agile principles, allows continuous refinement and improvement of the prototype's design.

Methodology for Future Work

The future work for AcademIa of ARistarchus involves several key steps: iterative development cycles were applied, incorporating expert review and theoretical refinement at each stage. This report serves as the Preliminary Design Report of the project. Final development will integrate feedback to refine and expand the game, ensuring it meets quality standards before full-scale implementation.

The evaluation process will involve a mixed-methods approach to collect both quantitative and qualitative data. A sample of 120 students aged 10-12 from different schools will participate in the study. The game's effectiveness will be measured using pre- and post-tests to assess knowledge acquisition, Likert-scale questionnaires to gauge engagement and enjoyment, and semi-structured interviews to gain deeper insights into student experiences. A triangulation approach will enhance validity and reliability by cross-referencing questionnaires, observations, and interviews. Quantitative data will be analysed statistically using t-tests to compare mean scores, while qualitative data will be thematically coded to identify recurring patterns and insights.

The study, scheduled for September 2025, aims to evaluate how the AR/AI game enhances students' understanding of the greenhouse effect and planetary atmospheres. Findings will guide

future improvements to optimize the game's learning potential and user experience.

Discussion

Integrating AR and AI in education enhances students' understanding of complex concepts through interactive learning. However, challenges such as technological complexity, accessibility, and equipment requirements must be addressed. Some key questions to consider are as follows: What are the pros and cons of AR/AI in teaching? How can these technologies be optimized for better learning outcomes? What challenges emerge when teaching complex scientific concepts/phenomena? How could the game be adapted to different age groups?

References

- Al-Azawi, R., Ayesh, A., & Obaidy, M. A. (2014, January). Towards agent-based agile approach for game development methodology. In 2014 *World Congress on Computer Applications and Information Systems (WCCAIS)* (pp. 1-6). IEEE. <https://doi.org/10.1109/WCCAIS.2014.6916626>
- Ayeni, O. O., Al Hamad, N. M., Chisom, O. N., Osawaru, B., & Adewusi, O. E. (2024). AI in education: A review of personalized learning and educational technology. *GSC Advanced Research and Reviews*, 18(2), 261-271. <https://doi.org/10.30574/gscarr.2024.18.2.0062>
- Bopp, M. M. (2008). *Storytelling and motivation in serious games*. Part of the Final Consolidated Research Report of the Enhanced Learning Experience and Knowledge Transfer-Project ELEKTRA (Report No. 27986)
- Georgiou, A., Karatza, A., Babassidis, G. E., & Galani, A. (2024). Sol(ar) system: Augmented reality in teaching astronomy concepts in the Geography curriculum. *Proceedings of the Panhellenic Conference on Science Education and New Technologies in Education*, 13, 608–614. <https://doi.org/10.12681/codiste.6841>
- Georgiou, A., Karatza, A., Mavrikai, E., & Galani, A. (2025). Applications of Artificial Intelligence in the Physical Sciences: A Literature Review. *Proceedings of the Panhellenic Conference on Science Education and New Technologies in Education*, 14(1). <https://doi.org/10.12681/codiste.7642>
- Krathwohl, D. R. (2002). A Revision Bloom's Taxonomy: An Overview. *Theory into Practice*, 41(4), 212–218.
- Ruparelia, N. B. (2010). Software development lifecycle models. *ACM SIGSOFT Software Engineering Notes*, 35(3), 8-13.
- Yen, J. C., Tsai, C. H., & Wu, M. (2013). Augmented reality in the higher education: Students' science concept learning and academic achievement in astronomy. *Procedia-social and behavioural sciences*, 103, 165-173. <https://doi.org/10.1016/j.sbspro.2013.10.322>
- Wang, C., & Huang, L. (2021). A Systematic Review of Serious Games for Collaborative Learning: Theoretical Framework, Game Mechanic and Efficiency Assessment. *International Journal of Emerging Technologies in Learning*, 16(6). <https://doi.org/10.3991/ijet.v16i06.18495>

Talking To Or Talking Through Teachers - An Analysis Of A Digital Teacher's Guide

Charlotta Nordlöf¹, Susanne Engström² and Pernilla Sundqvist³

¹Linköping university, Sweden

²KTH Royal Institute of Technology, Sweden

³Mälardalen university, Sweden

Teaching materials not only support students but also support teachers in how they transfer their subject knowledge to teaching. Teacher guides are often digital and can address teachers by talking through or talking to the teacher, thus supporting teachers. In this paper a digital teacher's guide for grades 7-9 is analysed to see how it communicates with the teachers. The teacher's guide in technology education was analysed by content analysis in two steps. Firstly, on whether it talks to the teacher or through the teacher, and secondly the units that talk to the teacher were analysed by using six design principles. The results show that most of the guide (75 %) is text that talks through the teacher, e.g. by giving direct instructions for the teacher on how to carry out the teaching. In the remaining sections of the guide, text that talks to the teacher, there are examples of all six design principles. These sections include the purpose and structure of each chapter, describe key content, highlight challenges, suggest teaching adaptations and provide definitions of important terms and concepts. The analysed guide does not take advantage of the benefits of digitalization. Teacher guides should enable teachers to adapt and expand upon the material. It is important to design teacher guides with more text that speaks to teachers, not just through them.

Keywords: Teaching materials, K-12 Education, STEM education

Introduction

Teaching materials, such as textbooks and teacher guides, play a central role in the way teachers transfer their subject knowledge into teaching. Teaching materials support teachers and influence the choices they make in planning and delivering lessons (Oates, 2014). Teachers with low subject knowledge are particularly dependent on teaching materials. Teacher guides in particular have the potential to support teachers in different ways.

Teacher guides are often digital. The digital format provides the possibility to redesign the material, i.e. to do what was not possible with the analogue format (Puentedura, 2020; Blundell et al., 2022). Remillard (2000) distinguishes between how the teacher guides communicate with teachers. Teacher guides that talk through the teacher and tell the teacher what to do, while teacher guides that talk to the teacher explain the underlying ideas of these suggestions.

In this paper, we analyse a digital teacher's guide for grades 7-9. The purpose is to give an account of the ways in which the examined teacher's guide communicates with the teachers, in terms of talking through and talking to the teacher.

Theoretical perspectives

Educative Curriculum Materials

Remillard (2000) points out that teacher guides commonly talk through teachers but argues that they should be designed to speak to teachers. However, van Steenbrugge and Ryve (2018) emphasise that teacher guides that talk through teachers can also be a support, especially for teachers who feel insecure about the subject or are unaccustomed to teaching situations.

Distinctive for educative curriculum materials (talk to teachers) is that they include educative features specifically designed to support teachers when adapting their teaching, both regarding content and how to conduct teaching. Based on Davis and Kracjik's (2005) guidelines for how and why teacher guides should be designed, and studies where versions of these guidelines were tested (Arias et al., 2016; Davis et al., 2014), Davis et al. (2017) constructed six design principles to characterise teacher guides in science education. These can shortly be described as:

1. Educative features should suggest ways to adapt lessons without losing their purpose.
2. Educative features should explain the reasoning behind pedagogical ideas and support teachers in adopting these ideas through contextualised examples.
3. Educative features should help make important content visible and explain it, showing how it connects across lessons.
4. Educative materials should offer various types of features to meet diverse needs.

Principles 5-6 are excluded since they are not used in this study.

The SAMR-Model

The SAMR-model (Substitution, Augmentation, Modification, Redefinition) is a model for integrating digital technology in education. Puentedura (2009) describes it as a guide for how teachers can design tasks and learning processes when using digital technology in education and states that digital tools can either be used to improve existing analogue teaching practices or to fundamentally transform them to a digital format. He describes four levels of integration:

- Substitution: at this level, the digital format is simply replacing the analogue format and offers no improvement or functional change.
- Augmentation: existing practice is improved by the digital tool.
- Modification: Instead of teaching in the same way and using the same tasks, the digital tool is used to adapt and change the tasks.
- Redefinition: The digital technology allows for and is used to create new tasks, and is transforming learning.

Learning activities categorised as substitution or augmentation are understood to improve or support learning, whereas those categorised as modification or redefinition are considered to fundamentally transform the learning experience (Romrell et al., 2014, with reference to Puentedura).

Since we study teacher's guides as potentially educative for teachers, we find the SAMR-model useful for analysing how the design of the teacher's guide benefits from the digital format in order to enhance teachers' understanding of how to organise and enact the teaching.

Method

The Teacher's Guide

The teacher's guide of a widely used curriculum material for technology education in grades 7-9 in Sweden is analysed (Lovén & Citrohn, 2022). It is a digital resource and can be bought as a complement to the textbook. Its size is equivalent to about 200 pages in book form, consisting of an introduction and six chapters based on the same structure. First, introduction, connections to the curriculum, and assessment. This is followed by sub-chapters corresponding to the textbook, covering the content (facts and theoretical information) and projects (tasks and assignments) contained in the book.

Analysis

To study how the teacher's guide benefits from the digital format, we looked at all the features of the digital guide and categorised them according to the four levels of the SAMR-model.

To examine how the teacher's guide communicates with the teacher, content analysis was performed in two steps. First, we coded each unit (phrase, sentence or paragraph) based on whether it talks *to* the teacher or *through* the teacher (Remillard, 2000). Then, we further coded the units that talk *to* the teacher using an adapted version of the design principles created by Davis and colleagues (2017). We used the original design principles 1-4 but modified 5-6 to fit the technology subject, inspired by Robert's vision I and II: like other researchers in technology education (e.g. Stoor & Popov, 2021).

1. In technology education, students should try, analyse, and evaluate working methods in various technical areas, such as design processes and life cycle analyses. The teacher's guide should provide examples of how work in different technical disciplines is carried out in practice and how this can be part of the teaching.
2. One objective for technology education is to prepare students to become conscious citizens in a society where technology plays a central role. The teacher's guide must support teachers in motivating the importance of technology education for society and everyday life and provide examples that highlight technical knowledge for making decisions at both individual and societal levels.

All three researchers coded the first chapter of the teacher's guide individually and subsequently discussed the coding agreement until consensus. The remaining chapters were divided among the researchers.

Findings

First, we present the findings related to the digital format of the teacher's guide. Second, the findings are summarised based on whether the text talks *through* or talks *to* the teacher.

The Digital Format Of The Teacher's Guide

This teacher's guide is digital, but it resembles a printed version. It does not offer any additional functionality compared to what an analogue guide can do. Thus, it is a substitute for a printed guide and placed at the substitution level in the SAMR-model.

Text That Talks *Through* The Teacher

Approximately 75% of the teacher's guide is text that talks *through* the teacher, where a large part is the answers to the questions in the textbook. Also, there are direct instructions for the teacher on how to perform the teaching. These are clear and explicit: *Let the students read the text on pages 48–55 silently or read the text aloud to them* or *Use the worksheet and have students write a list of various technical solutions*.

Overall, most of the guide talks directly *through* the teacher, who can essentially use it as a script. Alternatively, it can be employed as a practical step-by-step guide: *Distribute the worksheet to the students, then ask them to answer the questions*.

Text That Talks *To* The Teacher

The teacher's guide contains about 25 % text that talks *to* the teacher, mainly in the Introduction sections. These sections outline the purpose and structure of each chapter, describe key content, highlight challenges, suggest teaching adaptations and provide definitions of important terms and concepts. Additionally, the guide connects the content to the syllabus of technology education.

Another part of the teacher's guide where there is text that talks *to* the teacher is in the Assessment sections. These sections include a rubric providing examples of student responses at passing level and at higher grade levels. Findings and examples from each principle are presented in Table 1.

Table 1. Design principles 1 to 6 with examples.

Design principle (DP)	Findings	Example
DP1 – suggestions on how lessons can be adapted in several ways without losing the intended purpose.	Not extensively covered, occurs as suggestions for adjustments based on time, alternative projects, and strategies to address common difficulties students face.	<i>Students often find it challenging and complicated to create drawings. If they are given sufficient time, they usually manage to produce drawings that show simple objects with two views and measurements after some practice. Allow students extra time for this section.</i>
DP2 – the reasoning behind, and contextualised concrete examples of, the pedagogical ideas included in the teacher's guide.	Not very present overall. However, it is visible in the assessment section included in each chapter. It is also found in some introductory sections.	<i>Many students prefer to make drawings on graph paper, which can be a good starting point. However, when adding measurements and drawing lines between the grids, a drawing on graph paper can become unclear. [...] Make students aware of this and let them try drawing their sketches and drawings on white paper.</i>
DP3 – supporting teachers' subject knowledge, making important content visible, and explaining it.	Appears to varying degrees. E.g. in the chapter about technological development work, there is a strong focus on essential subject content. Also found in some introduction sections.	<i>The circular motion refers to the frequent need to go back and improve technical solutions during development</i>
DP4 – support teachers' different needs for themselves and their students.	Examples are found when the guide points out content that may be difficult for students.	<i>Understanding the system boundary can be difficult.</i> (No support to the teachers on how to teach about this or explain boundaries in different situations.)
DP5 – provide teachers with examples of how work within technical disciplines is carried out in practice and can be part of the teaching.	Links to professions as a way of making lessons meaningful and relevant to the world outside school.	<i>Students may think that block programming is something you do in school, but block programming is also common in industry, for example, to control industrial robots or manufacturing machines.</i>
DP6 – support teachers to make the importance of technology for societal and everyday life visible.	Found in the chapter addressing the consequences of technology, also examples of historical perspectives and technological systems.	<i>Emphasise to students that the best approach is to minimise waste, meaning we should only buy products we truly need.</i>

Discussion

The analysed guide predominantly speaks *through* the teacher, as teacher guides often do (Remillard, 2000). The structure and content in this guide provide direct instructions that leave little room for adaptation, e.g. for diverse student groups. Fixed structures and specific instructions in guides can limit students' exposure to diverse perspectives and the broadness of the subject, as a previous study found that a textbook does not sufficiently cover a subject and that supplementary materials are needed (Engström et al., 2023). Lack of explicit explanations for methods can hinder teachers' understanding and ability to make informed decisions. Teacher guides can be a support for insecure teachers (van Steenbrugge & Ryve, 2018), but over-reliance on guides can lead to inadequate explanations and support for students. Even though guides should not replace teacher education or professional development, they can be useful for substitutes and as part of professional development if designed educatively.

The teacher's guide in this study is digital, which is a common format of teacher guides within this field. The digital format offers great opportunities and possibilities for redefining the materials (Puentedura, 2020), such as visualisations in the form of images, simulations, and video clips, both when talking *to* and *through* the teacher. However, the current guide does not take advantage of the benefits of digitalisation. Rather, it acts like a substitute for the analogue format (c.f. Puentedura, 2020). It mainly consists of text and images that are copies from the textbook. This is because the structure of the guide is designed to support the teacher by speaking through the teacher and linking directly to the book's pages, tasks, and so on.

Teacher guides should empower teachers to adapt and expand upon the material by speaking *to* teachers, not just *through* them. Teacher education programs should provide opportunities to explore different types of guides, encouraging reflection on their approaches and implications for teaching practice. This fosters flexible and reflective teaching practices.

References

- Arias, A. M., Davis, E. A., Marino, J. C., Kademian, S. M., & Palincsar, A. S. (2016). Teachers' use of educative curriculum materials to engage students in science practices. *International Journal of Science Education*, 38(9), 1504-1526.
- Blundell, C. N., Mukherjee, M., & Nykvist, S. (2022). A scoping review of the application of the SAMR model in research. *Computers and Education Open*, 3, 100093.
- Davis, E. & Krajcik, J. (2005). Designing Educative Curriculum Materials to Promote Teacher Learning. *Educational Researcher* 34(3), 3-14.
- Davis, E. Palincsar, A. Arias, A. Bismack, A. Marulis, L. & Iwashyna, S. (2014). Designing Educative Curriculum Materials: A Theoretically and Empirically Driven Process. *Harvard Educational Review*, 84(1).
- Davis, E., Palincsar, A., Smith, S., Arias, A., & Kademian, S. (2017). Educative Curriculum Materials: Uptake, Impact, and Implications for Research and Design. *Educational Researcher* 46(6), 293–304.
- Engström, S., Sundqvist, P., Nordlöf, C., & Klasander, C. (2023). Approaches to industrial processes in technology textbooks. In S. Davies, M. McLain, A. Hardy & D. Morrison-Love (Eds), *The 40th International Pupils' Attitudes Towards Technology Conference Proceedings 2023*, 31 October-3 November, Liverpool John Moores University, Liverpool, UK.
- Lovén, K., & Citrohn, B. (2022). *Stella teknik 7–9*. Natur och kultur.
- Oates, T. (2014). *Why textbooks count*. Cambridge Assessment.
- Puentedura, R.R. (2020, January 25). SAMR - A research perspective. *Ruben R. Puentedura's Blog*. http://hippasus.com/rpweblog/archives/2020/01/SAMR_AResearchPerspective.pdf
- Puentedura, R.R. (2009). *As we may teach: educational technology, from theory into practice*. Apple.
- Remillard, J. T. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *The elementary school journal*, 100(4), 331-350.
- Romrell, D., Kidder, L. C., & Wood, E. (2014). The SAMR model as a framework for evaluating Learning. *Journal of Asynchronous Learning Networks*, 18(2).

- van Steenbrugge, H., & Ryve, A. (2018). Developing a reform mathematics curriculum program in Sweden: Relating international research and the local context. *ZDM*, *50*, 801-812.
- Stoor, M., & Popov, O. (2021). Societal Aspects of Technology: The Swedish Lower Secondary School Curricula in International Comparison. *Techne series: Research in sloyd education and crafts science. A*, *28*(2), 140-148.

The Relationship Between Critical Thinking And Generative Artificial Intelligence: A Two-Way Street For Science Education

Kalliopi Meli¹, Yaron Lehavi², Itzhak Aharon³

¹University of Patras, Greece

²David Yellin Academic College of Education, Israel

³Jerusalem College of Technology (Machon Lev), Israel

The emergence of Generative Artificial Intelligence (GenAI) has presented new opportunities and challenges for developing critical thinking skills in science education. From the learning perspective, students' critical thinking skills are challenged during content creation (input) and evaluation (output) when interacting with GenAI. This process establishes a continuous loop that implicitly requires critical thinking to achieve desirable outcomes. From a teaching perspective, GenAI can be used as a tool to explicitly develop and assess critical thinking skills. During an exploratory seminar, we invited participants (N=26, working in pairs) to (a) rank science-related questions by the degree of critical thinking they are likely to provoke, (b) evaluate alternative GenAI-generated rankings as 'good', 'weak', or 'random' critical thinking, and (c) reconsider their own rankings after exposure to GenAI outputs and group reflection. We examined how difficult it is to judge critical thinking demands even without GenAI support, and how GenAI suggestions can shift participants' decisions. Descriptively, most participants revised at least one ranking, and a slight majority reported that GenAI influenced their final choices. We discuss these findings, arguing that GenAI can support critical thinking when positioned as an object of critique and a partner for structured reasoning rather than an answer engine.

Keywords: Critical Thinking; Generative Artificial Intelligence; Science Education

Introduction

The importance of critical thinking (CT) in science education is reflected in curriculum standards and educational goals across many countries and jurisdictions, highlighting its perceived value in developing scientifically literate individuals capable of reasoned analysis and decision-making in scientific contexts (Bailin, 2002). While CT is often cited as a key competency, its definition has become too broad to be helpful in science education, which stresses what scientists do: generate, test, and evaluate claims, data, and theories. This can be suggested as a definition of CT in the context of science education (Schmaltz et al. 2017). Such CT skills involve awareness of cognitive biases to avoid them, conducting impartial analyses, and using logical reasoning and evidence to support claims. These skills may enable students to question established social norms, challenge conventional wisdom, and expose societal inequities (O'Reilly et al., 2022; Larson et al., 2024).

Technology provides innovative tools and platforms for developing such CT skills, necessitating new teaching and assessment approaches in digital environments (Saadé et al., 2012). This necessity has become increasingly evident with the emergence of Generative Artificial Intelligence (GenAI) (Verma et al., 2023). CT and GenAI interaction encompasses both the learning and teaching perspectives. During learning, GenAI can challenge students' CT skills in (a) content creation (student input) and (b) content evaluation (GenAI output) (Larson et al., 2024). This continuous loop can extend as far as the student permits. While this interactive process between students and GenAI necessitates engaging students' CT to achieve desirable outcomes, the teaching perspective is also viable in utilizing GenAI specifically and explicitly

for CT skills development and assessment (Noroozi et al., 2024).

Given the relative novelty of this technology, only a limited number of research studies have investigated the impact of GenAI on students' critical thinking abilities (Bahroun et al., 2023), predominantly through indirect means, and even fewer have exploited GenAI as a tool for teaching and evaluating critical thinking. In the pertinent literature, scholars need to extend this research to address these questions and elucidate the potential outcomes that may be derived from the relationship between critical thinking and GenAI (Larson et al., 2024). In the context of science education, such research may yield a profound perspective for the development of future scientists and the advancement of broad scientific literacy.

This paper reports on an ESERA 2025 exploratory seminar that treated GenAI not primarily as a tutor but as a stimulus for reflection on what counts as 'critical thinking' in science-related questioning. Participants completed a brief activity sequence: initial ranking of four climate–agriculture related questions by how much critical thinking they provoke, classification of three GenAI-generated rankings as good/weak/random, followed by post-reflection re-ranking and a short survey on whether GenAI influenced their final decisions. We use the seminar outputs to illustrate (a) the intrinsic difficulty of evaluating critical thinking demands and (b) the ways GenAI can reshape judgments even in a meta-level task about critical thinking itself.

Theoretical framework

Critical thinking is a crucial 21st-century skill that enhances scientific understanding, promotes democratic citizenship, and develops higher-order cognitive abilities (Santos, 2017). Together with the above definition of CT skills in science education, it has the potential to humanize sciences and situate them in broader contexts, fostering inquiry and creativity through questioning, critique, and argumentation. It prepares individuals to tackle complex future challenges, supports informed decision-making on socio-scientific issues, and involves complex reasoning and evidence-based argument construction. CT skills are increasingly needed and challenged in an environment progressively influenced by advancing GenAI technologies (Zhai et al., 2024).

Prompting, or prompt engineering, provides a natural-language-based interface for GenAI, facilitating iterative and interactive exchanges between humans and GenAI systems in a collaborative creative process (Oppenlaender et al., 2024). Initial prompts are evaluated, and GenAI responses are assessed against students' intended outcomes. If GenAI's output is inadequate, the prompt is modified and re-evaluated until the desired result is achieved (Cain, 2023). This process implicitly promotes learning by reflecting on students' objectives, GenAI's capabilities and limitations, and language nuances.

While GenAI offers numerous opportunities to enhance education, critical thinking remains crucial to ensuring its effective, ethical, and pedagogically sound implementation in learning environments (Giannakos et al., 2024). The potential for biased or misleading information in GenAI outputs underscores the importance of critical thinking skills in recognizing and addressing these issues. The limitations of GenAI in nuance, context, and common sense necessitate human interpretation and evaluation to ensure meaningful learning experiences. Additionally, there are potential ethical concerns regarding data privacy, security, and intellectual property rights when using GenAI in educational settings that students should learn to address critically.

The overreliance on solutions generated by GenAI, without applying critical thinking skills in input development and output evaluation, may compromise students' capacity for independent

learning and conceptual understanding (Thiga, 2024). Therefore, there is a growing focus on developing effective teaching strategies and assessment methods to foster skills, including critical thinking, across all educational levels, from early childhood to higher education (O'Reilly et al., 2022) and teacher education (Verma et al., 2023). Recent research in the STEM education context demonstrates some attempts to develop frameworks for explicitly cultivating critical thinking utilizing GenAI tools as skill-teaching or assessment assistants (Lee & Low, 2024; Pantazatos et al., 2024; Shanto & Ahmed, 2024; Thiga, 2024).

Shanto & Ahmed (2024) proposed a framework that offers structured guidance for focused inquiry, gathering diverse perspectives, evaluating responses, synthesizing insights, and reflective learning to analyse and build upon GenAI-generated output critically. Lee and Low (2024) demonstrated the application of GenAI tools in parallel workshop activities, utilizing the Paul-Elder framework to foster critical thinking development among students. In concurrence, Thiga (2024) posited that critical thinking skills can be cultivated through GenAI by employing systematic and in-depth interrogation of tools' outputs. Similarly, Pantazatos et al. (2024) investigated the integration of a GenAI-based virtual assistant, with particular emphasis on its influence on students' critical thinking abilities.

Research consistently shows that GenAI can both enhance and inhibit students' critical thinking. Applied within structured educational frameworks, GenAI tools notably improve critical thinking (Lee & Law, 2024; Pantazatos et al., 2024). However, misuse can reduce motivation for independent inquiry and reflection, indicating a need for careful integration into curricula (Thiga, 2024). Research suggests that a deliberate, well-designed framework can significantly elevate critical thinking (Shanto & Ahmed, 2024). These findings highlight the importance of strategically guiding GenAI use to foster critical thinking skills in education.

Research studies on the impact of critical thinking on students' interaction with GenAI (learning perspective) and the utilization of GenAI to develop critical thinking skills (teaching perspective), explicitly call for further investigations to explore this intersection (e.g., Bahroun et al., 2023; Larson et al., 2024). While there have been attempts to identify the role of critical thinking in several STEM education domains, there is a deficiency in addressing the potential and challenges specifically in science education fields, where critical thinking is important (Bailin, 2002; Santos, 2017).

To begin addressing this gap, we designed an exploratory seminar that examines the two-way relationship between critical thinking (CT) and GenAI in science education: first, how participants' CT is mobilised when they evaluate and justify the quality of GenAI-generated reasoning (learning perspective), and second, how GenAI's outputs can shape or steer participants' subsequent judgements about what counts as "critical-thinking demanding" work (teaching perspective). In the context of this exploratory seminar study, we developed two research questions to be used as a principled starting point for future, more tightly operationalised research questions considering the participants' responses, feedback, and concerns.

RQ1 (CT → GenAI evaluation): To what extent can participants discriminate between GenAI-generated rankings that are presented as reflecting different levels of critical thinking, and how consistently do they make these classifications?

RQ2 (GenAI → CT evaluation): How does exposure to GenAI-generated rankings influence participants' subsequent judgements of critical-thinking demand?

Method

Seminar Participants

Participants ($N=26$) were attendees of the ESERA 2025 conference who voluntarily selected this exploratory seminar session. They comprised science education researchers and practitioners of diverse ages and national backgrounds, bringing a range of perspectives to the activity. All participants worked in self-selected pairs to facilitate collaborative discussion and decision-making during the activities.

Description Of The Seminar Activity

The activity was delivered through a structured online form (developed in Typeform) and consisted of five sequential phases. The scenario and the related questions were generated by GenAI to situate the task in a science-education context. An expert-derived rubric, based on the Holistic Critical Thinking Scoring Rubric ((Facione & Facione, 1994; Ramlo, 2019), was used to judge the CT demand of each question on a 1–4 scale (1 = least CT required, 4 = most CT). ChatGPT 5.0 was instructed to assess the CT provoked by these questions using the rubric and the result was calibrated by the researchers. Ultimately, the four questions had increasing CT complexity in the order: Q1 (score 1) < Q3 (2) < Q2 (3) < Q4 (4). This “ground truth” ranking (Q1 → Q3 → Q2 → Q4, from less to more provoking) was later used to generate example AI rankings and to evaluate participant responses.

Phase 1: Scenario and Initial Ranking
Participants first read a brief GenAI-generated scenario on climate change and agriculture, then individually ranked four GenAI-generated questions (Q1–Q4) from 1 (provokes the least critical thinking) to 4 (provokes the most critical thinking). This captured each pair’s baseline judgment of which questions demanded more critical thinking.

The scenario text read:

Uneven shifts in climate and atmospheric CO₂ are altering agricultural conditions. Evidence indicates both potential gains and losses that vary by method, timescale, and context, with cascading effects across production, trade, and food access amid uncertainty. Regional responses differ as growing seasons and crop suitability shift. Inferences about magnitude and direction often depend on data sources (observations versus model projections) and underlying methodological choices.

The four questions were:

1. What potential agronomic benefits associated with climate change have been observed or projected, and how can these be empirically evaluated across regions?
2. What lines of evidence both support and challenge the hypothesis that climate change may improve agricultural conditions in some locales, and how might methodological choices influence these findings?
3. How might shifts in climatic suitability and growing seasons affect crop selection, global trade flows, and food distribution, and what key uncertainties should be considered?
4. Which evidence-based decision-making framework could guide policies that both leverage potential local benefits and mitigate risks to food security, and how would trade-offs be justified?

Phase 2: Evaluation of GenAI Rankings

Next, participants reviewed three anonymized GenAI-generated rankings of the same questions, labeled “Ranking A,” “Ranking B,” and “Ranking C.” They were told that one ranking was generated by a “Good Critical Thinker,” one by a “Weak Critical Thinker,” and one randomly, though the identities were not disclosed. Participants were asked to classify each ranking as *Good*, *Weak*, or *Random*, and identify which one they believed represented the best critical thinking ordering.

Phase 3: Justification

Participants were prompted to justify their choice of the “Good CT” ranking by writing a brief explanation of the criteria they used. This open-ended response captured their reasoning process and CT criteria (e.g., evidence use, synthesis, complexity) in their own words.

Phase 4: Re-Ranking

After completing the evaluation and justification phases, participants returned to their original question ranking and were invited to revise it based on the GenAI examples and their reflections. This phase allowed researchers to observe any shifts in judgement influenced by GenAI exposure.

Phase 5: Reflection On GenAI Influence

Finally, participants answered a Yes/No question on whether the GenAI-generated rankings influenced their revised decisions. If they answered “Yes,” they were prompted to elaborate on what factors caused their change in ranking. This phase captured participants’ perceived influence of GenAI on their thinking and decisions.

Data Collection and Analysis

The online form recorded all responses, yielding both quantitative and qualitative data. Quantitative measures included each pair’s initial and final question rankings (for comparing changes), and their classification of the GenAI rankings (to assess identification accuracy of “Good/Weak/Random” rankings). Qualitative data included the textual justifications for selecting a particular ranking as “Good CT” and the open-ended reflections on GenAI’s influence.

For analysis, we calculated simple descriptive statistics: the number of participants who correctly identified the AI’s good-vs-poor CT rankings, the proportion who revised their question ordering, and the degree of alignment between participant rankings and the expert rubric before vs. after the GenAI intervention. The open-ended responses were reviewed to extract common themes regarding how participants judged critical thinking and why they did or did not alter their rankings.

Results

Discrimination Of GenAI-Generated Rankings (RQ1)

Overall, participants showed moderate success in distinguishing the quality of the AI-proposed question rankings. As summarized in Table 1, 9 out of 13 pairs (69%) correctly identified the “Good Critical Thinker” ranking generated by GenAI (the sequence aligning with the expert rubric). The remaining 4 pairs misidentified a suboptimal ranking as the “good” one – notably, 3 pairs were confused by the reversed (weak) ranking, labelling it as the best. Participants were most adept at spotting the nonsensical ordering: 11 pairs (85%) recognized the random ranking for what it was. Identification of the weak ranking was more challenging (only 8 pairs, 62%, labeled it correctly), with some of its misinterpretation being mistaken for either a good or random order. In fact, just over half of the groups (54%) managed to classify all three AI-generated rankings correctly. These results indicate that while many participants could discern obviously

poor vs. strong question orderings, a substantial minority struggled, particularly confusing the weakest ranking with the optimal one.

Table 1. Participant accuracy in classifying GenAI-generated rankings (N=13 pairs).

GenAI Ranking Type	Correct Identifications (n)	Accuracy (%)
“Good” ranking	9	69%
“Random” ranking	11	85%
“Weak” ranking	8	62%
All three classified correctly	7	54%

Participants’ open-ended justifications shed light on the criteria they used to evaluate these rankings. Many participants based their judgments on the depth and complexity of thinking that questions in each ranking would demand. For example, responses frequently mentioned the inclusion of multiple perspectives, evaluation of evidence, and higher-order reasoning as signs of a “good” critical thinking question order. One pair explicitly referenced Bloom’s taxonomy levels, noting that questions asking “how” and requiring synthesis or solution planning signaled higher cognition, whereas questions focusing on recall of details indicated lower critical thinking. Others emphasized the importance of evidence-based reasoning: e.g. “Being able to collate data, interpret data, and apply knowledge” was cited as a hallmark of the stronger ranking. In contrast, rankings where questions were simply factual or one-dimensional were seen as reflecting weaker critical thinking. These qualitative insights confirm that participants were attending to meaningful aspects of the questions (such as breadth of analysis, need for justification, and consideration of uncertainties) when judging the AI-generated rankings as strong or weak in CT terms.

Influence Of GenAI On Final Rankings (RQ2)

Exposure to the GenAI-generated examples had a notable impact on many participants’ own question rankings. After evaluating the GenAI’s rankings and reflecting, a majority of the pairs (8 of 13, ~62%) revised their original ranking of the four questions before final submission. In other words, only five pairs chose to keep their initial ordering unchanged through the activity. To gauge whether these revisions moved participants closer to the expert-defined ranking, we compared the alignment of each pair’s ordering with the rubric-based “correct” order pre- and post-activity. Table 2 shows that alignment improved overall. Initially, only 1 out of 13 pairs (8%) had an exact match to the rubric’s ideal order (Q1 < Q3 < Q2 < Q4). After the GenAI intervention and re-ranking, this number rose to 5 pairs (38%). Similarly, a simple alignment metric (total rank difference from the rubric, on a 0–8 scale) showed a substantial reduction: before the activity, participants’ rankings deviated from the expert order by an average of 3.1 points; after revision, the average deviation dropped to 1.4. This indicates that many participants’ final rankings were more in line with the expected difficulty order of the questions, suggesting a positive influence of the GenAI examples on their critical evaluation of the questions’ demands.

Table 2. Alignment of participants’ rankings with expert rubric (initial vs. final).

Alignment Metric	Initial Ranking	Final Ranking
Mean total difference (0–8)	3.1 ± 1.8	1.4 ± 1.1
Pairs with exact rubric order (n)	1 (8%)	5 (38%)

Qualitative self-reports from the reflection stage help explain how and why GenAI influenced (or did not influence) participants’ thinking. Among the 13 pairs, 7 explicitly answered “Yes” – the GenAI rankings influenced their final decisions – while 6 reported no influence. Those who acknowledged influence often said that reviewing the GenAI-generated orderings prompted them to re-evaluate the criteria for critical thinking. For instance, one pair described how the AI

examples made them “think about the characteristics of critical thinking” and led them to consider “the depth of thinking and the amount of analysis” required by each question, even attempting to tie their reasoning to Bloom’s taxonomy. Others noted that the contrast between their initial ranking and the GenAI “good” ranking revealed gaps or mistaken assumptions in their original thinking, prompting adjustments for a more evidence-based or multi-faceted approach. In a few cases, participants admitted the GenAI rankings caused confusion or doubt: “The results led me to get confused,” one pair wrote, while another even joked that seeing an unexpected ordering made them “panic and question everything [we] knew about critical thinking.” Nevertheless, even these instances of surprise or uncertainty often triggered deeper reflection on why a given question might be harder or easier than initially thought.

Conversely, the 6 pairs who reported no influence tended to stick with their initial reasoning despite the GenAI input. Some of them had already ranked the questions in an order close to the rubric’s, so the GenAI examples may have simply affirmed their choices. Others indicated scepticism about the GenAI outputs, noting (for example) that the GenAI did not provide a rationale for its rankings, which limited its persuasive power. One pair commented that none of the AI-proposed orders matched exactly what they initially thought, but they did not fully trust those examples enough to overhaul their original ranking. This highlights that while GenAI can introduce new perspectives, participants with strong prior convictions or justified reasoning might resist changing their answers without a clear justification.

In the seminar activity, participants engaged with critical thinking in two distinct but complementary ways. First, they deployed critical thinking when analysing the cognitive demands of the four questions—drawing on reasoning, comparison, and synthesis to judge which required deeper thought. Second, they evaluated critical thinking when interpreting the logic and quality behind GenAI-generated rankings, reflecting on the reasoning implicit in each proposed sequence. This dual engagement allowed participants to not only practice critical thinking skills but also refine their understanding of what constitutes critical thinking in a science education context.

Discussion

This study contributes to a growing body of work examining how GenAI can be harnessed to support the development of critical thinking in science education. As previous research has noted, while CT is a central aim of science learning, its definition and operationalization often remain ambiguous (Bailin, 2002; Schmaltz et al., 2017). The seminar activity addressed this challenge by creating a concrete task in which participants had to weigh the cognitive demand of different questions and assess GenAI’s ability to do the same. The difficulty many experienced in making these judgments reinforces the complexity of critical thinking as a construct, aligning with concerns raised by O’Reilly et al. (2022) and Larson et al. (2024) about the need for more precise and context-sensitive tools.

The outcomes also resonate with literature that has framed GenAI as a partner in learning rather than a passive tool (Lee & Low, 2024; Shanto & Ahmed, 2024). Participants who altered their rankings described how the AI examples prompted deeper reflection on what constitutes a high-quality, thinking-oriented question. Rather than adopting AI outputs uncritically, they weighed the underlying reasoning, compared it to their own, and sometimes changed course. This kind of iterative, evaluative engagement reflects the kind of interaction that Oppenlaender et al. (2024) and Giannakos et al. (2024) have highlighted as central to using GenAI to enhance learning.

Still, the variation in responses underscores the need for robust scaffolding and transparency

when incorporating GenAI in critical-thinking tasks. Some participants reported confusion or distrust toward the AI's rankings, particularly in the absence of explanations. This confirms the concerns raised by Thiga (2024) regarding potential misuse or misinterpretation of GenAI outputs and points to the importance of embedding such tools within pedagogically grounded frameworks. Educators must help learners develop both the cognitive and digital literacies required to critically assess GenAI-generated content.

Looking ahead, future studies should explore how this kind of seminar-based approach could be scaled, adapted for student populations, or integrated into pre-service teacher training. Further research might incorporate control conditions, longitudinal designs, or direct assessments of critical thinking before and after engagement with GenAI. In practical terms, educators might adopt similar ranking and justification tasks in classrooms to promote deeper reasoning about what makes a question cognitively demanding and how AI-generated reasoning can be critically assessed.

Ultimately, this work highlights a promising path forward: rather than viewing GenAI as a challenge to critical thinking, we can frame it as a site where critical thinking is activated, tested, and refined. By making GenAI both a partner in reasoning and an object of analysis, science educators can better prepare students (and themselves) for navigating the increasingly complex interplay of knowledge, technology, and judgment in a digital age.

Conclusions

The exploratory seminar provided valuable insights into how science education professionals engage with GenAI-generated content when evaluating and re-evaluating the critical-thinking demand of science-related questions. Participants were moderately accurate in distinguishing between question rankings of varying quality, and their reasoning often drew on meaningful indicators such as depth of analysis, evidence use, and synthesis. Moreover, many participants revised their original rankings after reflecting on the GenAI-generated examples, resulting in greater alignment with expert expectations. These findings suggest that while distinguishing levels of critical thinking remains a nuanced and challenging task, especially when mediated by GenAI, structured activities such as this can foster improved judgment and metacognitive awareness.

Crucially, the results point to the value of positioning GenAI not as an authoritative answer engine but as a prompt for reflection and critique. Participants' engagement with the AI rankings influenced their decisions in ways that highlighted both the potential and the limitations of GenAI in supporting critical thinking. This two-way interaction, where GenAI both elicits and is subject to critical evaluation, offers a promising direction for future teaching and research practices in science education.

Acknowledgement

The authors would like to thank all the participants of the ESERA 2025 exploratory seminar for their enthusiastic engagement and thoughtful contributions.

References

- Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis. *Sustainability*, 15(17), 12983. <https://doi.org/10.3390/su151712983>

- Bailin, S. (2002). Critical Thinking and Science Education. *Science and Education*, 11(4), 361–375. <https://doi.org/10.1023/a:1016042608621>
- Cain, W. (2023). Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. *TechTrends*, 68(1), 47–57. <https://doi.org/10.1007/s11528-023-00896-0>
- Facione, P. A., & Facione, N. C. (n.d.). *Using the Holistic Critical Thinking Scoring Rubric to train discovery of evidence of critical thinking*. Insight Assessment. Retrieved from <https://www.insightassessment.com>
- Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour & Information Technology*. <https://doi.org/10.1080/0144929x.2024.2394886>
- Larson, B. Z., Caza, A., Moser, C., Colombo, L. A., & Muehlfeld, K. (2024). Critical Thinking in the Age of Generative AI. *Academy of Management Learning & Education*. <https://doi.org/10.5465/amle.2024.0338>
- Lee, C. C., & Low, M. Y. H. (2024). Using genAI in education: the case for critical thinking. *Frontiers in Artificial Intelligence*. <https://doi.org/10.3389/frai.2024.1452131>
- Noroozi, O., Farrokhnia, M., Banihashem, S. K., & Soleimani, S. (2024). Generative AI in Education: Pedagogical, Theoretical, and Methodological Perspectives. *International Journal of Technology in Education*, 7(3), 373–385. <https://doi.org/10.46328/ijte.845>
- Oppenlaender, J., Linder, R., & Silvennoinen, J. (2024). Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2024.2431761>
- O'Reilly, C., Devitt, A., & Hayes, N. (2022). Critical thinking in the preschool classroom - A systematic literature review. *Thinking Skills and Creativity*, 46, 101110. <https://doi.org/10.1016/j.tsc.2022.101110>
- Pantazatos, D., Grammatikou, M., & Maglaris, V. (2024). Enhancing Soft Skills in Network Management Education: A Study on the Impact of GenAI-Based Virtual Assistants. 1–5. <https://doi.org/10.1109/educon60312.2024.10578597>
- Ramlo, S. (2019). *Critical thinking and physics concepts: Indicators for general education science course assessment*. *The Journal of General Education*, 68(3–4), 191–215. <https://doi.org/10.5325/jgeneeduc.68.3-4.0191>
- Saadé, R. G., Morin, D., & Thomas, J. D. E. (2012). Critical thinking in E-learning environments. *Computers in Human Behaviour*, 28(5), 1608–1617. <https://doi.org/10.1016/j.chb.2012.03.025>
- Santos, L. F. (2017). The role of critical thinking in science education. *Online Submission, Journal of Education and Practice*, 8(20), 160-173.
- Schmaltz, R. M., Jansen, E., & Wenckowski, N. (2017). Redefining critical thinking: Teaching students to think like scientists. *Frontiers in psychology*, 8, 459.
- Shanto, S. S., Ahmed, Z., & Jony, A. I. (2024). Enriching Learning Process with Generative AI: A Proposed Framework to Cultivate Critical Thinking in Higher Education using Chat GPT. *Tuijin Jishu/Journal of Propulsion Technology*, 45(1), 3019-3029.
- Thiga, M. M. (2024). Generative AI and the development of critical thinking skills. *IRE Journals*, 7(9), 83–90.
- Verma, G., Campbell, T., Melville, W., & Park, B.-Y. (2023). Navigating Opportunities and Challenges of Artificial Intelligence: ChatGPT and Generative Models in Science Teacher Education. *Journal of Science Teacher Education*, 34(8), 793–798. <https://doi.org/10.1080/1046560x.2023.2263251>
- Zhai, X., Ma, W., & Nyaaba, M. (2024). Can Generative AI and ChatGPT Outperform Humans on Cognitive-Demanding Problem-Solving Tasks in Science? *Science & Education*. <https://doi.org/10.1007/s11191-024-00496-1>

Exploring The Use Of ChatGPT In Science Lesson Planning

Giorgos Peikos and Dimitris Stavrou

Department of Primary Education, University of Crete, Greece

Generative AI (GenAI) is increasingly recognized for its potential to support teachers in tasks such as lesson planning. However, its limitations, including inaccuracies in generated responses, make it essential for teachers to critically evaluate its outputs. This study explores the use of GenAI in science lesson planning through the lens of Pedagogical Content Knowledge (PCK). The research is structured in two phases. In the first phase, we examined the use of ChatGPT in science lesson planning by engaging with the tool as researchers, focusing on how lesson plan suggestions generated reflect components of PCK. Building on the findings from this phase and insights from related literature, the second phase investigates how primary school teacher students utilize ChatGPT for lesson planning informed by PCK within the context of a university seminar. Findings from the first phase, based on qualitative content analysis, indicate that when ChatGPT is guided by PCK-informed layer prompts and supported by relevant reference texts, it generates more pedagogically appropriate lesson plan suggestions. Results from the second phase show that, after being introduced to PCK and prompt engineering strategies, primary school teacher students increasingly employed layer prompts, particularly Natural Language Instruction and Follow-up prompt types, and incorporated PCK components in more meaningful ways. Overall, the findings support the conceptualization of ChatGPT as a co-designer in science lesson planning, rather than as a substitute for teachers' pedagogical judgement.

Keywords: artificial intelligence, pre-service teacher education, instructional design

Introduction

Generative AI (GenAI) is widely recognized for its generative capabilities in producing human-like text (Chan & Colloton, 2024). In the field of education, GenAI tools such as ChatGPT show potential to support teachers in lesson planning and creating teaching materials (Calvo-Utrilla et al., 2025; Choi, 2025; Cooper, 2023; ElSayary, 2023; ElSayary et al., 2025). However, concerns about inaccuracies, biases, and ethical issues in its outputs (Avraamidou, 2024; Calvo-Utrilla et al., 2025) highlight the importance of critical evaluation and responsible use by educators (Blonder & Feldman-Maggor, 2024; UNESCO, 2023). Prompt engineering has emerged as an essential skill for educators to optimize GenAI outputs, forming a critical part of their Technological Pedagogical Content Knowledge (TPACK) in the age of AI (Blonder & Feldman-Maggor, 2024; UNESCO, 2023). For instance, effective strategies may include instructing GenAI to adopt a specific persona, using step-by-step layer prompts also referred to as chain of thought prompting, and supplying reference texts such as scientific papers to reduce inaccuracies and minimize fabricated responses (Atlas, 2023; Blonder & Feldman-Maggor, 2024; Knoth et al., 2024; Wei et al., 2023). Nevertheless, human oversight remains imperative to address these limitations (UNESCO, 2023).

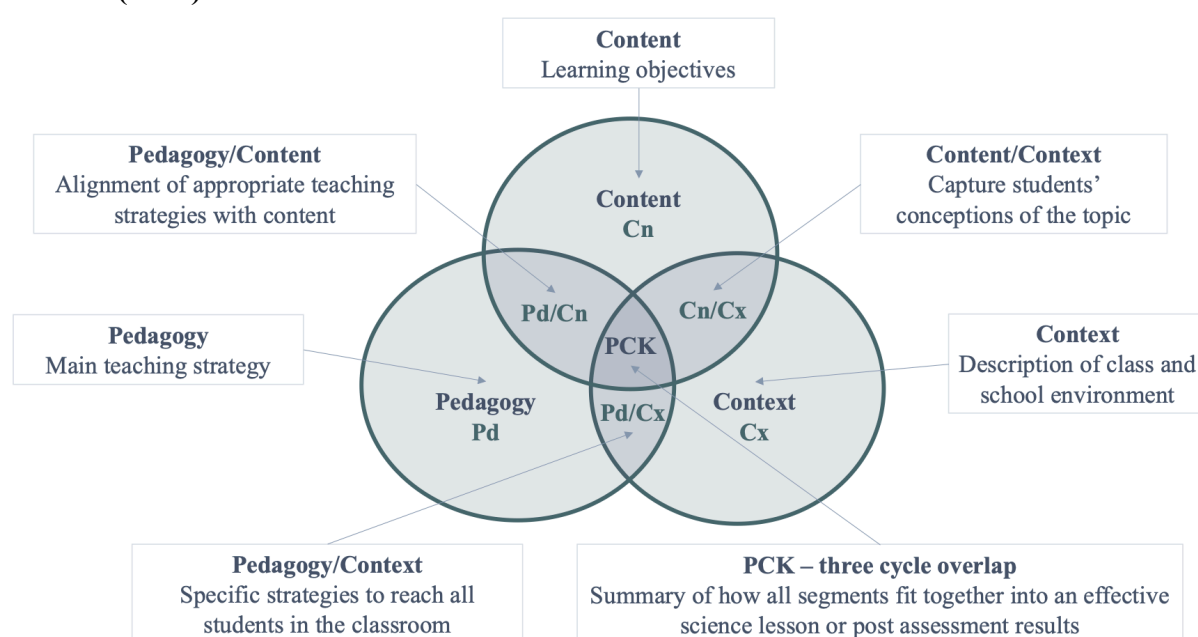
Research on the use of GenAI in lesson planning demonstrates its ability to create lesson plans for specific science topics based on inquiry-based teaching approaches (Cooper, 2023; Großmann et al., 2025), with evidence showing that its output improves when guided by iterative prompts (Hashem et al., 2024; Peikos & Stavrou, 2025). From the teachers' perspective, GenAI tools such as ChatGPT are valuable for generating innovative ideas, enhancing lesson quality through engaging activities, supporting personalized learning, aligning lessons with real-life scenarios to boost student motivation and saving time. Nevertheless, both researchers and teachers emphasize that GenAI should serve as an assistant rather than a substitute for teachers. They also raise

concerns about the reliability of its outputs, such as potential weaknesses in content accuracy and in the provision of meaningful pedagogical suggestions (Cooper, 2023; ElSayary, 2023; ElSayary et al., 2025; Großmann et al., 2025; Uğraş et al., 2025).

Building on the above literature findings, it is evident that GenAI holds significant potential for assisting with lesson planning, although this potential is interwoven with limitations regarding the accuracy of information. Therefore, it is essential for teachers to critically evaluate GenAI outputs when using it to support lesson planning (UNESCO, 2023), drawing on PCK (Blonder et al., 2025; Feldman-Maggor et al., 2024; Großmann et al., 2025; Peikos & Stavrou, 2025) as a foundational knowledge base to enhance its responsible use.

PCK has been established as a core construct for understanding teaching effectiveness and informing teacher professional development. PCK is “the knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students for enhanced student outcomes” (Gess-Newsome, 2015, p. 36). Over more than four decades of PCK research following Shulman’s seminal work (Shulman, 1986), several PCK models have been proposed, leading to the formulation of the Refined Consensus Model of PCK (Carlson et al., 2019). Across these models, key PCK components have been consistently identified, including knowledge of students’ understanding, instructional strategies, curriculum, assessment, context, content knowledge, pedagogical knowledge, and orientations toward teaching science. These components are intrinsically interconnected in practice, and their integrated enactment constitutes the core of effective science teaching (Carlson et al., 2019; Chan & Hume, 2019; Gess-Newsome, 2015; Magnusson et al., 1999; Park & Oliver, 2008). The Refined Consensus Model (Carlson et al., 2019) conceptualizes PCK across three interrelated realms: collective PCK, referring to shared professional knowledge; personal PCK, representing the knowledge and skills teachers develop through experience; and enacted PCK, referring to the knowledge and skills drawn on when planning, teaching, and reflecting on instruction and student outcomes.

Figure 1. Venn diagram of PCK for science lesson planning, adapted from Otto and Everett (2013).



Focusing more closely on enacted PCK, particularly when teachers plan instruction (Carlson et al., 2019), Otto and Everett (2013) have developed an innovative teaching strategy for introducing PCK concepts to pre-service elementary teachers by utilizing a three-circle Venn diagram for

science lesson planning (Figure 1). The diagram featured overlapping circles representing Pedagogy (e.g., teaching strategies), Content (e.g., learning objectives), and Context (e.g., classroom environment), Pedagogy/Context (e.g., specific strategies to reach all students in the classroom), Pedagogy/Content (e.g., aligning appropriate teaching strategies with content), Content/Context (e.g., capturing students' conceptions of the topic) as well as the PCK overlap summarizing how these components fits together into an effective science lesson (Otto & Everett, 2013).

Lesson planning is a key component of teachers' professional competence, requiring the application of PCK to design and justify pedagogical decisions (Großmann & Krüger, 2024). In the era of GenAI, within the context of a teacher-AI co-design process for lesson planning (UNESCO, 2023), PCK is proposed as a valuable framework for crafting effective prompts and critically evaluating the pedagogical quality of GenAI outputs (Blonder & Feldman-Maggor, 2024; Peikos & Stavrou, 2025). To explore these opportunities and challenges, this research is structured in two phases. In the first phase, we investigate the use of ChatGPT in science lesson planning, with the goal of identifying the characteristics of ChatGPT-generated lesson plan suggestions through the lens of PCK. Building on the findings from the first phase and insights from related literature, in the second phase, we explore how primary school teacher students integrate ChatGPT into their lesson planning process within the framework of PCK. The research is guided by the following research questions (RQs):

- RQ1: What are the characteristics of lesson plan suggestions produced in ChatGPT's outputs during collaboration between researchers and ChatGPT, in terms of PCK?
- RQ2: How do primary school teacher students utilize ChatGPT for lesson planning, focusing on (a) the types of prompts they used and (b) the PCK components incorporated into the prompts?

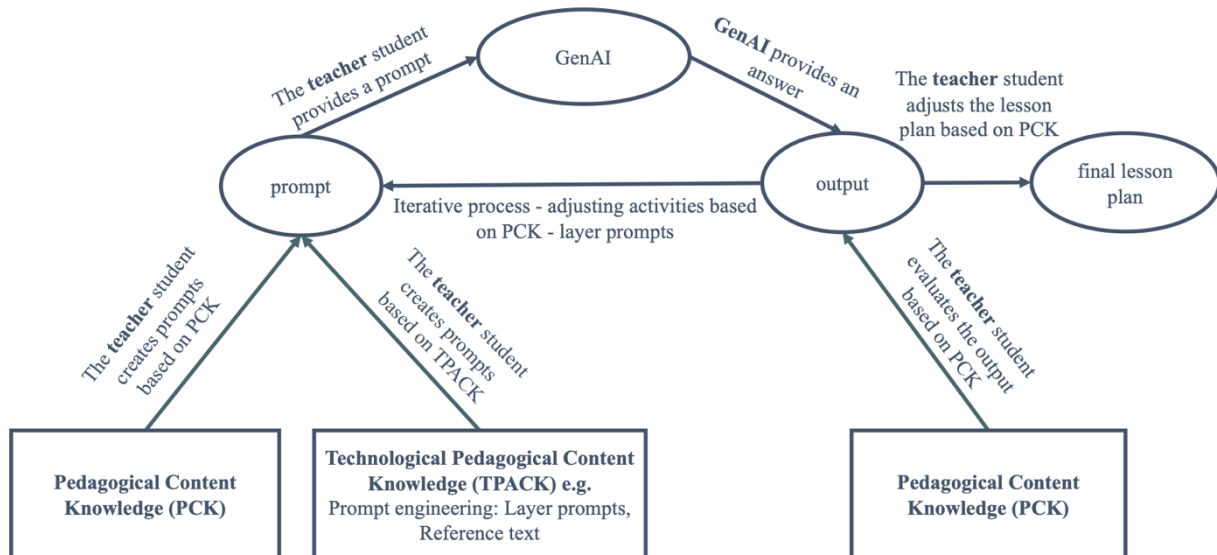
Methods

The Context

In the first phase, the researchers examined the use of ChatGPT for science lesson planning through the lens of PCK by using the tool themselves to design lesson plans. To this end, four interactions with ChatGPT were designed to generate lesson plans, each differing in prompt engineering strategies and PCK integration. ChatGPT was instructed to adopt the persona of an expert in science lesson planning for primary schools (Atlas, 2023; Chan & Colloton, 2024). The first interaction used a single prompt to create a lesson plan based on the 5Es instructional model for the topic of floating, sinking, and density, targeting 5th-grade primary school students. The second interaction introduced specific learning goals related to declarative and procedural knowledge. The third and fourth interactions employed layer prompts in a conversational manner to address several PCK aspects. For instance, ChatGPT was asked to describe activities to uncover students' misconceptions about floating and sinking (Content/Context overlap). The fourth interaction differed by providing a reference text i.e., an open-access scientific paper relevant to the topic. This phase resulted in the formulation of guidelines for the use of GenAI as a co-designer of lesson plans (Peikos & Stavrou, 2025) (Figure 2). Briefly, the guidelines indicate that teachers could initiate the process by formulating prompts that integrate components of PCK, including explicit learning goals, instructional strategies, and students' conceptions. Drawing on prompt engineering, a TPACK-related feature, they are encouraged to use layer prompts in a conversational manner, offering clear, step-by-step guidance that embeds PCK elements. In addition, ChatGPT is provided with relevant reference texts, such as scientific papers related to the instructional topic. The resulting output should be critically examined by the teacher through

the lens of their PCK and iteratively refined to address issues such as student misconceptions and grade-level appropriateness. Finally, teachers could select and adapt the proposed activities to their specific instructional context, leading to the development of the final lesson plan (Peikos & Stavrou, 2025).

Figure 2. Guidelines for student teacher–GenAI co-design of lesson plans based on PCK (adapted from Peikos & Stavrou 2025).



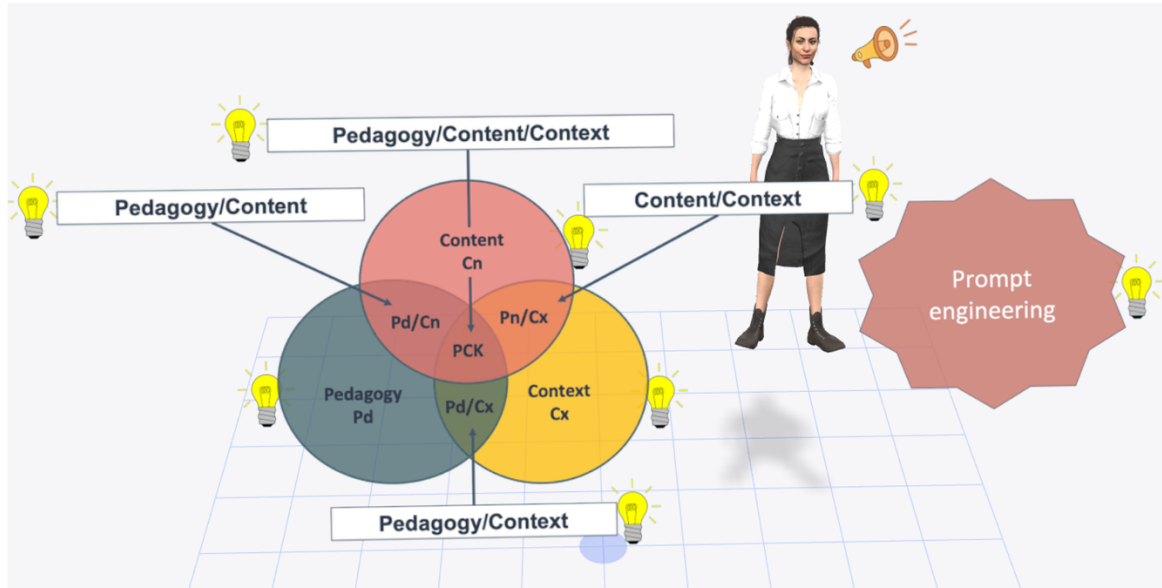
In the second phase, within the context of a semester-long university seminar, primary school teacher students were introduced to the use of GenAI for science lesson planning through the free web-based version of ChatGPT (GPT-4o). Based on the guidelines derived from the first phase, activities were designed to support the use of ChatGPT as a co-designer. The participants studied material related to the benefits and challenges of using ChatGPT, prompt engineering, and PCK. In addition, the researchers developed a pedagogical agent, Anna (Figure 3), within an augmented reality environment, which provided primary school teacher students with guidance during the development of lesson plans, particularly with respect to prompt engineering and the PCK diagram proposed by Otto and Everett (2013). For example, by selecting the Content/Context option (Figure 3), participants receive information, both in written and oral form, regarding the meaning of the Content/Context overlap, which focuses on the identification of students' ideas related to the Content, as well as on the elicitation of students' alternative ideas or prior knowledge. Moreover, they are provided with an example of a related prompt, such as: "Describe a structured activity aimed at eliciting students' alternative ideas about [insert the concept/phenomenon]".

Participants

The participants were 18 primary school teacher students in their third and fourth year of study, who attended most of the seminar sessions.

Data Collection And Analysis

For RQ1 (first phase), the data sources comprised texts produced through four interactions between the researchers and ChatGPT during the lesson planning process, with a focus on ChatGPT's outputs.

Figure 3. Augmented reality pedagogical agent Anna.

For RQ2 (second phase) qualitative data were collected through the discussions between primary school teacher students (working in teams of two) and ChatGPT during the lesson planning process. The focus was on the prompts formulated by the teacher students. Data were collected at three time points, as participants were engaged in the creation of three lesson plans: (a) before the introduction to GenAI and PCK, (b) after the introduction to GenAI and PCK, and (c) after the provision of feedback by the instructor on the primary school teacher students–AI co-design lesson planning process.

For RQ1 and RQ2b, the texts were analysed deductively in order to categorize coding units into predefined PCK components (Chaitidou et al., 2018; Otto & Everett, 2013; Peikos & Stavrou, 2025), namely Pedagogy, Content, Context, and their overlaps, with additional subcategories formed inductively from the data (Table 1).

Table 1. Data coding for RQ1 and RQ2b: PCK components, based on Chaitidou et al. (2018), Otto & Everett (2013) and Peikos & Stavrou (2025)

Category	Description	Examples of subcategories
Content	Learning objectives	Declarative knowledge, Procedural knowledge,
Context	Descriptions of class and school environment including resources and time constraints	Grade level, List of resources, Students' characteristics, Time constraints
Pedagogy	Main teaching strategy	5E format
Content/ Context	Capture students' conceptions of the topic	Activity to identify specific students' misconceptions, Activity to identify students' prior knowledge/ideas/common misconceptions
Pedagogy/ Context	Specific strategies to reach all students in the classroom	Activity to address common misconceptions, Activity to address a specific misconception
Pedagogy/	Alignment of appropriate	Activity aligned with specific learning objective,

Content	teaching strategy with content	Activity with specific teaching strategy for the content
PCK	Summary of how all segments fit together/proposed assessments	General assessment activity, Assessment activity based on specific characteristics

For RQ2a, concerning the types of prompts used by primary school teacher students, the texts were analysed deductively based on the prompt types proposed by Şimşek (2025), which distinguishes six types of prompts: (a) Natural Language Instruction (NLI), (b) Question (Q), (c) Follow-up (FU), (d) Extend (E), (e) Search Query (SQ), and (f) Emotive Feedback (EF), as well as combinations of these types such as Natural Language Instruction & Follow-up (Table 2).

Table 2. Data coding for RQ2a: Types of prompts (Şimşek, 2025, p. 11413).

Category	Definition
Natural Language Instruction (NLI)	A prompt formulated as an explicit instruction for a task in natural language.
Question (Q)	A prompt that begins with a question word and is formulated as a direct or indirect interrogative statement.
Follow-up (FU)	A prompt related to the response given by ChatGPT in the previous step.
Extend (E)	A prompt that encourages ChatGPT to extend or continue an input data example.
Search Query (SQ)	A prompt, such as a search query, typed into a search engine to access information.
Emotive Feedback (EF)	A prompt that provides no contribution to the task and offers positive or negative feedback to ChatGPT as if it were human.

Results

In the following paragraphs we present the preliminary findings of this study. Concerning RQ1 “How do lesson plans generated by ChatGPT reflect the components of PCK?” we found that, in relation to the overlap between Content/Context, ChatGPT demonstrated potential to create activities aimed at uncovering students’ prior knowledge and misconceptions. In the absence of detailed prompts, the model often provided generic activities. However, when guided by specific layer prompts, ChatGPT produced more structured and targeted strategies, such as concept mapping. Concerning Pedagogy/Context overlap layer prompts enhanced ChatGPT’s ability to propose strategies for addressing student misconceptions and accommodating learning diversity. In early interactions, the model’s outputs lacked targeted activities to address specific misconceptions. However, after providing related prompts, ChatGPT began generating inquiry-based tasks designed to challenge common misconceptions, such as the belief that “heavier objects always sink”. Furthermore, regarding the Pedagogy/Content overlap in its initial responses, ChatGPT frequently suggested both qualitative and quantitative approaches to

teaching the concepts of floating and sinking. However, when considering the Context, i.e., the grade level of primary school students, literature suggests that qualitative approaches, such as visual representations, are more appropriate than quantitative approaches, such as the formula $d=m/v$ (Zoupidis et al., 2021). With the use of layer prompts and the provision of reference text, ChatGPT's outputs in the fourth interaction became more aligned with students' grade, emphasizing qualitative instructional approaches. A detailed presentation of the results related to RQ1 is provided in Peikos and Stavrou (2025).

Regarding RQ2a, which focuses on the types of prompts used by primary school teacher students, a noticeable shift was observed during the second and third lesson planning processes. In particular, prompts identified as Natural Language Instruction became more prevalent. This suggests that, once introduced to GenAI, primary school teacher students increasingly framed their interaction with ChatGPT through explicit task-oriented instructions expressed in natural language. For example, one team formulated the following prompt:

Take the role of an expert in science education for the 5th grade science lesson. Help me create a one-hour lesson plan using simple everyday materials, to teach the phenomenon of floating and sinking by explaining it through density, according to the 5E instructional model...

Similarly, Natural Language Instruction combined with Follow-up was observed more frequently during the second and third lesson planning processes. This category refers to prompts formulated as explicit instructions in natural language that build directly on the response provided by ChatGPT in a previous step. For instance, a team of students wrote:

For the activity in the engagement phase, according to what we said above, I would like it to be in the form of a short story that narrates a problem related to the phenomenon of sinking and floating, so that students' alternative ideas can become evident based on the solutions they will give. Make it in the form of a worksheet that will include an image showing the problem of the story.

Regarding RQ2b "How do primary school teacher students utilize ChatGPT for lesson planning, focusing on the PCK components incorporated into the prompts?" notable changes were observed in the Content category. In the first lesson planning process, teacher students' prompts that explicitly referred to specific learning objectives were limited. Most prompts either addressed the topic in general terms or asked ChatGPT to define the learning goals. For example, a team of students submitted the following prompt to ChatGPT:

[Create a teaching scenario for the subject of Physics. It is intended for 5th grade primary school students. The topic is floating-sinking and density.] Write the objectives in detail.

In contrast, during the second and third lesson planning processes, teacher students clearly incorporated specific learning objectives into their prompts. These objectives were connected to both declarative and procedural knowledge, indicating a more explicit integration of Content. For instance, a team of students submitted the following prompt to ChatGPT:

[...Help me create a lesson plan on density and its relationship to floating and sinking for 5th-grade students, based on the 5E instructional model ...] The learning objectives are: (a) for students to understand the rule governing whether an object floats or sinks based on its density; (b) for students to state that density is a characteristic property of a material; and (c) for students to be able to apply strategies for controlling variables...

Examining more closely the overlaps among the PCK components, notable patterns were identified. With regard to the Pedagogy/Context overlap, most primary school teacher students'

prompts in the second and third lesson planning processes asked ChatGPT to propose resources to support the planned activities. These resources included, indicatively, videos, simulations, models, illustrations, storytelling, and worksheets. For example, one team of primary school teacher students wrote:

Propose the use of supportive educational materials such as simulations and models that contribute to the understanding of the concept of heat transfer by conduction in metals.

In addition, during the second and third lesson planning processes, primary school teacher students more frequently asked, through their prompts, for activities that would connect to students' experiences, foster collaboration, and elicit students' interest. Regarding the Content/Context overlap, prompts explicitly related to students' prior knowledge and misconceptions were very limited in the first lesson planning process. In contrast, during the second and third lesson planning processes, primary school teacher students asked ChatGPT to develop activities explicitly aimed at identifying students' prior knowledge and misconceptions. For example, a team of students wrote:

In the Exploration phase, propose one or more hands-on activities during which students will express their initial conceptions and reveal their misconceptions, focusing on the misconception that the temperature of water rises above 100 degrees Celsius while it is boiling.

Conclusion

In the first phase, the study examined the use of ChatGPT for science lesson planning through the lens of PCK by engaging the tool in a series of lesson planning interactions conducted by the researchers. These interactions varied in both prompt engineering strategies and levels of PCK integration. The findings demonstrate that incorporating PCK elements into prompts, using layer prompts, and providing ChatGPT with reference texts related to the instructional topic may enhance the integration of PCK components in GenAI-generated outputs related to the lesson plans. In this phase, the PCK framework appeared meaningful for supporting responsible interactions with ChatGPT in the context of teacher and GenAI lesson co-design. It provided a coherent knowledge base for both crafting pedagogically informed prompts and critically evaluating ChatGPT's responses in relation to science lesson planning (Peikos & Stavrou, 2025).

Building on these insights, the second phase examined how primary school teacher students utilized ChatGPT for science lesson planning within the context of a university seminar, following their introduction to prompt engineering strategies and the PCK framework. The findings indicate that the students were able to formulate prompts grounded in their PCK, with particular emphasis on overlaps such as Pedagogy/Context, as well as Content/Context. In parallel, they engaged with aspects of TPACK, particularly in relation to prompt engineering. Rather than limiting their interaction to ChatGPT's initial responses, the students engaged in an iterative prompting process guided by their PCK, progressively refining the generated outputs. Through the use of layer prompts, mainly in the form of Natural Language Instruction and Follow-up prompt types, they adapted the proposed activities to better align with pedagogical goals and classroom context.

Grounded in the Refined Consensus Model of PCK (Carlson et al., 2019), the findings indicate that teachers' engagement with GenAI tools for lesson planning, as an activity situated within enacted PCK realm, involves the design, reflection, and justification of lesson plans drawing primarily on their personal PCK, while also being shaped by collective PCK and the particularities of the learning context. Within this process, research-based literature on the

teaching and learning of specific science concepts, as an aspect of collective PCK, may be provided to GenAI tools as reference material, thereby informing lesson plan suggestions and potentially reducing inaccuracies. In parallel, responsibility for critically evaluating GenAI-generated outputs remains with the teacher. Drawing on their personal PCK, teachers are required to evaluate the pedagogical soundness of proposed activities, the accuracy of the scientific content, and their suitability for the specific classroom context. Taken together, these findings support the view that GenAI should be conceptualized as a co-designer in the lesson planning process, rather than as the primary designer of instructional plans.

The results underscore the importance of preparing teachers to engage with GenAI through a strong grounding in PCK, combined with prompt engineering skills. Teacher education and professional development programmes should therefore explicitly address this interplay, as it may support more responsible and pedagogically informed integration of GenAI into science lesson planning. Moreover, GenAI can be positioned not merely as a time-saving assistant, but as a tool for supporting teachers' reflection on their own PCK (Blonder et al., 2025). When used in a critical way and potentially supported by pedagogical agents that scaffold interaction between teachers and GenAI, such engagement may help raise teachers' awareness of their pedagogical reasoning and instructional decision-making in science lesson planning.

References

- Atlas, S. (2023). *ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI*. https://digitalcommons.uri.edu/cba_facpubs/548
- Avraamidou, L. (2024). Can we disrupt the momentum of the AI colonization of science education? *Journal of Research in Science Teaching*, 61(10), 2570–2574. <https://doi.org/10.1002/tea.21961>
- Blonder, R., & Feldman-Maggor, Y. (2024). AI for chemistry teaching: responsible AI and ethical considerations. *Chemistry Teacher International*, 6(4), 385–395. <https://doi.org/10.1515/cti-2024-0014>
- Blonder, R., Feldman-Maggor, Y., & Rap, S. (2025). Are They Ready to Teach? Generative AI as a Means to Uncover Pre-Service Science Teachers' PCK and Enhance Their Preparation Program. *Journal of Science Education and Technology*, 34(6), 1301–1310. <https://doi.org/10.1007/s10956-024-10180-2>
- Calvo-Utrilla, M., Paños, E., & Ruíz-Gallardo, J.-R. (2025). Chatbots in Science Education: A Scoping Review of Early Empirical Evidence. *Journal of Science Education and Technology*. <https://doi.org/10.1007/s10956-025-10260-x>
- Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Chan, K. K. H., Cooper, R., Friedrichsen, P., Gess-Newsome, J., Henze-Rietveld, I., Hume, A., Kirschner, S., Liepertz, S., Loughran, J., Mavhunga, E., Neumann, K., Nilsson, P., ... Wilson, C. D. (2019). The Refined Consensus Model of Pedagogical Content Knowledge in Science Education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (pp. 77–94). Springer.
- Chaitidou, M., Spyrtou, A., Kariotoglou, P., & Dimitriadou, C. (2018). Professional development in inquiry-oriented Pedagogical Content Knowledge among primary school teachers. *International Journal of Science, Mathematics and Technology Learning*, 25(2), 17–36. <https://doi.org/10.18848/2327-7971/CGP/V25I02/17-36>
- Chan, C. K. Y., & Colloton, T. (2024). *Generative AI in Higher Education*. Routledge. <https://doi.org/10.4324/9781003459026>
- Chan, K. K. H., & Hume, A. (2019). Towards a Consensus Model: Literature Review of How Science Teachers' Pedagogical Content Knowledge Is Investigated in Empirical Studies. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science* (pp. 3–76). Springer.
- Choi, Y. (2025). Integrating ChatGPT into the Design of 5E-Based Earth Science Lessons. *Education Sciences*, 15(7), 815. <https://doi.org/10.3390/educsci15070815>
- Cooper, G. (2023). Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *Journal of Science Education and Technology*, 32(3), 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- ElSayary, A. (2023). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for

- teaching and learning in the digital era. *Journal of Computer Assisted Learning*. <https://doi.org/10.1111/jcal.12926>
- ElSayary, A., Kuhail, M. A., & Hojeij, Z. (2025). Examining the Role of Prompt Engineering in Utilizing Generative AI Tools for Lesson Planning: Insights From Teachers' Experiences and Perceptions. *Human Behaviour and Emerging Technologies*, 2025(1). <https://doi.org/10.1155/hbe2/9986139>
- Feldman-Maggor, Y., Blonder, R., & Alexandron, G. (2024). Perspectives of Generative AI in Chemistry Education Within the TPACK Framework. *Journal of Science Education and Technology*. <https://doi.org/10.1007/s10956-024-10147-3>
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, P. J. Friedrichsen, & J. Loughran (Eds.), *Re-examining Pedagogical Content Knowledge in Science Education* (pp. 38–52). Routledge.
- Großmann, L., Koberstein-Schwarz, M., Krüger, D., & Krell, M. (2025). Lesson planning with ChatGPT for inquiry-based biology instruction – A(I) roll of the dice? *International Journal of Science Education*, 1–20. <https://doi.org/10.1080/09500693.2025.2567509>
- Großmann, L., & Krüger, D. (2024). Assessing the quality of science teachers' lesson plans: Evaluation and application of a novel instrument. *Science Education*, 108(1), 153–189. <https://doi.org/10.1002/sce.21832>
- Hashem, R., Ali, N., El Zein, F., Fidalgo, P., & Abu Khurma, O. (2024). AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Research and Practice in Technology Enhanced Learning*, 19, 023. <https://doi.org/10.58459/rptel.2024.19023>
- Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, 6, 100225. <https://doi.org/10.1016/j.caeai.2024.100225>
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge* (pp. 95–132). Springer. https://doi.org/10.1007/0-306-47217-1_4
- Otto, C. A., & Everett, S. A. (2013). An Instructional Strategy to Introduce Pedagogical Content Knowledge Using Venn Diagrams. *Journal of Science Teacher Education*, 24(2), 391–403. <https://doi.org/10.1007/s10972-012-9272-5>
- Park, S., & Oliver, J. S. (2008). Revisiting the Conceptualisation of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals. *Research in Science Education*, 38(3), 261–284. <https://doi.org/10.1007/s11165-007-9049-6>
- Peikos, G., & Stavrou, D. (2025). ChatGPT for Science Lesson Planning: An Exploratory Study Based on Pedagogical Content Knowledge. *Education Sciences*, 15(3), 338. <https://doi.org/10.3390/educsci15030338>
- Shulman, L. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Şimşek, N. (2025). Integration of ChatGPT in mathematical story-focused 5E lesson planning: Teachers and pre-service teachers' interactions with ChatGPT. *Education and Information Technologies*, 30(8), 11391–11462. <https://doi.org/10.1007/s10639-024-13258-x>
- Uğraş, H., Uğraş, M., Papadakis, S., & Kalogiannakis, M. (2025). Innovative Early Childhood STEM Education with ChatGPT: Teacher Perspectives. *Technology, Knowledge and Learning*, 30(2), 809–831. <https://doi.org/10.1007/s10758-024-09804-8>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. V., & Zhou, D. (2023). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *ArXiv*.
- Zoupidis, A., Spyrtou, A., Pnevmatikos, D., & Kariotoglou, P. (2021). Teaching and Learning Floating and Sinking: Didactic Transformation in a Density-Based Approach. *Fluids*, 6(4), 158. <https://doi.org/10.3390/fluids6040158>

Artificial Intelligence In Teacher Education: Attitudes, Usage, And The Road To AI Literacy

Wolfgang Lutz, Jens Damköhler and Thomas Trefzger
Julius-Maximilians-Universität Würzburg, Germany

The rapid development of artificial intelligence is also having an impact on the education sector and placing new demands on the professionalisation of future teachers. This study examines the extent to which teacher trainees use generative AI in their studies and how they assess their attitudes and skills with regard to this technology. To this end, two cross-sectional surveys were conducted via online questionnaire in April 2024 (n = 221) and April 2025 (n = 204). The results show a significant increase in AI use between 2024 and 2025, especially in the study-related use outside of lectures, while its use in courses remains comparatively low despite the increase. It also shows that higher use is associated with an increase in positive and a decrease in negative attitudes. There is also a positive correlation between usage rates and subjectively perceived AI literacy, AI self-efficacy and AI self-competence. In 2025, positive attitudes and all recorded perceptions of competence increase significantly, while negative attitudes remain largely stable as compared to 2024.

Keywords: AI usage, AI attitudes, AI literacy

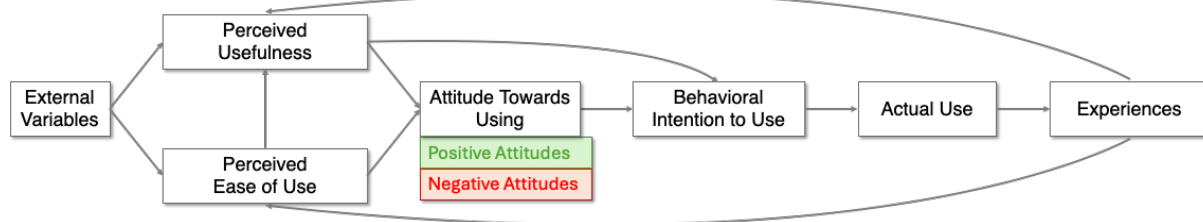
Introduction And Theoretical Background

The advancing digitalisation of schools is expanding the possibilities for the didactic and methodological design of lessons. To exploit this potential, it does not suffice to simply replace analogue media with digital media. Rather, teaching and learning processes must be transformed (Puentedura, 2006). These changes keep presenting teachers with new challenges (Eickelmann, 2018; Scheiter, 2017). It is thus not only important to master digital technologies but above all to use them in a didactically reflective manner. The introduction of ChatGPT at the end of 2022 marked the beginning of a rapid and ongoing development of artificial intelligence (AI), which has also triggered changes in the field of education. In this context, current and future teachers bear a great responsibility (Europäische Kommission, 2022) as they are expected to systematically integrate teaching and learning with and about AI into their subject teaching and prepare their students to use this technology competently (Huwer et al., 2024). In a rapidly evolving digital world, the continuous training and further education of teachers appears to keep gaining importance. Teacher trainees are the prospective teachers who must take on these tasks in future. Even though as digital natives they are considered to be tech-savvy, research paints a more differentiated picture: During their studies, teacher trainees often gain too little practical experience with digital technologies (Eickelmann et al., 2019), show competence deficits in the area of digitalisation compared to students of other disciplines (Senkbeil et al., 2020), and rate their digital competences as only mediocre (Drossel & Eickelmann, 2018). The discrepancy between digital requirements and insufficient experience in teacher training programmes highlights an urgent need for action.

Therefore, it seems important that teacher trainees are willing to engage with AI in order to develop AI competences (Huwer et al., 2024). In the Technology Acceptance Model (Figure 1), two factors influencing a person's attitudes and thus their acceptance of a technology can be identified: perceived usefulness and perceived ease of use. These two factors result in positive and negative attitudes which shape a person's overall attitude. A positive attitude promotes the intention to use the technology, whereas a negative attitude reduces the intention to use. In

addition, the intention to use a technology can also be influenced independently of a person's individual attitudes. This is the case, for example, when the added value or efficiency gains offered by the technology are obvious. The intention to use ultimately determines the actual use of a technology. Through every use of a technology, a person automatically gains experience which is why this component may be added to the Technology Acceptance Model. The successful application of a technology can also increase its perceived usefulness and have a positive effect on attitudes, as, for example, experienced users recognise more possible applications for a technology and find it more intuitive to use due to their familiarity with it.

Figure 1. Technology acceptance model enhanced by experiences (according to Davis, 1986).



Research Questions And Hypothesis

This article focuses on the use of generative AI in teacher training programmes and the attitudes and perceived competence of students. It also examines changes between 2024 and 2025. The following research questions and hypotheses specify this research interest.

Q1: To what extent do teacher trainees use AI?

New technologies, such as AI, can often only be used in university courses after a delay, especially when licences must first be purchased and data protection issues clarified. Furthermore, the use of AI in higher education has so far been described mainly through isolated best practice examples, while systematic evaluations of teaching concepts are still pending (Huwert et al., 2024). While universities must first clarify formal and legal criteria, students can immediately access free offers and use AI for interactive and adaptive learning, for example, or as a personal learning coach (Baidoo-Anu & Owusu Ansah, 2023; Baig & Yadegaridehkordi, 2024; von Garrel & Mayer, 2025). According to the TAM model (Davis, 1986), the perceived usefulness and low-threshold accessibility of AI tools increases the likelihood of use. This leads to the following hypothesis:

H1.1: Teacher trainees' out-of-courses use of AI tools is higher than their use in courses.

Q2: To what extent are teacher trainees' attitudes and perceived AI competence influenced by their AI use?

Surveys on the use of generative AI show that a large number of students already use this technology for their studies (Gottschling et al., 2024; Lutz et al., 2025; von Garrel & Mayer, 2025). Based on the extended TAM model (Figure 1), it can be assumed that positive attitudes promote the intention to use a technology and thus the actual use, while negative attitudes tend to inhibit both. At the same time, every experience with a technology contributes to a change in perceived usefulness and perceived ease of use, which in turn affects a person's attitudes. While successful applications have a positive effect, unsuccessful application can have a negative impact. In addition, little or no experience with a technology can be associated with increased uncertainty and fear, which inhibits technology acceptance (Dönmez-Turan & Kır, 2019). This leads to following hypothesis:

H2.1: Students with greater AI use rate the chances of AI more positive and the risks less negative.

Competences are not built by the acquisition of cognitive skills and abilities alone, but also by their successful and responsible application in various situations (Weinert, 2001). Competent AI use thus requires active use of the technology. Practical experience also helps to familiarise oneself with the possibilities and limitations of generative AI. This allows the technology to be used in a more targeted, effective, and reflective manner. As a result, the successful application of AI tools becomes more likely. In the study by Süße & Kobert (2023) high correlation effects between the use of generative AI and perceived learning success were found among some students. According to Bandura (1997), perceived successes strengthen one's self-efficacy and thus confidence in one's competence, even in challenging situations. This leads to following hypothesis:

H2.2: Students with a higher usage behaviour rate their AI competences, their AI self-efficacy, and their AI self-competency higher.

Q3: How did usage behaviour, attitudes, and perceived AI skills change between 2024 and 2025?

Generative AI is subject to particularly dynamic technological development, which is accompanied by continuous improvement and increased performance. The further development of new language models means that the up-to-then most recent versions are made available free of charge as previous versions. That way, increasingly better versions are made available free of charge. For students, this means that AI tools with a growing range of functions can be integrated into everyday student life without financial barriers. Better and more powerful systems increase the perceived usefulness, which, according to the TAM model (Davis, 1986), leads to a more positive attitude and more frequent use of the technology. Based on a total of 49 study results on the usage behaviour of generative AI among students, Von Garrel & Mayer (2025) performed a regression analysis that shows an increase in usage behaviour from 2023 to early 2025. Thus, the following hypotheses arise:

H3.1: In 2025, teacher trainees' study-related use of AI is higher than in 2024.

H3.2: In 2025, teacher trainees associate fewer risks and more chances with the study-related use of AI tools than in 2024.

H3.3: In 2025, teacher trainees rate their AI competence, AI self-efficacy, and AI self-competency higher than in 2024.

Study description

Study Design

This study aims to record the usage behaviour, attitudes, and subjectively perceived skills of teacher trainees with regard to generative AI. To this end, cross-sectional surveys were conducted in April 2024 and April 2025 using an online questionnaire. Participation in the surveys was voluntary, and data collection was pseudonymised by creating an acronym from personal queries. This code shows that the overlap between students who participated in both surveys is less than 1 percent. A longitudinal analysis of the data therefore does not appear to be meaningful, so the data collected in 2024 and 2025 are considered statistically independent in the analysis.

Sample

Table 1 provides an overview of the teacher trainees participating in the surveys in 2024 and 2025. The sample sizes of both studies are comparable with $n = 221$ (2024) and $n = 204$ (2025). The gender distribution is similar in both years, with a significantly higher proportion of female students. There were only a few participants in both surveys who stated that they were non-binary. In Germany, teacher training is organised according to school type and comprises five types of

school, each of which has its own degree programme. The percentage distribution of students across the school types roughly corresponds to the number of students enrolled in each school type. The imbalance is therefore not due to the survey design. In both surveys, most students were in their third to sixth semester. Plausible explanations for this are that the standard time to degree vary between seven and nine semesters depending on the type of school and that the surveys were conducted by lecturers as part of didactic courses. These are typically attended by students in the middle of their studies.

Table 1. Sample composition.

	2024					2025				
	f	m	d			f	m	d		
Gender	161	59	1			144	57	3		
Type of school	GYM	GS	FS	RS	MS	GYM	GS	FS	RS	MS
	36.7%	34.8%	15.8%	7.7%	5.0%	40.2%	28.4%	14.7%	10.8%	5.9%
Semester	1 - 2	3 - 4	5-6	7-8	9-10	1 - 2	3 - 4	5-6	7-8	9-10
	10.4%	36.7%	22.2%	15.4%	10.4%	8.1%	25.3%	27.1%	14.5%	13.1%
	11-12	13-14	>14			11-12	13-14	>14		
	2.3%	1.4%	1.4%			3.2%	0.5%	0%		

Operationalisation And Test Instruments

The use of generative AI was assessed using two single-item measures, one relating to courses ("Was artificial intelligence used as a tool to support the learning process in courses you attended in previous semesters?") and the other relating to individual study-related out-of-course use ("Do you use AI tools in your studies?"). Both items were rated on a four-point frequency scale (never = 0, rarely = 1, occasionally = 2, regularly = 3).

To assess teacher trainees' attitudes towards generative AI, the General Attitudes towards Artificial Intelligence Scale (GAAIS, Schepman & Rodway, 2023) was adapted. The original scale comprises a total of twelve items on positive attitudes and eight items on negative attitudes, each rated on a five-point Likert scale (strongly disagree = 0, disagree = 1, neutral = 2, agree = 3, strongly agree = 4). For the present study, three positive items (e.g., "I am interested in using artificially intelligent systems in my daily life") and three negative items (e.g., "I think artificial intelligence is dangerous") were selected. The reliabilities are in an acceptable to good range, with $\alpha = 0.76$ (positive attitude) and $\alpha = 0.74$ (negative attitude).

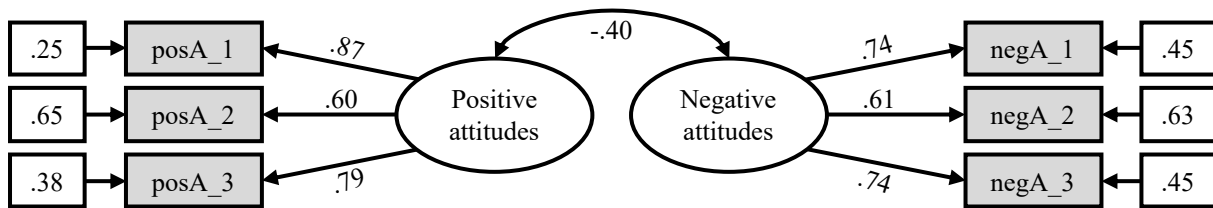
The subjectively perceived AI literacy (AILit), AI self-efficacy (AISEff) and AI self-competency (AISCom) of the teacher trainees were assessed using the Meta AI literacy scale (MAILS, Carolus et al., 2023). The 30 items are each rated on an eleven-point rating scale (not at all pronounced 0 – 10 very well pronounced). The scale is multidimensional. The AILit factor is composed of the factors use & apply AI (UsApAI, 6 items, $\alpha = 0.96$), know & understand AI (KnUnAI, 6 items, $\alpha = 0.88$), detect AI (DetAI, 3 items, $\alpha = 0.84$) and AI ethics (AIEth, 3 items, $\alpha = 0.87$). The factor AISEff is formed from the subcategories AI problem solving (3 items, $\alpha = 0.90$) and AI learning (3 items, $\alpha = 0.88$). The AISCom factor is composed of AI persuasion literacy (3 items, $\alpha = 0.73$) and AI emotion regulation (3 items, $\alpha = 0.83$). The reliabilities for each construct are in the acceptable to very good range.

Modelling

According to the multi-component model of attitude (Zanna & Rempel, 2008), a person's attitude is derived from the overall evaluation of cognitive, affective, and behavioural information. Attitudes can be ambivalent, i.e. both positive and negative (Haddock & Maio, 2014). For example, on a cognitive level, a person may perceive a technology as useful and yet at the same

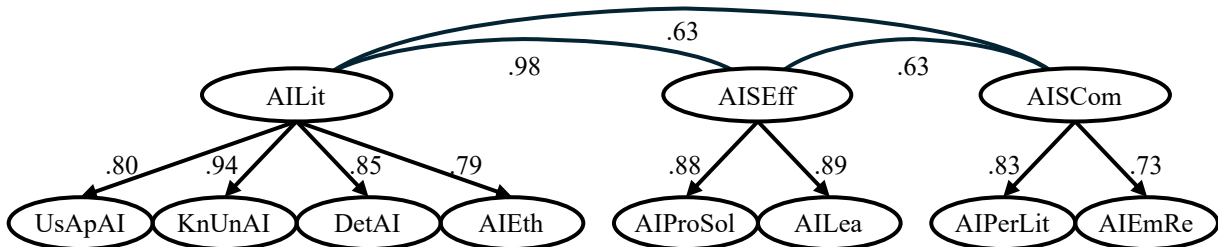
time recognise risks in its use (e.g. data protection). On an affective level, both positive emotions (e.g. joy) and negative emotions (e.g. frustration) can arise when the use of a technology leads to successes but also failures. On a behavioural level, a person may consciously reject the use of a technology in certain situations (e.g. sensitive data) despite an overall high willingness to use it. Against this background, the positive and negative attitudes of teacher trainees were modelled as two separate and mutually influencing factors using confirmatory factor analysis (see Figure 2). Both the fit indices (CFI = 0.998, RMSEA = 0.020, SRMR = 0.025) and the non-significant chi-square test ($\chi^2(8) = 9.325$, $p=.316$) indicate a very good fit of the model (Hu & Bentler, 1999; MacCallum et al., 1996).

Figure 2. Latent measurement model of positive and negative attitudes.



Based on the multidimensional approach of Carolus et al. (2023), the constructs AILit, AISEff, and AISCom were specified as latent variables using confirmatory factor analysis (Figure 3). Although the chi-square test is significant ($\chi^2(381) = 781.309$, $p<.001$), which is often observed in larger samples and complex models, the fit indices (CFI = 0.953, RMSEA = 0.048, SRMR = 0.053) indicate a good model fit (Hu & Bentler, 1999; MacCallum et al., 1996).

Figure 3. Latent measurement model for AILit, AISEff, and AISCom.



Results

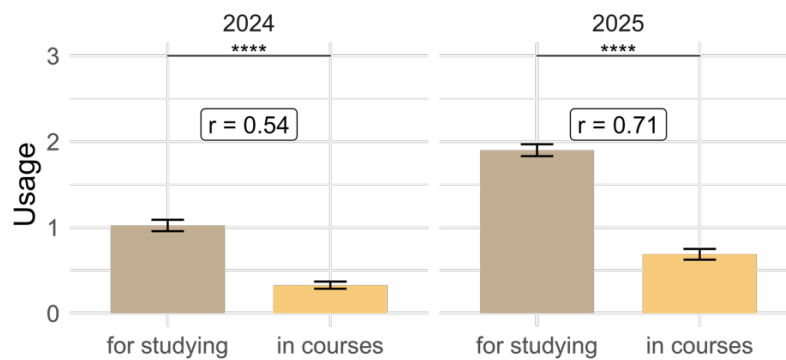
In 2024, the majority of students (74.5%) stated that generative AI is never used in courses. Only a small proportion of respondents answered that they rarely (19.1%), occasionally (5.5%), or regularly (0.9%) used AI. However, the frequency of use increased in 2025. More students stated that AI tools are used rarely (23.9%), occasionally (16.4%), or even regularly (4.0%) in courses. Nonetheless, the majority of respondents still stated that AI is never used in courses (55.7%).

The frequency of study-related out-of-class AI use has also increased. In 2024, most students stated that they never (40.5%) or rarely (23.6%) used AI for study-related matters. However, a significant number of students used AI tools occasionally (29.1%) or even regularly (6.8%). In 2025, only 10.0% of students said that they never used AI tools for their studies. Most respondents indicated they used AI rarely (23.4%), occasionally (33.3%), or even regularly (33.3%) for their studies.

Paired Wilcoxon tests were used to examine the extent to which teacher trainees rated the frequency of use of generative AI for their studies and in courses differently (Figure 4). The students indicated that they used generative AI more frequently outside of class for their studies (2024: $\mu = 1.02$, $SE = 0.067$, $Mdn = 1$ | 2025: $\mu = 1.90$, $SE = 0.069$, $Mdn = 2$) than in courses (2024: $\mu = 0.33$, $SE = 0.042$, $Mdn = 0$ | 2025: $\mu = 0.69$, $SE = 0.063$, $Mdn = 0$). According to the

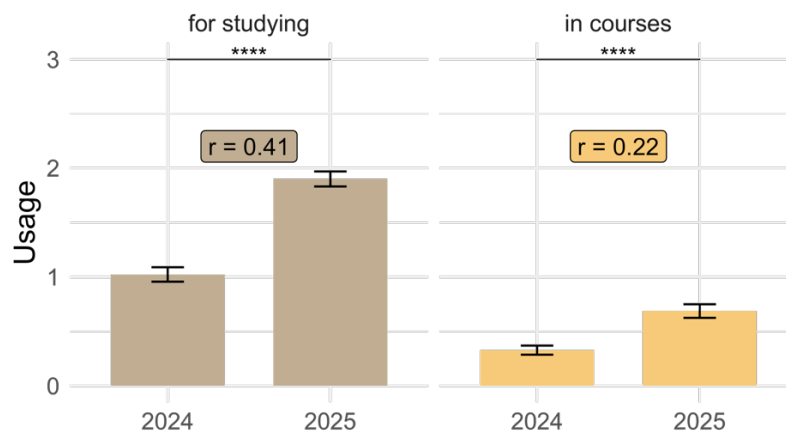
Wilcoxon rank sum test, this difference is significant in both surveys, with a large effect size in each case (2024: $V = 7196$, $p < .001$, $r = .54$ | 2025: $V = 11753$, $p < .001$, $r = 0.71$).

Figure 4. Comparison of frequency of out-of-course use and use in courses.



To analyse the changes in frequency of use, unpaired Wilcoxon rank sum tests were calculated based on statistically independent samples. It can be seen that both the study-related out-of-course use of AI ($V = 12096$, $p < .001$, $r = 0.41$) and in courses ($V = 17408$, $p < .001$, $r = 0.22$) increased significantly over time with a medium or small effect size (Figure 5).

Figure 5. Comparison of the frequency of generative AI use in 2024 and 2025.



To examine the relationship between attitudes and the frequency of use reported by students (Figure 6), linear regressions were calculated based on data from both surveys. There is a significant correlation between positive attitudes and frequency of use ($b = 0.45$, $SE = 0.04$, $t = 12.58$, $p < .001$), i.e. students who report higher usage tend to have more positive attitudes. Frequency of use explains 27% of the variance in positive attitudes ($R^2 = 0.27$). Negative attitudes also show a significant correlation with frequency of use ($b = -0.14$, $SE = 0.03$, $t = -4.08$, $p < .001$), i.e. students with higher usage rates have fewer negative attitudes towards the technology. However, the explained variance is significantly lower at $R^2 = 0.04$ compared to positive attitudes.

A multivariate analysis of variance was calculated to compare the attitudes of students in 2024 and 2025 (Figure 7). The attitudes differ significantly with a medium effect size ($\Lambda = .94$, $F(2, 418) = 13.95$, $p < .001$, $\eta^2 = 0.06$). In the follow-up tests, t-tests with Bonferroni correction were calculated for the positive and negative attitudes. The positive attitudes of the students grew significantly from 2024 ($\mu = -0.177$, $SE = 0.061$) to 2025 ($\mu = 0.225$, $SE = 0.063$) with a small to medium effect size ($t(419) = -4.57$, $p < .001$, $d = -0.45$). The negative attitudes towards AI among students did not change significantly between 2024 ($\mu = -0.009$, $SE = 0.056$) and 2025 ($\mu = -0.002$, $SE = 0.053$) ($t(419) = -0.09$, $p = .93$, $d < -0.01$).

Figure 6. Mean values of attitudes with standard errors in relation to the frequency of use.

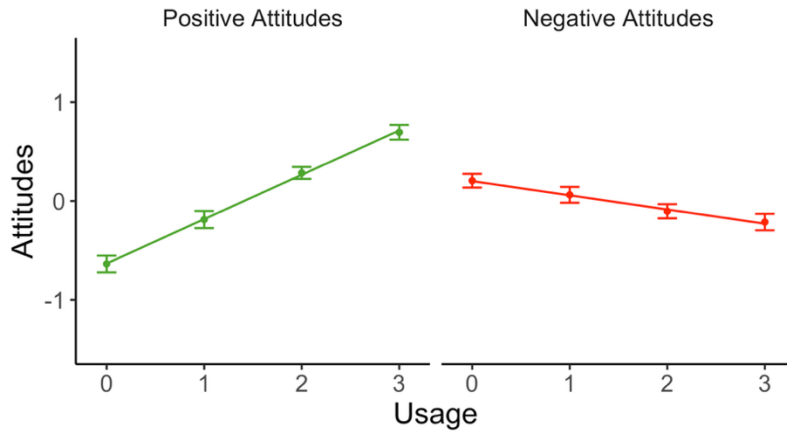
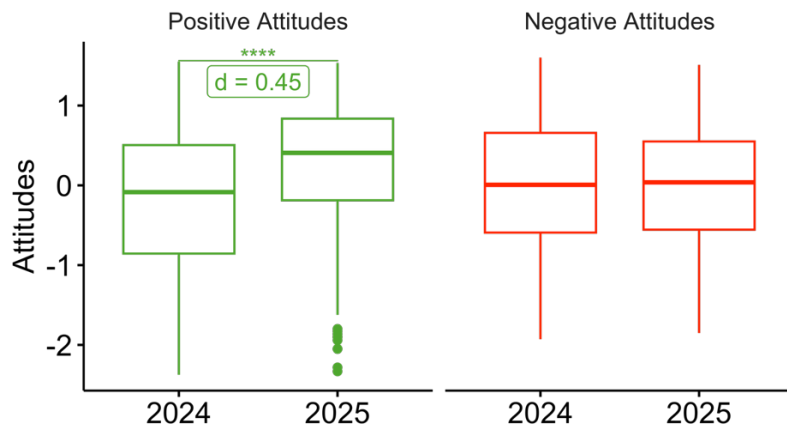
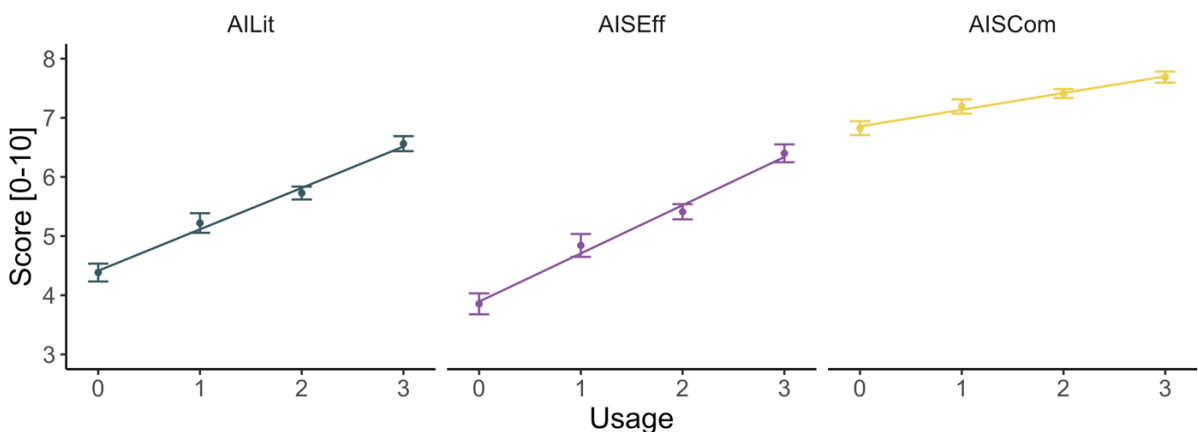


Figure 7. Comparison of attitudes towards generative AI in 2024 and 2025.



As with the attitudes, the correlations between frequency of use and AI literacy, AI self-efficacy and AI self-competence (Figure 8) were examined using linear regression analyses. All data from the surveys were used for this purpose. All factors showed a significant correlation with frequency of use (AILit: $b = 0.70$, $SE = 0.06$, $t = 10.86$, $p < .001$ | AISEff: $b = 0.81$, $SE = 0.08$, $t = 10.77$, $p < .001$ | AISCom: $b = 0.28$, $SE = 0.05$, $t = 5.92$, $p < .001$). The variance explained by usage behaviour is $R^2 = 0.22$ (AILit), $R^2 = 0.21$ (AISEff), and $R^2 = 0.08$ (AISCom).

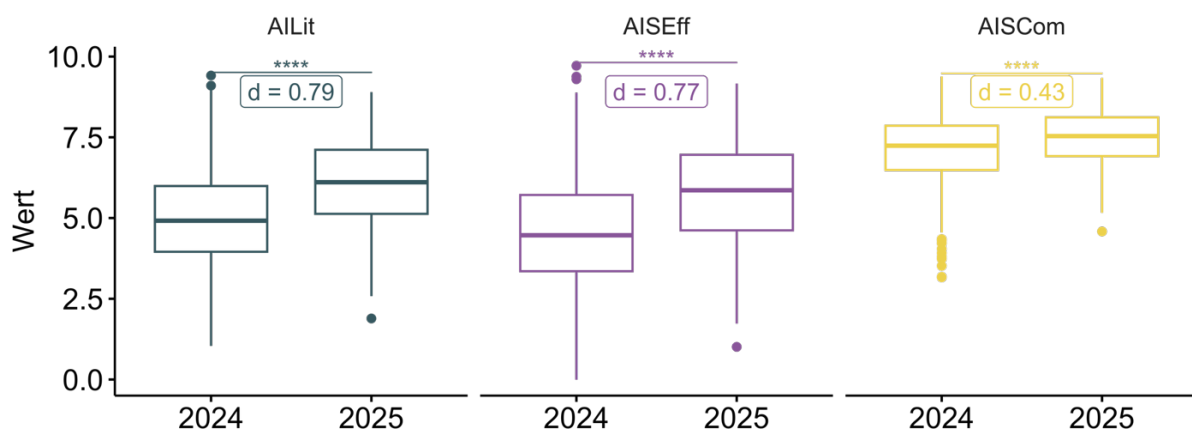
Figure 8. Correlation between the perceived AI competences and the frequency of use.



In a final study, the students' perceived competence in 2024 and 2025 (Figure 9) was compared. For this purpose, a multivariate analysis of variance with Pillai's Trace was calculated, as the significant results of the Levene tests (AILit: $F(1, 419) = 6.63$, $p = .01$ | SCom: $F(1, 419) = 8.43$, $p < 0.01$ | SEff: $F(1, 419) = 6.35$, $p = .01$), showed no variance homogeneity between the groups.

The analysis shows that the students' perceptions of competence differ significantly between years with a large effect size ($V = 0.15$, $F(3, 417) = 24.52$, $p < .001$, $\eta_p^2 = 0.15$). Welch's tests with Bonferroni correction were performed for each variable to analyse this effect in more detail. A comparison of the students' AI literacy ratings reveals a significant difference between the years 2024 ($\mu = 4.86$, $SE = 0.11$) and 2025 ($\mu = 6.04$, $SE = 0.09$) with a medium to large effect size ($t(410) = -8.14$, $p < .001$, $d = 0.79$). A similar picture emerges when comparing the AI self-efficacy ratings, which were lower on average in 2024 ($\mu = 4.42$, $SE = 0.13$) than in 2025 ($\mu = 5.77$, $SE = 0.11$). This difference is significant with a medium to large effect size ($t(412) = -7.97$, $p < .001$, $d = 0.77$). The AI self-competence rating also increased significantly from 2024 ($\mu = 7.04$, $SE = 0.08$) to 2025 ($\mu = 7.50$, $SE = 0.06$) with a small to medium effect size ($t(399) = -4.42$, $p < .001$, $d = 0.43$).

Figure 9. Comparison of perceived AI competences in 2024 and 2025.



Discussion

The results of this study show that the use of generative AI in teacher training has increased and that the majority of students now use this technology occasionally or regularly for their studies. Even though the use of AI tools in courses has increased significantly, this technology is still not used in most courses. The results of the study show that generative AI is used significantly more often outside of courses than within courses. Thus, hypotheses H1.1 and H3.1 can be confirmed.

In the context of this study, increased use of generative AI is associated with more positive attitudes and fewer negative attitudes. Hypothesis H2.1 can thus be confirmed. However, it is striking that the correlation between use and positive attitudes is significantly stronger than the correlation between use and negative attitudes. This suggests that experiences primarily contribute to the perception of potential, while concerns and perceptions of risk only decrease slightly. This effect is also evident when comparing the attitudes of students in 2024 and 2025. With increased usage in 2025, there is also an increase in positive attitudes among students, while negative attitudes remain unchanged. Hypothesis H3.2 can therefore only be accepted to a limited extent for positive attitudes.

There is a positive correlation between frequency of use and the level of AI literacy, AI self-efficacy, and AI self-competence. Hypothesis H2.2 can therefore be confirmed. Students' experience with generative AI leads to greater confidence in their own abilities and the perception that they can use AI in a targeted manner. However, it should be noted that this study deals with subjectively perceived competencies. Frequent use can therefore also lead to a misjudgement of competence, especially if students accept the information generated by AI without systematic verification. This is particularly relevant in the case of generative AI, because the technology produces outputs that are written in a convincing manner but often incorrect in terms of content.

It therefore remains unclear to what extent the increased perceptions of competence are actually accompanied by objective gains in competence. A comparison of students' perceptions of competence between 2024 and 2025 also shows a significant increase in the perception of competence in all categories considered; hypothesis H3.3 can thus be accepted.

Limitations Of The Study

Despite the meaningful results, some limitations must be taken into account. For example, the frequency of use is recorded through single-item measurements, which limits the objectivity and differentiation of the measurement. Furthermore, the data is cross-sectional, which reflects differences between the survey dates in 2024 and 2025, but does not allow conclusions to be drawn about individual changes. Finally, the measurement of skills was based exclusively on self-assessments and not on skills tests. The extent to which experience with generative AI influences actual skills development cannot be answered within the scope of this study.

Conclusion

Overall, the results of this study show that the use of generative AI in teacher training has increased significantly within a short period of time. While students are developing more positive attitudes towards the technology and rate their AI literacy, AI self-efficacy, and AI self-competence higher, negative attitudes remain unchanged. In addition, the use of generative AI in courses has been limited so far, with students primarily using AI outside of institutional learning settings. On the one hand, this indicates an increasing normalisation of generative AI in everyday student life, but on the other hand, it also shows that this development has so far been largely self-organised and less supported by universities in a targeted manner. Thus, it becomes clear that although increasing use is an important foundation, it does not automatically lead to pedagogically and didactically reflective application in later careers. In order for teacher trainees to be prepared for the use of generative AI in a school context, more than just user experience is needed. They must develop the skills to use AI responsibly, didactically, and reflectively. At this point, the university has a central responsibility to prepare future teachers for this task. The work of (Damköhler et al., 2024a, 2024b, 2025) or Herz et al. (2025) shows how generative AI can be successfully implemented in teaching.

References

- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, 7(1), 52–62. <https://doi.org/10.61969/jai.1337500>
- Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, 102411. <https://doi.org/10.1016/j.ijer.2024.102411>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.
- Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS-Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change-and meta-competencies. *Computers in Human Behaviour: Artificial Humans*, 1(2), 100014.
- Damköhler, J., Lutz, W., & Trefzger, T. (2024a). Das Würzburger KI-Projekt: ChatGPT als Reflexionscoach im Lehr-Lern-Labor-Seminar Physik. In J. Huwer, S. Becker-Genschow, C. Thyssen, L.-J. Thoms, A. Finger, L. von Kotzebue, E. Kremser, M. Meier, & T. Bruckermann (Eds), *Kompetenzen für den Unterricht mit und über Künstliche Intelligenz: Perspektiven, Orientierungshilfen und Praxisbeispiele für die Lehramtsausbildung in den Naturwissenschaften* (pp. 70–73). Waxmann.
- Damköhler, J., Lutz, W., & Trefzger, T. (2024b). Der digitale Dozent: ChatGPT als Co-Pilot in der Lehrpersonenbildung. In H. Grötzebauch & S. Heinicke (Eds), *PhyDid B - Didaktik der Physik—Beiträge zur DPG-Frühjahrstagung Physik Greifswald 2024* (pp. 91–96).
- Damköhler, J., Lutz, W., & Trefzger, T. (2025). ChatGPT als Reflexionscoach—Einblicke in das

- Würzburger Lehr-Lern-Labor. *Entdecken, Lehren Und Forschen Im Schülerlabor. Gesellschaft Für Didaktik Der Chemie Und Physik Jahrestagung 2024 in Bochum*, 45, 41–44.
- Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* [Thesis, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/15192>
- Dönmez-Turan, A., & Kır, M. (2019). User anxiety as an external variable of technology acceptance model: A meta-analytic study. *Procedia Computer Science*, 158, 715–724. <https://doi.org/10.1016/j.procs.2019.09.107>
- Drossel, K., & Eickelmann, B. (2018). Die Rolle der Lehrerprofessionalisierung für die Implementierung neuer Technologien in den Unterricht—Eine Latent-Class-Analyse zur Identifikation von Lehrertypen. *MedienPädagogik: Zeitschrift Für Theorie Und Praxis Der Medienbildung*, 31, 166–191.
- Eickelmann, B. (2018). Digitalisierung in der schulischen Bildung Entwicklungen, Befunde und Perspektiven für die. *Digitalisierung in Der Schulischen Bildung: Chancen Und Herausforderungen*, 11.
- Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (2019). *ICILS 2018# Deutschland: Computer-und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking*. Waxmann Verlag.
- Europäische Kommission. (2022). *Ethische Leitlinien für Lehrkräfte über die Nutzung von KI und Daten für Lehr- und Lernzwecke*. Amt für Veröffentlichungen der Europäischen Union. <https://data.europa.eu/doi/10.2766/494>
- Gottschling, S., Seidl, T., & Vonhof, C. (2024). Nutzung von KI-Tools durch Studierende. Eine exemplarische Untersuchung studentischer Nutzungsszenarien. *Die Hochschullehre*, 10. <https://doi.org/10.3278/HSL2411W>
- Haddock, G., & Maio, G. R. (2014). Einstellungen. In K. Jonas, W. Stroebe, & M. Hewstone (Eds), *Sozialpsychologie* (pp. 197–229). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41091-8_6
- Herz, P., Damköhler, J., Lutz, W., & Trefzger, T. (2025). ChatGPT im Lehr-Lern-Labor: Potenziale eines KI-basierten Assistenten bei der Entwicklung von Experimentierumgebungen. In G. H. & H. S. (Eds), *DPG-Frühjahrstagung 2025*. PhyDid B.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huwer, J., Becker-Genschow, S., Thyssen, C., Thoms, L.-J., Finger, A., von Kotzebue, L., Kremser, E., Meier, M., & Bruckermann, T. (2024). *Kompetenzen für den Unterricht mit und über Künstliche Intelligenz: Perspektiven, Orientierungshilfen und Praxisbeispiele für die Lehramtsausbildung in den Naturwissenschaften*. <https://doi.org/10.57668/phtg-000572>
- Lutz, W., Damköhler, J., & Trefzger, T. (2025). Künstliche Intelligenz—Akzeptanz und AI-Literacy unter Lehramtsstudierenden. *Entdecken, Lehren Und Forschen Im Schülerlabor. Gesellschaft Für Didaktik Der Chemie Und Physik Jahrestagung 2024 in Bochum*, 45, 45–48.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modelling. *Psychological Methods*, 1(2), 130.
- Puentedura, R. (2006). *Transformation, technology, and education*.
- Scheiter, K. (2017). *Lernen mit digitalen Medien: Potenziale und Herausforderungen aus Sicht der Lehr-Lernforschung*.
- Schepman, A., & Rodway, P. (2023). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust. *International Journal of Human-Computer Interaction*, 39(13), 2724–2741.
- Senkbeil, M., Ihme, J. M., & Schöber, C. (2020). Empirische Arbeit: Schulische Medienkompetenzförderung in einer digitalen Welt: Über welche digitalen Kompetenzen verfügen angehende Lehrkräfte? *Psychologie in Erziehung Und Unterricht*, 68(1), 4–22.
- Süße, T., & Kobert, M. (2023). *Generative KI an Schulen—Eine Studie über die Nutzung generativer KI aus Sicht von Schülerinnen und Schülern unter Berücksichtigung handlungsleitender Eigenschaften und ausgewählter sozialer Kontextfaktoren*. <https://doi.org/10.5281/ZENODO.10210312>
- von Garrel, J., & Mayer, J. (2025). *Künstliche Intelligenz im Studium—Eine quantitative Längsschnittstudie zur Nutzung KI-basierter Tools durch Studierende* [Application/pdf]. 9790 KB. https://doi.org/10.48444/H_DOCS-PUB-533
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen—Eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* (pp. 17–33). Beltz.

Zanna, M. P., & Rempel, J. K. (2008). Attitudes: A new look at an old concept. In *Attitudes: Their structure, function, and consequences*. (pp. 7–15). Psychology Press.

Augmented Reality For Chemistry Education: Visualizing The Invisible In The Nitrogen Biogeochemical Cycle – A Pilot Study

Luka Ribič, Luka Vinko and Iztok Devetak

University of Ljubljana, Faculty of Education, Ljubljana, Slovenia

Chemistry is a subject that students usually find very complex, also due to the triple nature of chemical concepts. Learners have to bridge, in their working memories, different levels of chemical phenomena explanations in order to adequately understand them. These learning processes can cause difficulties. One way to address these difficulties is through the use of technology. Virtual learning environments, such as augmented reality (AR) can provide important information, enable connections between observable (macroscopic) and unobservable (submicroscopic) processes and thereby might contribute to a better understanding of chemistry. The aim of this pilot study was to gain insight into pre-service primary school teachers' attitudes toward the use of AR technology in the context of environmental chemistry and sustainable development as well as their perceptions of the usability of the AR application. 10 participants completed the Technology Acceptance Model questionnaire and the System Usability Scale. The results showed that the pre-service teachers had positive attitudes toward AR technology, perceived the AR application easy to use, and expressed willingness to participate in additional courses that integrate AR into chemistry teaching and learning. These findings primarily indicate the feasibility and acceptance of AR supported laboratory activities in environmental chemistry. The results provide a foundation for further research examining the effects of AR on conceptual understanding, particularly in relation to supporting connections between macroscopic, submicroscopic and symbolic representations of chemical phenomena.

Keywords: augmented reality, environmental chemistry, pre-service teachers

Introduction

In science education, assessing students' conceptual understanding is essential for enhancing learning outcomes (Halloun & Hestenes, 1985). It is well known that students experience difficulties in constructing and understanding of chemical concepts (Johnstone, 1991). Previous research (Horvat et al., 2023; Tan et al., 2020) has highlighted the importance of conceptual understanding in science education. Therefore, it is crucial for learners to develop deep comprehension of scientific ideas, rather than merely memorizing facts. The learning difficulties students face in different science subjects should be taken into account when planning instructional approaches (Taber, 2008). To identify these difficulties and address them effectively, an understanding of both the discipline and the philosophy of chemistry is required (Kaya & Erduran, 2013). Consequently, chemistry educators need to be familiar with the nature of chemical phenomena, not only with the chemistry content they teach (Erduran, 2007). However, previous studies (Allen et al., 2024; Pavlin et al., 2009; Syahrial et al., 2023) have shown that primary school and chemistry teachers often hold misconceptions about various chemistry topics. These findings suggest that pre-service teachers in particular possess significant gaps in their conceptual chemistry content (Allen et al., 2024).

The major source of complexity in chemistry lies in the nature of chemical concepts, which often extend beyond human senses (Johnstone, 1991). To explain chemical phenomena effectively, learners must be able to bridge different levels of chemical phenomena (Sjöström et al., 2020). Johnstone (1991) proposed a model consisting of three different levels of chemical presentations (microscopic, submicroscopic and symbolic), which is reflecting the triple nature of chemical

phenomena. Integrating all three levels of representations is essential for effective chemistry teaching (Sjöström et al., 2020) and can be supported through laboratory work (Gurung & Gurung, 2023). Due to its inherent complexity, chemistry concepts have been challenging for both learners and educators (Wilsson & Bernie, 1999). Nevertheless these challenges can be mitigated through the use of technology, which can enhance students' motivation and support conceptual understanding by providing meaningful visualizations and information (Mei-Hung et al., 2018). Emerging technology with considerable potential for chemistry education is augmented reality (AR) (Akçayir & Akçayir, 2017). Several studies have investigated the use of AR in chemistry teaching (Coduto et al., 2023; Tarng et al., 2021; Yamtinah et al., 2023). However, no previous studies have examined the use of AR to enhance students' understanding of the triple nature of chemical phenomena within the context of environmental chemistry.

Chemistry Triplet

Students often perceive chemistry as a difficult subject, due to the triple nature of chemical phenomena (Johnstone, 1991). Achieving an adequate understanding of chemical phenomena, requires learners to comprehend chemical concepts across different levels of representation. This implies that educators themselves need a thorough understanding of how chemical phenomena are represented and how the use of visual representations can support chemistry learning (Gilbert & Treagust, 2009). Johnstone (1982) who proposed that chemical phenomena can be understood at three distinct levels:

- (1) Descriptive and functional: The level at which phenomena are directly experienced through the senses and can be described.
- (2) Representational: The level at which symbols are used to represent entities and processes that cannot directly be observed.
- (3) Explanatory: The level at which phenomena are explained using models.

In later work (Johnstone, 1991) clarified these levels as the macroscopic level, referring to observable phenomena, symbolic level which includes chemical equations, formulas and mathematical representations and submicroscopic level, which involves models of particles and interactions. Experts in chemistry are able to operate fluently across all three levels, whereas novice learners tend to function primarily at the macroscopic level and experience difficulties in relating it to the submicroscopic and symbolic levels (Talanquer, 2010). Research has shown the integrating all three levels of representation, particularly through the use of interactive visualization tools can lead to improved conceptual understanding of chemistry content (Ferk Savec et al., 2009). Despite this, chemical educators often emphasizes only one level of representation at a time, most commonly the symbolic level, which can hinder students' ability to construct coherent mental models of chemical phenomena (Sjöström et al. 2020). To bridge this gap, the use of interactive visualization elements is particularly important (Wu et al., 2001). Two-dimensional and three-dimensional visualizations, such as diagrams, molecular models, and animations are commonly used as visualization tools in chemistry education (Atkins et al., 1995; McMurry & McFay, 2001). These representations function as analogue models of chemical phenomena, helping learners to connect abstract concepts with observable effects (Harrison & Treagust, 1998).

The use of emerging technologies has the potential to further address difficulties associated with understanding abstract chemical concepts (Winkellman et al., 2017). In particular, virtual learning environments enable students to explore processes that cannot be observed with the naked eye (De Jong et al., 2013). During learning chemistry, a substantial portion of information is processed through visual channels (Sanders & McCormick, 1987). These channels can be

effectively supported through virtual technologies that provide real-time visualization, three-dimensional representations and opportunities for learners to manipulate and interact with virtual or real objects.

Augmented Reality

Augmented reality (AR) is a virtual learning environment that enables the digital enrichment of learning content through interactive experiences by overlaying the computer-generated elements onto the real world. Unlike virtual reality, which fully immerse users in a simulated environment, AR enhances real world without completely replacing it (Ardiny & Khanmirza, 2018). Azuma (1997) defined AR as technology that fulfils three main criteria: (1) it combines real and virtual environments, (2) it is interactive in real time and (3) it is registered in three-dimensional space. Accordingly, AR represents the continuum between the real and the virtual world (Milgram et al., 1994). In chemistry education, AR technology allows the visualization of abstract chemical phenomena particularly at a submicroscopic level of representation (Mazzuco et al., 2022). AR applications have been developed across all educational levels (Bacca et al., 2019), and studies indicate that AR can enhance students' learning outcomes (Tarnng et al., 2022). However, the effectiveness of AR is influenced by multiple factors, including technology acceptance (Liu et al., 2022). Inappropriate use of AR applications may hinder learning. Making it essential to evaluate the usability and suitability of each application within its specific learning context (Peeters, 2022). The effectiveness of an AR application in the learning process can be assessed based on three criteria: (1) effectiveness, that is, whether it enables task completion and the attainment of high-quality outcomes; (2) efficiency, meaning that it requires minimal resources for its use; and whether it provides; and (3) user satisfaction, that is whether it is engaging and motivating for learners (Brooke, 1995). Ribič and Devetak (2024) highlighted in their literature review that no previous studies have investigated the effectiveness of AR in supporting students' understanding of the chemical triplet. Furthermore, AR has not been previously used to linked selected chemical processes demonstrated in a laboratory setting to the context of environmental chemistry. Nonetheless, studies included in this review emphasized that ease of use and acceptance of technology are critical factors influencing the learning experience with AR. The authors also noted the lack of research on the application of AR specifically within the field of environmental chemistry.

Aim of the Research

Despite the potential of AR use to enhance chemistry learning, pre-service primary school teachers often hold misconceptions about both the triple nature of chemical phenomena (Pavlin et al., 2009), and the biogeochemical cycles of the elements, which are fundamental for understanding environmental chemistry (Ribič et al., 2025). Effective chemistry teaching can also be hindered by the lack of appropriate learning models and instructional resources (Timilsena, 2022). Previous studies have shown that AR technology can improve students' understanding of chemical concepts by providing interactive visualizations that support learning at submicroscopic and symbolic levels (Tarnng, 2022; Tsai et al., 2021). However, the effectiveness of AR-based learning depends on multiple factors, including learners' acceptance of the technology (Tarnng, 2022) and the usability of the application (Peeters et al., 2022). Laboratory work has been shown to further support students' understanding of chemical phenomena by allowing observation of processes at the macroscopic level. Integrating AR with laboratory activities therefore provides an opportunity to bridge the gap between observable phenomena and abstract chemical representations, offering learners information that cannot be directly observed with the naked eye. Based on this rationale, the aim of this pilot study was to evaluate pre-service primary school

teachers' technology acceptance and perceptions of the usability of an AR application designed to enhance learning during laboratory work in environmental chemistry. Specifically, the study addressed the following research questions:

- (1) What is the attitude of pre-service primary school teachers towards the use of AR in the implementation of the environmental chemistry learning module?
- (2) How do the pre-service primary school teachers perceive usability of the AR application?

Participants

A total of 10 undergraduate second year pre-service primary school teachers (20 years old) from the Faculty of Education at the University of Ljubljana participated in this pilot study.

Instruments

Two paper-based questionnaires were used to collect the data in this study. The System Usability Scale (SUS) by Brooke (1996) was employed to evaluate the usability of the AR application within the environmental chemistry module. The SUS questionnaire is widely used tool for assessing usability, perceived usefulness, user attitude, intention to use and perceived ease of use. Responses were recorded on a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). For positively formulated items responses of 3 or higher were considered positive feedback, whereas for negatively formulated items, responses of 3 or lower were treated as positive feedback (see Table 1).

The Technology Acceptance Questionnaire (TAM) by Holden & Rada (2014) was used to investigate participants' acceptance of the AR system when learning specific chemical phenomena. The questionnaire included 15 items assessing three dimensions: perceived ease of use (PEU), perceived ease of use and usability (PEUU) and perceived usefulness (PU). Responses were recorded on a 7-point Likert-type scale ranging from 0 (*strongly disagree*) to 6 (*strongly agree*). In addition, five items using semantic differential scale ranging from 1 (*bad*) to 7 (*good*) were included to assess participants' attitudes toward the use of the technology (AT).

Research Design

This pilot study introduced a new environmental chemistry module on the nitrogen cycle to pre-service primary school teachers. The module included ten different activities, a relatively large number intended to provide participants with broader contextual knowledge of nitrogen cycling across different environmental spheres. Four of these activities were laboratory activities enhanced with AR visualizations: (1) the simulation of the Haber-Bosh process, (2) the dissolving of ammonia in water (the fountain), (3) spectrometry to determine the concentration of ammonium ions in water, and (4) the adsorption of ammonium and nitrate ions on soil particles. These AR visualizations represented processes at the submicroscopic level, providing learners with interactive, otherwise unobservable views of chemical phenomena. Prior to the module, participants attended a 10-minute training session to familiarize themselves with the AR technology and associated hardware. After completing the module, participants fulfilled paper-based questionnaires (SUS and TAM). For reporting purposes, the items measuring perceived ease of use (PEU) and perceived ease of use and usability (PEUU) were combined into a single category (PEU+PEUU). Given the small sample size, a descriptive approach was used to analyse the data.

Findings and Discussion

The multimedia environment was evaluated using two instruments. In this pilot study, the primary aim of the SUS was to assess the usability of the AR app and to determine its suitability for the

implementation in the main study. Overall, participants evaluated the AR application positively, indicating that it was easy to use and supportive of the learning process (*Table 1*).

Table 1. Students' evaluation of the items in the SUS instrument.

Category	No. of items evaluated good	No. of items asevaluated neutral	No. of items asevaluated poor	Total No. of items
Usability	5	3	0	8
Perceived usefulness	2	0	0	2
Attitude	2	2	0	4
Intension to use	2	1	0	3
Perceived ease of use	2	1	0	3
Total number of items				20

Based on participants' responses across all SUS categories, the AR application demonstrated good overall usability. The students reported positive attitudes toward the use of AR in chemistry learning, expressed willingness to attend additional classes using this technology, and perceived the application easy to use and easy to learn. These findings are consistent with previous studies (Peeters et al., 2022; Akçayir & Akçayir, 2017), which have shown that well-designed AR applications can positively influence students' usability perceptions and learning experiences. From a usability perspective, Brooke (1995) emphasized that effective system should enable task completion, require minimal resources for use and provide satisfying user experience. The results of this study suggest that the developed AR application meets these criteria, as participants found it engaging, motivating and supportive during laboratory work. This is particularly important in chemistry education, where visualization tools are needed to support the integration of macroscopic observations with submicroscopic representations (Gilbert & Treagust, 2009; Johnstone, 1991). However, these findings should be interpreted with caution as this was a pilot study and participants were exposed to AR for the first time. The positive evaluations may be partially influenced by a novelty effect, whereby learners initially respond more favourably to new and unfamiliar technologies (Mazzuco et al., 2022). Consequently, while the current results indicate that the AR application is usable and well accepted, further investigation in a larger scale study is required to evaluate its long-term impact on learning and conceptual understanding.

Results of questions on questionnaire on technology acceptance are presented in *Table 2*.

Table 2. Descriptive results of technology acceptance factors (TAM)

TAM dimension	No. of items	Mean range	Interpretation*
PEU + PEUU (Perceived Ease of Use & Usability)	9	5.3 – 6.6	High
PU (Perceived Usefulness)	5	5.6 – 6.8	High
AT (Attention to Use)	5	3.2 – 6.9	Mainly positive

*Interpretation is based on item mean scores relative to the scale midpoint. Scores above the midpoint are interpreted as positive (Liu et al., 2022).

As shown in *Table 2*, participants responses indicated mainly positive perceptions across all TAM factors. No student reported neutral or negative ratings for any statement related to technology acceptance, suggesting that the AR application did not negatively affect students' engagement

with or understanding of the learning content (Tarnig et al., 2022). These results suggest that the AR application was generally well received. The high PEU + PEUU scores indicate that the students found the system easy to use and supportive of learning, while the high PU scores demonstrate that learners recognized its educational value. The AT scores, although slightly lower for one item, indicate that most participants were willing to engage with the AR application in future learning activities. From a theoretical perspective, the positive usability and technology acceptance results suggest that AR has the potential to support learners in bridging the gap between macroscopic observations and submicroscopic or symbolic representations in chemistry, consistent with the chemical triplet framework (Johnstone, 1991; Gilbert & Treagust, 2009). By providing visualizations of the processes such as the Haber-Bosch reaction, ammonia dissolution, and ion adsorption, the AR application may offer learners interactive ways to explore chemical phenomena that are otherwise difficult to observe, which could be particularly valuable in environmental chemistry education.

Conclusions

This pilot study evaluated a newly developed AR application for teaching the nitrogen cycle in an environmental chemistry module to pre-service primary school teachers. Overall, participants reported positive perceptions of usability, perceived usefulness, and intention to use. The System Usability Scale (SUS) indicated that the application was easy to use, engaging, and supportive of learning, while the Technology Acceptance Questionnaire (TAM) demonstrated that students were willing to engage with the application and recognized its educational value.

The findings suggest that the AR application is well received and has the potential to support chemistry learning, particularly in providing interactive visualizations of chemical processes that are difficult to observe directly. However, future research with a larger sample will allow examination of additional variables, such as conceptual understanding, motivation, and learning outcomes, to more fully assess the educational impact of AR in environmental chemistry. As a pilot study, these findings should be interpreted with caution. The small sample size limits generalizability. Future studies are needed to directly assess whether these visualizations improve conceptual understanding and learning outcomes. Also, previous studies on emerging technologies in education highlighted the importance of longitudinal studies to determine whether positive attitudes toward AR are sustained over time. Therefore, larger study is required to evaluate long term impact on attitudes towards AR technology.

Acknowledgement

The authors acknowledge that the project »Augmented reality to achieve better understanding of the triple nature of chemical concepts« (Grant No. J5-50155) was financially supported by the Slovenian Research and Innovation Agency.

References

- Akçayir, M., & Akçayir, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Education Research Review*, 20, 1–11. <https://doi.org/10.1016/j.edurev.2016.11.002>
- Allen, A. E., Koperová, D., Kuhnová, M. & Rusek, M. (2024). Preservice chemistry teachers' conceptual understanding and confidence judgment: Insights from a three-tier chemistry concept inventory. *Journal of Chemical Education*, 102(1), 53–65. <https://doi.org/10.1021/acs.jchemed.4c01146>
- Ardiny, H., & Khanmirza, E. The role of AR and VR technologies in education developments: Opportunities and challenges. In *Proceedings of the 2018 6th RSI International Conference on Robotics and Mechatronics (Icrom)* (pp. 482–487). Tehran, Iran. [10.1109/ICRoM.2018.8657615](https://doi.org/10.1109/ICRoM.2018.8657615)
- Atkins, P. W., Clugston, M. J., Frazer, M. J., & Jones, R. A. Y. (1995). *Kemija zakonitosti in uporaba*. Tehniška založba Slovenije.

- Azuma, R. T. (1997). A survey of augmented reality. *Presence*, 6(4), 355–383. <https://doi.org/10.1162/pres.1997.6.4.355>
- Bacca, J., Baldiris, S., Fabregat, R., & Kinshuk, K. (2019). Frameworks for designing motivational augmented reality applications in vocational education and training. *Australian Journal of Education Technology*, 35(3), 102–117. <https://doi.org/10.14742/ajet.4182>
- Brooke, J. (1995). *SUS: A quick and dirty usability scale*. London, CRC Press.
- Brooke, J. (1996). “SUS - A quick and dirty usability scale”. *Usability evaluation in industry*, 189(3).
- Coduto, J. R., Lazicki, A., & Leddy, J. (2023). Visualizing 3D objects in analytical chemistry. *Journal of Chemical Education*, 101(1), 77–87. <https://doi.org/10.1021/acs.jchemed.3c00821>
- De Jong, T., Linn, M. C., & Zaccharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, 340(6130), 305–308. [10.1126/science.1230579](https://doi.org/10.1126/science.1230579)
- Erduran, S. (2007). Breaking the law: Promoting domain-specificity in chemical education in the context of arguing about the periodic law. *Foundations of Chemistry*, 9(3), 247–263. <https://doi.org/10.1007/s10698-007-9036-z>
- Ferk Savec, V., Sajovic, I., & Wissiak Grm, K. S. (2009). Action research to promote the formation of linkages by chemistry students between the macro, submicro, and symbolic representational levels. V J. K. Gilbert (eds.), *Multiple Representations in Chemical Education* (pp. 309–331). Springer, Dordrecht.
- Gilbert, J. K., & Treagust, D. (2009). *Models and modelling in science education*. The University of Reading, Institute of Education.
- Gurung, R., & Gurung, B. B. (2023). Laboratory work and its impact in learning chemistry at middle secondary schools of Trngsa, Bhutan. *Asian Journal of Education and Social Studies*, 43(3), 175–184. <https://doi.org/10.9734/ajess/2023/v48i31078>
- Halloun, I. A. & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, 53(11), 1043–1055. <https://doi.org/10.1119/1.14030>
- Harrison, A. G., & Treagust, D. F. (1998). Modelling in science lessons: Are there better ways to learn with models? *School Science and Mathematics*, 98(8), 420–429. <https://doi.org/10.1111/j.1949-8594.1998.tb17434.x>
- Holden, H. & Rada, R. (2014). Understanding the influence of perceived usability and technology self-efficacy on teachers' technology acceptance. *Journal of Research on Technology Education*, 43(4), 343–367. <https://doi.org/10.1080/15391523.2011.10782576>
- Horvat, S., Rodić, D., Jović, N., Rončević, T. & Babić-Kekez, B. (2023). Validation of the strategy for determining the numerical complexity of exam items in the field of chemical kinetics. *CEPS Journal*, 13(4), 111–133. <https://doi.org/10.26529/cepsj.1235>
- Johnstone, A. H. (1982). Macro- and micro-chemistry. *School Science Review*, 64, 377–379.
- Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning*, 7(2), 75–83. <https://doi.org/10.1111/j.1365-2729.1991.tb00230.x>
- Kaya, E., & Erduran, S. (2013). Integrating epistemological perspectives on chemistry in chemical education: The cases of concept duality, chemical language, and structural explanations. *Science & Education*, 22(7), 1741–1755. <https://doi.org/10.1007/S11191-011-9399-3>
- Liu, Q., Ma, J., Yu, S., Wang, Q., & Xu, S. (2022). Effects of an augmented reality-based chemistry experiential application on student knowledge gains, learning motivation, and technology perception. *Journal of Science Education and Technology*, 32(3), 1–15. [10.1007/s10956-022-10014-z](https://doi.org/10.1007/s10956-022-10014-z)
- Mazzuco, A., Krassman, A. L., Reategui, E., & Gomes, R. S. (2022). A systematic review of augmented reality in chemistry education. *Review of Education*, 10(1), e3325. [10.1002/rev3.3325](https://doi.org/10.1002/rev3.3325)
- Mazzuco, A., Krassman, A. L., Reategui, E., & Gomes, R. S. (2022). A systematic review of augmented reality in chemistry education. *Review of Education*, 10(1), e3325. <https://doi.org/10.1002/rev3.3325>
- McMurry, J., in McFay, R. C. (2001). *Chemistry*. Prentice-Hall, New Jersey.
- Mei-Hung, C., Chin-Cheng, C., Yi-Hung, C., TaMin, H., Wie-Tian, T., Jin-Wie, H, Hongming, L. L., & Ming-Kang, T. (2018). Model-based learning about structures and properties of chemical elements and compounds via the use of augmented realities. *Chemistry Teacher International*, 1(1), 20180002. <https://doi.org/10.1515/cti-2018-0002>
- Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994). Augmented reality: A class of displays on the reality/virtuality continuum. *Telem manipulator and Telepresence Technologies, SPIE*, 2351, 282–292. <https://doi.org/10.1117/12.197321>
- Pavlin, J., Glažar, S. A., Slapničar, M., & Devetak, I. (2019). The impact of students' educational background, interest in learning, formal reasoning and visualisation abilities on gas context-based

- exercises achievements with submicro-animations. *Chemistry Education Research and Practice*, 20(3), 633–649. <https://doi.org/10.1039/C8RP00189H>
- Peeters, H., Habig, S., & Fechner, S. (2023). Does augmented reality help to understand chemical phenomena during hands-on experiments? – Implications for cognitive load and learning. *Multimodal Technologies and Interaction*, 7(2), 9. <https://doi.org/10.3390/mti7020009>
- Ribič, L & Devetak, I. (2024). Augmented reality in developing students' understanding of chemistry triplet: A systematic literature review. *Chemistry Teacher International*. <https://doi.org/10.1515/cti-2024-0060>
- Ribič, L., Devetak, I. & Poročnik, R. (2025). Pre-service primary school teachers' understanding of biogeochemical cycles of elements. *Education Sciences*, 15(1), 110. <https://doi.org/10.3390/educsci15010110>
- Sanders, M., & S. McCormick, E. J. (1987). *Human factors and engineering and design* (7. ed.). McGraw-Hili.
- Sjöström, J., Eilks, I., & Talanquer, V. (2020). Didaktik models in chemistry education. *Journal of Chemical Education*, 97(4), 910–915. <https://doi.org/10.1021/ACS.JCHEMED.9B01034>
- Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M. & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of Serbian Chemical Society*, 88(2), 211–221. <https://doi.org/10.2298/JSC220117073S>
- Taber, K. S. (2008). Exploring conceptual integration in student thinking: Evidence from a case study. *International Journal of Science Education*, 30(14), 1915–1943. <https://doi.org/10.1080/09500690701589404>
- Talanquer, V. (2010). Macro, submicro and symbolic: The many faces of the chemistry "triplet". *International Journal of Science Education*, 33(2), 179–195. <https://doi.org/10.1080/09500690903386435>
- Tan, R. M.; Yangco, R. T. & Que, E. N. (2020). Students' conceptual understanding and science process skills in an inquiry-based flipped classroom environment. *Malaysian Journal of Learning Instructions*, 17(1), 159–184. <https://doi.org/10.32890/mjli2020.17.1.7>
- Tarng, W., Lin, Y., & Ou, K. (2021). A virtual experiment for learning the principle of Daniell cell based on augmented reality. *Applied Sciences*, 11(2), 762. <https://doi.org/10.3390/app11020762>
- Tarng, W., Tseng, Y. C., & Ou, K. L. (2022). Application of augmented reality for learning material structures and chemical equilibrium. *Systems*, 10(5), 141. <https://doi.org/10.3390/systems10050141>
- Tarng, W., Tseng, Y. C., & Ou, K. L. (2022). Application of augmented reality for learning material structures and chemical equilibrium. *Systems*, 10(5), 141. <https://doi.org/10.3390/systems10050141>
- Timilsena, N. P., Maharjan, K. B., & Devkota, K. M. (2022). Teachers' and students' experiences in chemistry learning difficulties. *Journal of Positive School Psychology*, 6(10), 2856–2867. <https://journalppw.com/index.php/jpsp/article/view/13764/8923>
- Tsai, C., Ho, Y., & Nisar, H. (2021). Design and validation of a virtual chemical laboratory – An example of natural science in elementary education. *Applied Sciences*, 11(21), 10070. [10.3390/app112110070](https://doi.org/10.3390/app112110070)
- Wilsson, S. M., & Bernie, J. (1999). Chapter 6: Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, 24(1), 173–209. <https://doi.org/10.3102/0091732X0240011>
- Winkelmann, K., Keeney - Kennicut, W., Fowler, D., & Macik, M. (2017). Development, implementation and assessment of general chemistry lab experiments performed in the virtual world of second life. *Journal of Chemical Education*, 94(7), 849–858. [10.1021/acs.jchemed.6b00733](https://doi.org/10.1021/acs.jchemed.6b00733)
- Wu, H. K., Krajcik, J. S., & Soloway, E. (2001). Promoting understanding of chemical presentations: Students' use of a visualisation tool in the classroom. *Journal of Research in Science Teaching*, 38(7), 821–842. <https://doi.org/10.1002/tea.1033>
- Yamtinah, S., Susanti, E. V. H., Saputro, S., Dwi Arian, S. R., Shidiq, A. S., Sari, D. R., & Ilyasa, D. G. (2023). Augmented reality learning media based on tetrahedral chemical representation: How effective in learning Process? *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), 1305–8223. <https://doi.org/10.29333/ejmste/13436>

Beware The Slider's Range: A Design Lesson From Physics Simulation Development

Jared Arnell, Hillary Swanson, Boyd Edwards, Sadra Jafari Ghalehkohneh, Kaden Hart, and John Edwards
Utah State University, United States

Digital simulations are a critical component of modern education, yet guidance for the many microscopic design decisions that accompany their creation is incredibly rare. During our own development of a physics simulation, our team encountered one such decision regarding the range of the sliders in the simulation. We recount how an initial 'intuitive' choice resulted in significant changes to the simulations' functionality, which ultimately overshadowed rich details of the phenomenon. From this experience, we devised a sequence of design principles for choosing slider boundaries, which includes stating learning goals, identifying necessary inclusions/exclusions, quantifying relationships, and checking maximizing/minimizing cases. Through this explication, we illuminate the utility of small-scale design principles for creating digital simulations for educational applications.

Keywords: Physics, Simulations, Design Principles

Introduction

Digital educational simulations play a key role in modern STEM education and have proved extremely effective at improving students' conceptual understanding (Rutten et al., 2012; Wieman & Perkins, 2005). Today, a wide array of such simulations exists, covering a multitude of domains and topics including physics, chemistry, and mathematics (Banda & Nzabahimana, 2021; Fan & Geelan, 2013; Molina-Toro et al., 2019; Moore et al., 2014; Yuriy et al., 2022). Given their resounding empirical support and wide applicability, developing educational simulation technologies remains a vital goal in educational research.

Although there are countless studies examining the application and effectiveness of digital simulations across disciplines, the literature on how to *create or design* such simulations remains drastically underdeveloped by comparison. There are myriad instructional design methodologies which are (in theory) generalizable, but the unique challenges presented in developing a computer-based technology mean many of these methods lack the specificity or detail to be helpful at a granular level. For example, Becker & Parker (2012) synthesized general principles from instructional and game design frameworks to develop an overarching design process for digital simulations. Their process—which follows a familiar phase-based plan with such recognizable steps as “preparation,” “design,” and “testing”—may offer a helpful roadmap for large-scale planning, but its limited guidance within these phases leaves designers on their own when facing the deluge of choices and commitments that must be made when building a simulation from the ground up.

Some researchers have sought to look across empirical studies which showcase digital simulations in educational environments in order to construct a sort of “best practices” list of affordances and features employed by effective simulations. Blake & Scanlon (2007) compiled one such list by comparing three simulations—two physics-based and one biology-based—and concluded that effective simulations must a) be based on real-world phenomena, b) use multiple representations like graphs and visualizations, and c) offer appropriate support and scaffolding to guide inquiry. A similar analysis by Podolefsky et al. (2010) presented a detailed case study of

two students using a physics simulation and explored how the simulation's ability to make invisible phenomena observable and provide real-time feedback contributed to the students' learning. These analyses begin to identify design priorities and goals that are specific to simulation development, but their advice still leans towards the birds-eye view—primarily considering the targets of study for simulations and the instructional environment in which they will be situated.

A larger study was performed by Scalise et al. (2011), who reviewed 79 studies of virtual labs and simulations in secondary school science classrooms. From their analysis, they composed three rubrics: one for software/hardware elements, one for the visualizations themselves, and one for the scientific inquiry which the simulation scaffolded. While many elements of their rubrics arrive at the same conclusions of the works mentioned previously (such as showing scale models of large/small phenomena or providing lesson plans), they also unpack several fine-grain components of simulation design, like API compatibility and the inclusion of auditory elements. Such atomized dissections of minute components of simulation design are quite rare and often only mentioned in passing. For instance, Podolefsky et al. (2010) did mention how the affordance of the pause button was useful for one student, but an interrogation of the simulation at that scale was obviously not their main focus.

One examination of simulation design at this grain-size is found in Williams (2003), who drew upon theories from design and learning research to critique the common practice of defaulting to interface choices which seem to be the most intuitive at first glance. Instead, Williams examines the ways in which miniscule trade-offs in modes of interaction—like swapping direct manipulation of an object with typing in a command prompt or delaying the speed at which the simulation updates or responds to user inputs—may have a subtle but significant effect on the learner's cognitive experience. While Williams still devotes substantial time to discuss larger-scale issues (like the designer's "intent"), his considerations on interface design remain one of the few instances of such zoomed-in granularity.

By far, the most detailed guidance on small-scale design choices for digital simulations comes from the Physics Education Technology group (PhET, n.d.), who have published in-depth accounts of both their approach to designing educational simulations (Adams et al., 2008b) as well as implementing them in practical settings (Adams et al., 2008a). These treatises condense results from over 200 clinical interviews of students interacting with digital simulations, and so touch on many of the themes already mentioned: real-world connections, guiding prompts, and scaffolding inquiry. In addition, Adams et al. (2008b) shed much-needed light on a variety of design choices that arise during the piece-by-piece construction of such computer simulations: default/start-up settings, stylization, color choice, distinguishing "control panels" from "play areas," and even the use of backgrounds.

One of the conclusions Adams et al. (2008) draw regarding interface design is that controls which "resemble using the mouse as a simple extension of direct manipulations by hand" (p. 568) are particularly intuitive. This philosophy of prioritizing direct manipulation in simulations has been understood for some time (Shneiderman, 1997), but Adams et al. go a step further to share at length the subtle differences they observed between users' interactions with various types of manipulatable elements, such as click-and-drag objects, radio buttons, and checkboxes. For designers new to the computer simulation landscape, deconstructions like these can prove to be an invaluable asset.

For the last three years, our research team has embarked on just such a maiden voyage into digital simulation design. Our goal has been to develop simulations for upper-division physics curricula

(Jafari Ghalehkohne et al., 2025), a niche we feel is critically underexplored. In our nascent efforts to learn the tricks and tools of the trade, we inadvertently stumbled into a design oversight which, in hindsight, may provide a useful example for unwrapping a previously undiscussed element of simulation micro-design: the decision of what *range* to use on grabbable sliders. By recounting the factors which led to this hurdle, the turbulence it caused, and the way we corrected it, we hope to contribute to the literature on micro-scale design choices for digital simulations.

In part one of this work, we will relate the development process of our first simulation, the circumstances of our design oversight, and the impact it had on users during our pilot interviews. In part two, we will outline how we used this learning opportunity to construct a sequence of design principles for decision making in this narrow but critical element of simulation design.

Part One: Simulation And Study Context

Our research group has been developing a suite of physics simulations called PhysMath (PhysMath, n.d.), which go beyond traditional visualizations of physics phenomena to incorporate interactive equations as well. These mathematical inclusions aim to address a distressingly common finding in physics education research: students who can successfully perform mathematical tasks in physics contexts often have little to no ability to explain the purpose or meaning of such actions (Arnell, Swanson, Edwards, & Edwards, 2024; Mason & Singh, 2010; Tuminaro & Redish, 2007).

Our development team consists of individuals from three fields: physicists as content experts, computer scientists to build and program the simulations, and learning scientists to base the design in education theories and conduct user testing. Our design process begins with the physicists crafting a detailed explanation of the phenomenon, which is communicated to the computer scientists to develop a basic (functional) simulation. Then, the simulation is presented to the team for feedback to improve its aesthetics and usability. When the team feels the simulation has reached a satisfactory level of polish after several iterative refinements, the learning scientists conduct clinical interviews with volunteer students to observe their behaviour and receive user feedback.

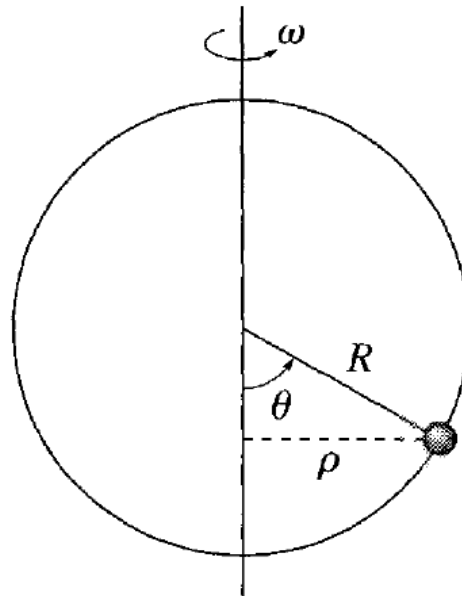
The first simulation we chose to design was the “bead on a spinning wire hoop” problem (Taylor, 2005, p. 261) from Classical Mechanics. We chose this target because it was conceptually and mathematically rich yet nearly impossible to demonstrate physically (Jafari Ghalehkohne et al., 2025). The design experience we share in this work took place during the earliest versions of this simulation. In the following section, we will cover the necessary background for understanding the phenomenon and how a seemingly innocuous design choice on slider ranges compromised our simulation’s instructional utility.

The PhysMath Bead-On-Hoop Simulation

Fortunately, a rigorous mathematical understanding of the Bead-on-Hoop is not required for our purposes in this work; instead, a cursory, qualitative explanation will suffice. For this problem, a bead is threaded onto a wire hoop with radius R , which is then spun around a vertical axis with angular velocity ω , as diagramed in Figure 1. The position of the bead at any moment can be described with a single parameter: the angle θ which the bead makes with the bottom of the hoop.

As the bead is allowed to slide around the hoop, its motion is governed by two forces. The first is gravity (traditionally symbolized by g), which always points downward and has a constant magnitude. The second force is the centrifugal force, which always points horizontally outward away from the axis of rotation and gets larger as the hoop spins faster and as the bead moves further away from the axis of rotation (a distance denoted by ρ in Figure 1).

Figure 1: The bead on a spinning hoop (Taylor, 2005, p. 261)



On its own, the gravitational force would pull the bead to the lowest point on the hoop (at $\theta = 0^\circ$), and this is indeed the primary equilibrium point for the bead when the hoop is not spinning. As the hoop begins to spin (and if the bead is not perfectly at $\theta = 0^\circ$), the centrifugal force will begin to grow. For small values of ω , the centrifugal force will be present but too small to overcome gravity, and the bead will continue to find equilibrium at the bottom of the hoop in spite of the hoop's rotation. But as the hoop spins faster, the centrifugal force can eventually grow strong enough to overcome gravity. When this occurs, the bead will rise away from $\theta = 0^\circ$. As the bead approaches $\theta = \pm 90^\circ$, the centrifugal force—although growing larger due to the increased distance ρ from the rotational axis—will approach an angle normal to the hoop, causing its influence on the bead's motion around the hoop to diminish. At the same time, the gravitational force will approach an angle which is tangential to the hoop and thus become more influential. This shifting balance will result in some 'critical angle' θ_{crit} at which the effects of gravity and the centrifugal force are equivalent, and the bead can remain stationary at this position on the hoop indefinitely; as the hoop rotates faster, the critical angle will approach 90° . Importantly, this critical angle can only exist if the centrifugal force is sufficiently strong, and so there is some 'critical angular velocity' ω_{crit} at which the hoop must spin before the bead can achieve this additional equilibrium point.

This rich bifurcation of behaviour was one of the reasons the bead-on-hoop problem was such a tantalizing target for our simulations. Our hope was that students could use the PhysMath simulation to explore how the parameters of the system—namely, the strength of gravity and the size of the hoop—would affect this critical angular velocity and observe how the motion of the bead is significantly altered as the rotation rate of the hoop crosses this crucial threshold.

As our programmers crafted the earliest versions of the PhysMath Bead-on-Hoop simulation, they followed the well-established practices for physics simulations (Adams et al., 2008b) and included draggable sliders for each of the system's parameters. Unfortunately, our team had not had any explicit discussion about what the minimum and maximum values for each slider should be. As this was the first PhysMath simulation to be built and there was no previous experience to draw upon, the programmers chose values which seemed intuitive: a minimum of 0 (0.1 for the radius slider, which cannot be 0) and a maximum of 100.

While these values may be sensible in other circumstances, they turned out to be quite problematic for the Bead-on-Hoop in particular, especially for the radius slider! The magnitude

of the centrifugal force is dependent on ρ , and thus dependent on R as well. At a radius of only a few meters, a hoop could rotate at an incredibly small rate and still create a substantial centrifugal force. This would mean that the critical angular velocity ω_{crit} —the threshold point around which we wanted our students’ explorations to revolve—resides so close to zero that it practically vanishes. With the slider’s maximum set at 100, even positions on the slider that would intuitively appear to be quite low (e.g., 5–10% of the slider’s total length) would still represent radii that were inappropriately large and would result in an overwhelming centrifugal force.

Unfortunately, the rest of our design team did not recognize the eventual consequences of these slider ranges during our rounds of iterative feedback on the programmer’s prototypes. It was only when our learning scientists sat down with students to observe their interactions with the simulation that the impact of the slider ranges was revealed. In the next section, we will share excerpts from the students’ interviews to demonstrate how our design oversight hindered their explorations into the simulated phenomenon.

Pilot Study Observations

Eventually, the simulation prototype was ready for user testing. Students who were enrolled in Classical Mechanics at our institution were recruited for clinical interviews (Arnell, Swanson, Edwards, Hart, et al., 2024), in which they shared their understanding of the phenomenon and then explored the PhysMath Bead-on-Hoop simulation. On the whole, the students’ experience was positive, and our observations concurred with Adams et al. (2008b) regarding the intuitiveness of controls that mimic physical manipulations. However, during these pilot interviews, we finally noticed the size of the radius slider impeding the students’ learning. Here, we will share excerpts from interviews with four students: Alena, Armando, Jake, and Ronan (pseudonyms).

Alena’s first act when she began exploring the simulation was to change the radius of the hoop, saying, “what happens if we make [the radius]... like, five?” As the hoop spun, the bead sat relatively stable near 90° , at which Alena commented, “It seems to just hang out right here.” She then began increasing and decreasing ω , but without readjusting R to a smaller value. When this had little effect on the bead, Alena changed the bead’s starting position before varying ω again. All of Alena’s experiments yielded the same result: the bead found its equilibrium near 90° , regardless of any other system parameter. Alena concluded that “if you start it at the halfway mark, it will tend to stay there”—a conclusion which is only true if the hoop is spinning above the critical angular velocity.

Armando was also drawn to the radius slider first, saying “let’s increase R , see what that does.” He dragged the slider up about a third of the way (34 m). After watching the simulation for a moment, he moved on to other variables without resetting R . Armando played around with the sliders for gravity, friction, and angular velocity, checking the hoop’s behaviour when each slider was near its minimum and maximum; in all cases, the bead’s equilibrium point never shifted from the super-critical 90° position. Armando commented, “it definitely has this critical point right there, at the very side... I wonder if I increase gravity, if the position at that point changes at all?” Armando’s intuition was aimed in a productive direction: a stronger-than-usual gravitational force might overpower the centrifugal force enough for him to find sub-critical behaviours. However, the influence of the absurdly large radius was so great that even when Armando slid the gravity slider to near its maximum (around 75 m/s^2), the bead’s behaviour remained unchanged. Armando responded “No, it doesn’t. It still stays there.” In his experimenting, the only time Armando saw the bead come to rest at the bottom of the hoop was when he set ω to zero—a trivial solution, as *no* rotation results in *no* centrifugal force.

Jake took a unique approach, varying some of the last sliders on the list first; this meant he only reached the radius slider after he had already experimented with the starting position and velocity, angular velocity, and friction. Jake increased R up to around 30 m, then decreased it as he had the others. However, due to the length of the slider, Jake's 'low' value was still 8 m (giving a hoop as tall as a two-story building!). As Jake continued to experiment with the simulation, he tried removing gravity by setting g to zero. When the bead's equilibrium stayed around 90° , he remarked "it doesn't change it terribly much."

Ronan made the typical choice to interact with the radius slider first; he dragged the slider to its absolute maximum, then back to a lower value—but ultimately left it at 20m. From there, Ronan varied the gravity, friction, angular velocity, and starting position/velocity sliders. Through all his experiments, the bead maintained an equilibrium point at roughly 90° . Summing up his observations, he said "there's this side motion... it's the furthest to the side that it can be, and it just stays there."

The interviewer (first author) noticed the issue of the range slider's excessive maximum after the interviews with Alena and Armando. Unfortunately, an updated version of the simulation was not available in time for the interviews with Jake and Ronan, so the interviewer elected to not immediately intervene if the participants unwittingly set R too high. This was done to give the students an opportunity to detect the error and correct it themselves, though neither did. Eventually, the interviewer stepped in to correct the issue for Jake and Ronan once they needed to move on to the next phase of the interview.

Alena, Armando, Jake, and Ronan all shared a very similar experience during their exploration of the PhysMath Bead-on-Hoop simulation. They each increased R from its default value and performed subsequent tests while R remained at a relatively high value. In spite of their attempts at varying many of the other parameters, the bead's equilibrium at approximately 90° could not be disturbed (outside of removing the rotation of the hoop altogether). As they observed the system's behaviour, they all made comments regarding the apparent permanence of the bead's equilibrium position.

From these pilot interviews, it was immediately clear that the range of the radius slider (and possibly others) must change. Not only had the students been unable to discover any of the sub-critical behaviour examples, but they appeared to conclude from this that such behaviour may not even exist! This tendency for students to trust simulations to a fault was also observed by Adams et al. (2008a), who noticed that students may even attribute meaning to cases when simulations broke or encountered bugs, believing that the simulations *must* be portraying the phenomenon correctly.

What was far less clear to our design team was what the new slider range should be. In part two, we will share how we constructed a sequence of design principles for making such a determination—a sequence we hope will be useful for any designers who may be struggling with a similar decision.

Part Two: Design Principles For Determining Slider Ranges

For digital simulation designers who must choose upper and lower bounds for sliders representing physical parameters in a modelled phenomenon, we offer the four-stage design principle sequence we constructed from our problem-solving experience: 1) state your learning objective, 2) designate allowed/disallowed input values, 3) quantify your parameter relationships, and 4) use maximizing/minimizing cases to calculate slider boundaries.

1. State Your Learning Objectives And How They Relate To The Slider Values

As is the case with many instructional design methods (Branch, 2009), our first priority was to make our learning objectives explicit. As mentioned previously, our intent was that students who were exploring the PhysMath Bead-on-Hoop simulation should be able to find angular velocity values which were both above and below ω_{crit} , allowing them to observe and compare the differences in the bead's behaviour on either side of this threshold. To make these kinds of discoveries as natural as possible, our preference would be for ω_{crit} to fall somewhere near the middle of the angular velocity slider and remain away from the slider's edges as much as possible as the other system parameters were varied. This would increase the likelihood of users finding both categories of motion organically as they instinctively check the limits of the simulation (Adams et al, 2008a).

With this target articulated, we realized that we would need to adjust not only the scale of the radius slider which had been the apparent cause of the issue, but also the angular velocity slider so that it was scaled correctly to accommodate whatever ω_{crit} values were made possible by the values of the other system parameters. Similarly, we would also need to fine-tune the range of our gravity slider, as g affects ω_{crit} by providing a benchmark for the growing centrifugal force to overpower.

And so, our ultimate goal would be to find minimum and maximum boundaries for this trio of interconnected parameters that would restrict the possible ω_{crit} values to a reasonable range as near the middle of the angular velocity slider as we could manage.

2. Designate Specific Input Values Which Must Be Allowed/Disallowed

With the radius, gravity, and angular velocity sliders in our sights, our next step was to establish if there were any specific values for these three parameters that we must include (or exclude) for the simulation to make sense. We agreed that we must let g go to zero, as we wanted to allow students to “turn off” gravity and explore how the complete removal of that force affected the system. Congruently, we needed to allow for ω to have a minimum of zero to let students observe the (albeit trivial) cases when the hoop is not spinning and there is no centrifugal force. On the other hand, R could not be allowed to reach zero, as doing so would essentially ‘break’ the formulae used to calculate the bead's motion; thus, the radius slider would need a non-zero minimum value. Our final observation at this stage was that the gravity slider needed to include the value 9.8, as this is the standard g value for earth; we would then prefer that the gravity slider extend beyond that point by some amount to allow for testing gravitational forces which were stronger than earth's.

By clarifying these must-have points, we had already established the minimum for two of our three sliders, as well as getting an approximate feel for where the maximum for g and minimum for R may wind up. To find the rest of the boundaries, we would need to quantify the way changes to any one parameter would affect the others.

3. Quantify The Relationship Between Your Parameters

The relationship between the radius, gravity, and the critical angular velocity for the Bead-on-Hoop is given by Equation 1:

$$\omega_{crit}^2 = \frac{g}{R} \quad (1)$$

This expression quantifies the proportionalities we have already mentioned, and so now allows us to plug in individual values to calculate an exact critical angular velocity. By plugging in

hypothetical maximum and minimum values for g and R , we can construct a range of possible values for ω_{crit} and iteratively home in on a set of ranges for all three parameters which coincides as closely to our desired learning objectives as we can manage.

4. Use Maximizing/Minimizing Cases To Determine Slider Boundaries

Equation 1 can be looked at in two ways: trying to either maximize or minimize ω_{crit} . For our calculations, we first looked at the cases where ω_{crit} would minimize (in hindsight, if we had tried to start with the maximizing case, we would not have had enough information yet and would have resorted to the minimizing case instead).

As we had already established a minimum value of zero for the gravity slider, we began by checking how this boundary affected the possible values for ω_{crit} . Unfortunately, Equation 1 demonstrated that ω_{crit} would approach zero as g does, which means that ω_{crit} would inevitably drift to the bottom edge of the angular velocity slider regardless of the slider's overall range. While this goes against our preference to keep ω_{crit} away from the slider's edge, we realized that there would be no way to avoid this particular case without abandoning our commitment to allow g to go to zero, and so we would need to try to work around this case as best we could.

Since we could not prevent ω_{crit} from reaching zero, we elected to check what would happen at a small but non-zero value for g (we chose $g = 1$). Our calculations revealed that if we wanted to keep the minimum ω_{crit} values out of the decimals, then the absolute maximum for our radius slider would need to be 1—which speaks to why our student users were so unable to find any sub-critical behaviour with their R set so much higher! If our radius slider has a maximum of 1, then a value for g closer to the default/standard would yield $\omega_{crit} \approx 3$. This let us know that our angular velocity slider maximum shouldn't be so large as to make values in the low, single-digit range imperceptible.

With the maximum R value established, we could now look to the maximizing cases of Equation 1 to find the last of our boundaries. If we kept g at a standard value and made R smaller, then by the time we reached $R = 0.1$, ω_{crit} had already climbed to around 10. Knowing that ω_{crit} would climb even higher if g was increased as well, we felt 0.1 would be a sufficient minimum for the radius slider to keep it from pushing ω_{crit} too large.

At this point, the only boundaries we had not established were the maximums for the gravity and angular velocity sliders. As Equation 1 shows, ω_{crit} and g are directly proportional, so allowing higher values of g would continue to raise the maximum possible value for ω_{crit} . This meant the maximum values for these sliders could be as low or high as we preferred. We resolved to cap the two sliders low enough that single-digit values for ω_{crit} wouldn't be completely washed out, and so chose 20 for a maximum ω and 25 for a maximum g (though these values could have been slightly different and the overall functionality of the simulation would remain relatively unchanged). These particular values meant that the maximum possible ω_{crit} would be around 16, which we felt would keep ω_{crit} acceptably far away from the top edge of the angular velocity slider in most cases.

Conclusion

In this work, we have shared an account of a challenge our research team encountered while designing a physics simulation: due to the interdisciplinary nature of the task, our programmers chose values for the range of our simulation's parameters sliders which were ill-suited for the modelled phenomenon. When this oversight was not noticed by our team's physicists and learning scientists, we conducted pilot interviews in which the improper slider ranges had a significant impact on the students' exploratory sessions with the simulation. Learning from this

mistake, we have since generated a sequence of design principles to help other designers calculate optimal slider ranges for their own educational simulations. This sequence asks designers to articulate their learning objectives, designate necessary input values, quantify their parameter relationships, and using maximal/minimal cases to arrive at slider ranges based on concrete justifications.

Through this experience, our team came to appreciate how minor details in interface design can have a substantial effect on the utility of educational simulations—especially when modelling complex phenomena. While there are simulation design teams which have been open about sharing their own design lessons (Adams et al., 2008a; 2008b), the fine-grain operationalization of design choices for computer-based simulations could use substantially more explication. We hope our detailed deconstruction and analysis of the process and rationalization behind one micro-design choice can assist other educational simulation designers when faced with the same conundrum. Furthermore, we encourage the publication of similar micro-design deconstructions to illuminate the otherwise imperceptible challenges presented by digital simulation design.

Acknowledgement

The PhysMath research project was supported by the National Science Foundation (2235569).

References

- Adams, W. K., Reid, S., LeMaster, R., McKagan, S. B., Perkins, K. K., Dubson, M., & Wieman, C. E. (2008). A Study of Educational Simulations Part I — Engagement and Learning. *Journal of Interactive Learning Research*, 19(3), 397–419. <https://www.learntechlib.org/primary/p/24230/>
- Adams, W. K., Reid, S., LeMaster, R., McKagan, S. B., Perkins, K. K., Dubson, M., & Wieman, C. E. (2008). A Study of Educational Simulations Part II — Interface Design. *Journal of Interactive Learning Research*, 19(4), 551–577. <https://www.learntechlib.org/primary/p/24364/>
- Arnell, J., Swanson, H., Edwards, B., Hart, K., Ghalehkohneh, S. J., & Edwards, J. (2024). “Instead of gravity pointing down, it’s now pointing up”: Enhancing physics students’ connection between mathematics and mechanism. In Q. X. Ryan, A. Pawl, & J. P. Zwolak (Eds.), *2024 Physics Education Research Conference Proceedings* (pp. 28–33). Boston, Massachusetts: American Association of Physics Teachers. <https://doi.org/10.1119/perc.2024.pr.Arnell>
- Arnell, J., Swanson, H., Edwards, B., & Edwards, J. (2024). Operational routes: A construct for characterizing intermediate learners’ cognitive navigational strategies. In R. Lindgren, R. I. Asino, E. A. Kyza, C. K. Looi, D. T. Keifert, & E. Suárez (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences—ICLS 2024* (pp. 825–832). Buffalo, New York: International Society of the Learning Sciences. <https://doi.org/10.22318/icls2024.764144>
- Banda, H. J., & Nzabanimana, J. (2021). Effect of integrating physics education technology simulations on students’ conceptual understanding in physics: A review of literature. *Physical Review Physics Education Research*, 17(2), 023108. <https://doi.org/10.1103/PhysRevPhysEducRes.17.023108>
- Becker, K., & Parker, J. (2012). Serious instructional design: ID for digital simulations and games. In P. Resta (Ed.), *Proceedings of SITE 2012—Society for Information Technology & Teacher Education International Conference* (pp. 2480–2485). Association for the Advancement of Computing in Education. <https://www.learntechlib.org/primary/p/39955/>
- Blake, C., & Scanlon, E. (2007). Reconsidering simulations in science education at a distance: Features of effective use. *Journal of Computer Assisted Learning*, 23(6), 491–502. <https://doi.org/10.1111/j.1365-2729.2007.00239.x>
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer New York, NY. <https://doi.org/10.1007/978-0-387-09506-6>
- Candido, K. J. O., Gillesania, K. C. C., Mercado, J. C., & Reales, J. M. B. (2022). Interactive simulation on modern physics: A systematic review. *International Journal of Multidisciplinary: Applied Business and Education Research*, 3(8), 1452–1462. <https://doi.org/10.11594/ijmaber.03.08.08>
- Fan, X., & Geelan, D. (2013). Enhancing students’ scientific literacy in science education using interactive simulations: A critical literature review. *Journal of Computers in Mathematics and Science Teaching*, 32(2), 125–171. <https://www.learntechlib.org/primary/p/39468/>

- Jafari Ghalehkohneh, S., Arnell, J., Hart, K., Swanson, H., Edwards, B., and Edwards, J. (2025). Upper-division physics simulations with equation manipulation. *Physical Review Physics Education Research*, 21(1), 010151. <https://doi.org/10.1103/physrevphyseducres.21.010151>
- Mason, A., & Singh, C. (2010). Surveying graduate students' attitudes and approaches to problem solving. *Physical Review Special Topics—Physics Education Research*, 6(2), 020124. <https://doi.org/10.1103/PhysRevSTPER.6.020124>
- Molina-Toro, J. F., Rendón-Mesa, P. A., & Villa-Ochoa, J. A. (2019). Research trends in digital technologies and modelling in mathematics education. *EURASIA Journal of Mathematics, Science and Technology Education*, 15(8), em1736. <https://doi.org/10.29333/ejmste/108438>
- Moore, E. B., Chamberlain, J. M., Parson, R., & Perkins, K. K. (2014). PhET interactive simulations: Transformative tools for teaching chemistry. *Journal of Chemical Education*, 91(8), 1191–1197. <https://doi.org/10.1021/ed4005084>
- PhET. (n.d.). PhET: Free online physics, chemistry, biology, earth science, and math simulations. PhET Interactive Simulations. <https://phet.colorado.edu/>
- PhysMath (n.d.). PhysMath interactive physics simulations. PhysMath. <https://physmath.usu.edu/>
- Podolefsky, N. S., Perkins, K. K., & Adams, W. K. (2010). Factors promoting engaged exploration with computer simulations. *Physical Review Special Topics—Physics Education Research*, 6(2), 020117. <https://doi.org/10.1103/PhysRevSTPER.6.020117>
- Rutten, N., Van Joolingen, W. R., & Van Der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136–153. <https://doi.org/10.1016/j.compedu.2011.07.017>
- Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., & Irvin, P. S. (2011). Student learning in science simulations: Design features that promote learning gains. *Journal of Research in Science Teaching*, 48(9), 1050–1078. <https://doi.org/10.1002/tea.20437>
- Shneiderman, B. (1997). Direct manipulation vs. interface agents. *Interactions*, 4(6), 42–61. <https://doi.org/10.1145/267505.267514>
- Taylor, J. R. (2005). *Classical Mechanics*. University Science Books.
- Tuminaro, J., & Redish, E. F. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics - Physics Education Research*, 3(2), 020101. <https://doi.org/10.1103/PhysRevSTPER.3.020101>
- Wieman, C., & Perkins, K. (2005). Transforming physics education. *Physics Today*, 58(11), 36–41. <https://doi.org/10.1063/1.2155756>
- Williams, V. (2003). Designing simulations for learning. *E-Journal of Instructional Science and Technology*, 6(1).
- Yuriy, R., Huzchenko, S., Lobach, N., Karbovanets, O., Bokova, S., & Isychko, L. (2022). Modern digital learning and simulation technologies in higher medical education: Definitions, innovative potential. *Amazonia Investiga*, 11(60), 53–61. <https://doi.org/10.34069/AI/2022.60.12.6>

What Factors Influence Pre-Service Science Teachers' Ethical Use Of ChatGPT?

Radu Bogdan Toma and Iraya Yáñez-Pérez

Department of Specific Didactics, Area of Science Education, University of Burgos, Spain

This study examines factors influencing pre-service science teachers' ethical use of ChatGPT using the Theory of Reasoned Goal Pursuit (TRGP). Through directed content analysis of responses from 28 students, key determinants related to procurement goal beliefs, approval goal beliefs, attitudes, subjective norms, motivation, and perceived behavioural control are identified. The findings offer insights into how these constructs shape ethical engagement with ChatGPT, highlighting the need for targeted interventions to foster responsible use in education

Keywords: ChatGPT, academic integrity, pre-service teachers

Introduction

Large language models (LLMs), such as ChatGPT, are significantly reshaping how students approach academic tasks (Kikalishvili, 2023). These tools assist in developing teaching units, rubrics, and quizzes (Cooper, 2023). The growth of these models also relates to increased academic dishonesty and plagiarism (Kasneci et al., 2023; Milano et al., 2023). Recent studies confirm these trends in educational settings (Toma & Yáñez-Pérez, 2025). There is a clear need to investigate factors influencing ethical use.

Against this background, this study applies the Theory of Reasoned Goal Pursuit (TRGP) to explore determinants of ethical ChatGPT use (Ajzen & Kruglanski, 2019). The TRGP connects personal goals with social and environmental beliefs, which could position this framework as useful in explaining how students use LLM. The model suggests that behaviour results from the interaction between desired outcomes and perceived pressures. Therefore, the study examines proactive ethical use of LLM rather than just misuse. Specifically, the study aims to identify factors that foster ethical use during professional development. Therefore, it does not focus on predicting academic misconduct. In doing so, finding might inform design of professional development programs focused on fostering an ethical use of LLM like ChatGPT. Indeed, understanding these factors is necessary for teacher education programs. Pre-service teachers must learn to balance efficiency with academic standards. The results of this study can guide policy making within universities. This study addressed the following research question:

- What factors drive the ethical use of ChatGPT?

Theoretical Underpinnings

The TRGP is a recently developed comprehensive framework designed to predict and modify behaviours (Ajzen & Kruglanski, 2019). It integrates components from Ajzen's (1991) Theory of Planned Behaviour (TPB) and Kruglanski et al.'s (2002) Goal Systems Theory (GST) to incorporate the influence of individuals' goals. TRGP proposes that behaviours are primarily driven by two types of goals: (1) procurement goal beliefs and (2) approval goal beliefs. Procurement goal beliefs refer to the desired outcomes individuals aim to achieve through their behaviour, serving as determinants of attitudes. Approval goal beliefs, in contrast, represent the significant others whose approval individuals seek by performing the behaviour, acting as antecedents of subjective norms.

The theory suggests that procurement goal beliefs exert a stronger influence on attitudes than behavioural beliefs, while approval goal beliefs have a greater impact on subjective norms than normative beliefs. Consequently, individuals are more likely to adopt a behaviour if it aligns with both procurement and approval goals. For example, students may aim to complete university assignments efficiently (procurement goal) and gain peer approval (approval goal). Positive beliefs about ChatGPT, such as improved understanding of material, may not motivate its use if procurement goals remain unmet. Similarly, normative beliefs regarding individuals whose approval is not sought, such as professors, may not significantly influence behaviour.

The theory also posits that attitudes and subjective norms form the foundation for motivation to engage in a behaviour. Within this framework, motivation is defined as “a strong desire to perform the behaviour” (Ajzen & Kruglanski, 2019, p. 775). Higher levels of motivation are expected to lead to the formation of intentions, which subsequently predict and influence behaviour. Additionally, perceived behavioural control moderates motivation and also directly affects intentions. For instance, pre-service teachers with positive attitudes toward the ethical use of ChatGPT and strong subjective norms will exhibit greater motivation to use ChatGPT ethically. This motivation is further strengthened and moderated by high levels of perceived behavioural control over the behaviour.

Methodology

Participants

Participants were selected using purposive sampling to ensure experience with ChatGPT’s ethical use (Cohen et al., 2018). Recruitment criteria included: (1) undergraduate status, (2) attendance in an in-person course, (3) willingness to create a ChatGPT account and participate, and (4) extensive experience in ethical ChatGPT use.

Eligible students completed an intervention on ethical ChatGPT use for teaching science (Toma & Yáñez-Pérez, 2024). Of 37 first-year teaching students recruited, nine were excluded: six did not meet the attendance criterion, and three failed the experience criterion. The final sample included 28 students (89.3% female, mean age = 21.93, SD = 1.80). For university entrance, most pursued social sciences (67.9%), with others in science (17.9%) or vocational training (14.3%). Participants completed an elicitation questionnaire post-intervention.

Data Collection Instrument

An elicitation questionnaire, following standardized protocols (Fishbein & Ajzen, 2010), included eight open-ended questions. Participants were first provided a definition of ethical ChatGPT use: “Ethical use involves seeking information, clarifying concepts, or generating ideas for inspiration, as done in this course. Fraudulent use includes copying ChatGPT content or having it complete assignments”.

Participants listed their most salient procurement goal beliefs ('What is the most important personal goal you could achieve by using ChatGPT ethically this semester?') and approval goal beliefs ('Would you use ChatGPT ethically to gain approval from specific individuals or groups? List them.'). They then identified advantages and disadvantages of ethical ChatGPT use ('What are the advantages and disadvantages of using ChatGPT ethically this semester?'). Normative beliefs were elicited by listing individuals or groups who would approve or disapprove of their ethical use ('Who would approve or disapprove of your ethical use of ChatGPT this semester?'). Finally, control beliefs were identified by listing factors facilitating or hindering ethical use ('What factors would make it easier or harder to use ChatGPT ethically this semester?').

Data Analysis

Directed content analysis was employed to analyse students' responses based on predefined TRGP belief categories (Cohen et al., 2018). Guided by theory, two researchers independently coded responses into categories: personal goals, positive/negative behavioural beliefs, approval-seeking individuals, approval/disapproval normative influences, and facilitating/hindering control beliefs. Inter-rater agreement exceeded 80%, indicating high consistency. Discrepancies were resolved through discussion until consensus was achieved. Modal salient beliefs—determinants of ethical use—were identified using frequency counts, retaining those mentioned by at least 30% of participants.

Findings

A total of 26 beliefs were identified: 2 personal goals, 4 advantages, 3 disadvantages, 2 approval-seeking referents, 4 approving normative referents, 2 disapproving normative referents, 4 facilitating factors, and 5 barriers to ethical ChatGPT use (Table 1).

Table 1. Salient beliefs

Constructs	Belief category	Beliefs	N	%	
Procurement goal	Personal goal	Pass course	24	85.7	
		Save time on assignments	18	64.3	
Behavioural beliefs	Advantages	Pass course assignments with less effort	26	92.9	
		Improve grades	19	67.9	
		Creative ideas for assignments	16	57.1	
	Disadvantages	Improve understanding of course material	10	35.7	
		Crashed/inoperative	22	78.6	
		Risk of plagiarism	18	64.3	
Approval goal	Gain approval	Generic/unhelpful answers	11	39.3	
		University professors	19	67.9	
		Classmates	11	39.3	
		Classmates	23	82.1	
Normative	Approve	Parents	16	57.1	
		University professors	11	39.3	
		Employers	9	32.1	
		Disapprove	University professors	20	71.4
	Control beliefs	Easier	Employers	12	42.9
			Supportive professors/allowing its use	25	89.3
Clear regulation for use			17	60.7	
Difficult		Learn prompts/guidance for ethical use	12	42.9	
		Relevant/useful assignments	9	32.1	
		Time constraints	22	78.6	
		Unrealistic/challenging assignments	19	67.9	
		Excessive assignments	16	57.1	
		Ambiguity about ethical vs. unethical use	14	50	
		Assignments with a low-grade contribution	11	39.3	

Participants primarily pursue ethical ChatGPT use to achieve personal and academic goals. Twenty-four participants (85.7%) identify passing the course as a procurement goal. Eighteen participants (64.3%) use the tool to save time on assignments. Behavioural beliefs regarding advantages centre on efficiency and performance. Twenty-six participants (92.9%) believe it helps them pass assignments with less effort. Nineteen participants (67.9%) expect improved grades. Sixteen participants (57.1%) value the tool for generating creative ideas. Ten participants (35.7%) believe it improves their understanding of course material. Behavioural beliefs regarding disadvantages focus on technical and academic risks. Twenty-two participants (78.6%) identify system crashes as a disadvantage. Eighteen participants (64.3%) cite the risk of plagiarism. Up to eleven participants (39.3%) find the answers to be generic or unhelpful.

Social influence plays a complex role in approval and normative beliefs. Nineteen participants (67.9%) seek approval from university professors. Eleven participants (39.3%) seek approval from classmates. Normative beliefs show a conflict between different social groups. Twenty-three participants (82.1%) believe their classmates approve of the use. Sixteen participants (57.1%) believe their parents approve. Eleven participants (39.3%) perceive approval from university professors. Nine participants (32.1%) perceive approval from employers. Conversely, twenty participants (71.4%) believe university professors disapprove of the use. Twelve participants (42.9%) believe employers disapprove.

Control beliefs indicate various facilitators and barriers to ethical use. Twenty-five participants (89.3%) believe supportive professors make ethical use easier. Seventeen participants (60.7%) identify clear regulations as a facilitating factor. Twelve participants (42.9%) believe that learning prompts and guidance for ethical use makes the process easier. Nine participants (32.1%) find relevant and useful assignments to be a facilitator. Several factors make ethical use difficult for participants. Twenty-two participants (78.6%) cite time constraints as a barrier. Nineteen participants (67.9%) find unrealistic or challenging assignments difficult. Sixteen participants (57.1%) identify excessive assignments as a barrier. Fourteen participants (50%) report difficulty due to ambiguity about ethical versus unethical use. Eleven participants (39.3%) find it difficult to maintain ethical use for assignments with a low-grade contribution.

Discussion

The findings of this study identify specific advantages and concerns regarding ChatGPT use in university settings. These results align with the Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis proposed by Farrokhnia et al. (2023). Participants recognize several strengths including increased efficiency and improved grades. They also value enhanced creativity and better learning outcomes, which aligns with previous research suggesting that LLMs might have positive impact on creativity (Toma & Yáñez-Pérez, 2024). These behavioural beliefs suggest that pre-service teachers view LLMs as a functional asset for their professional development. However, participants also acknowledge significant weaknesses. Concerns remain regarding LLM reliability and the risk of plagiarism. Participants specifically question the quality of generated content and the potential for generic responses.

The data emphasize the critical role of professors in guiding ethical practices. Participants suggest that structured guidance and prompt engineering support from instructors would improve their usage. This finding highlights a gap between the desire for faculty approval and the perceived disapproval from the same group. While students seek to meet academic standards, the lack of clear institutional frameworks creates ambiguity. Providing specific guidance for ethical use would likely bridge this gap and reduce the difficulty associated with navigating new technology.

These results provide empirical evidence for the benefits of ChatGPT discussed by Cooper (2023)

and Dwivedi et al. (2023). The findings also confirm concerns about potential unethical use noted by Kasneci et al. (2023). University students demonstrate an active awareness of these ethical challenges and the technical limitations of LLMs. This study identifies key beliefs that educators can target in future interventions to promote ethical engagement. Specifically, addressing control beliefs such as time constraints and assignment design could reduce the pressure to use AI unethically.

The practical implications of these findings inform the development of strategies to foster an ethical learning culture. Future studies might design and implement interventions that support the integration of LLM tools like ChatGPT. Rather than focusing on misconduct, professional development programs should highlight how LLM can assist in achieving procurement goals ethically. These strategies should include clear regulations and supportive teaching practices to lower the barriers to responsible use. This approach ensures that pre-service teachers develop the necessary skills to use LLM in their future classrooms.

References

- Ajzen, I. (1991). The theory of planned behaviour. *Organizational Behaviour and Human Decision Processes*, 50(2), 179–211.
- Ajzen, I., & Kruglanski, A. W. (2019). Reasoned action in the service of goal pursuit. *Psychological Review*, 126(5), 774–786. <https://doi.org/10.1037/rev0000155>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education (8th edition)*. Routledge.
- Cooper, G. (2023). Examining Science Education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, 32(3), 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- Dwivedi, Y. K., ... Wright, R. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 1–15. <https://doi.org/10.1080/14703297.2023.2195846>
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behaviour: The reasoned action approach*. Psychology Press.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kikalishvili, S. (2023). Unlocking the potential of GPT-3 in education: Opportunities, limitations, and recommendations for effective integration. *Interactive Learning Environments*, 1–13. <https://doi.org/10.1080/10494820.2023.2220401>
- Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Woo Young Chun, & Sleeth-Keppler, D. (2002). A theory of goal systems. *Advances in Experimental Social Psychology* 34, 331–378. [https://doi.org/10.1016/s0065-2601\(02\)80008-9](https://doi.org/10.1016/s0065-2601(02)80008-9)
- Milano, S., McGrane, J. A., & Leonelli, S. (2023). Large language models challenge the future of higher education. *Nature Machine Intelligence*, 5(April), 333–334. <https://doi.org/10.1038/s42256-023-00644-2>
- Toma, R. B. & Yáñez-Pérez, I. (2024). Effects of ChatGPT use on undergraduate students’ creativity: a threat to creative thinking? *Discover Artificial Intelligence*, 4(74), 1-9. <https://doi.org/10.1007/s44163-024-00172-x>
- Toma, R. B., & Yáñez-Pérez, I. (2025). Factors influencing undergraduates’ ethical use of ChatGPT: A reasoned goal pursuit approach. *Interactive Learning Environments*, 33(6), 4000-4019. <https://doi.org/10.1080/10494820.2025.2457349>

Generative AI Training For Pre-Service Teachers: Impacts And Adoption Patterns

*Helcio Soares Padilha*¹, *Victor López Simó*² and *Agostinho Serrano de Andrade*³

¹Post-Graduate Program in Science and Mathematics Teaching, Universidade Luterana do Brasil (ULBRA), Brazil

²Department of Mathematical Didactics and Experimental Sciences, Universitat Autònoma de Barcelona (UAB), Spain

³Post-Graduate Program in Science and Mathematics Teaching, Universidade de Caxias do Sul (UCS), Brazil

This study investigates the impact of targeted training in the use of Generative Artificial Intelligence (GAI) among pre-service Science and Mathematics teachers. Grounded in Rogers' Diffusion of Innovations Theory (DIT), the research examines how structured training experiences influence adoption patterns, pedagogical perceptions, and ethical awareness related to GAI use in educational contexts. The study adopts a quantitative approach and involved 183 pre-service teachers enrolled in initial teacher education programs. Participants took part in practical workshops designed to explore the pedagogical use of GAI tools, particularly ChatGPT, for lesson planning, problem-solving activities, and reflective discussions on ethical implications. Data were collected through pre- and post-workshop questionnaires and analysed using descriptive statistics and cluster analysis. The findings identified six distinct GAI adoption profiles, with Innovative Adopters and Moderate Adopters demonstrating higher levels of pedagogical engagement and greater confidence in integrating GAI into teaching practices. These groups increasingly used GAI to design interactive and personalized educational materials aligned with the goals of science education, such as fostering critical thinking and student engagement. Conversely, more conservative profiles showed resistance to adoption, highlighting the need for differentiated training strategies. Overall, the study underscores the relevance of targeted GAI training in initial teacher education, emphasizing its potential to enhance digital literacy, pedagogical innovation, and ethical awareness among future Science and Mathematics teachers.

Keywords: Generative AI, Teacher Training, Science Education

Introduction

Generative Artificial Intelligence (GAI) is rapidly transforming how teachers plan lessons, create materials, and engage students. Tools like ChatGPT now automate routine tasks, generate personalized content, and support interactive learning—opening new pedagogical possibilities for Science and Mathematics education. Yet this technological promise comes with significant challenges: ethical concerns about bias and misinformation, methodological questions about appropriate use, and pedagogical dilemmas about balancing AI assistance with student learning (García Sánchez, 2023; Sok & Heng, 2024).

The risks of uncritical GAI adoption are well documented. Students may develop technological dependency rather than critical thinking skills, and AI systems can perpetuate biases embedded in their training data (Montesano de Talavera et al., 2023; Avraamidou, 2024). These concerns are particularly acute in Science and Mathematics education, where inquiry, problem-solving, and reasoning form the core of learning. Preparing teachers to integrate GAI reflectively and ethically—rather than merely technically—has therefore become a central challenge for teacher education.

Initial teacher education offers a strategic intervention point. During this formative period, future

teachers develop beliefs, competencies, and dispositions toward technology that will shape their entire professional practice (Koehler & Mishra, 2008; San Andrés et al., 2022). Understanding how pre-service teachers adopt emerging technologies like GAI is therefore essential—not just for individual professional development, but for the broader project of pedagogical innovation in science education.

To examine these adoption processes, this study draws on Rogers' Diffusion of Innovations Theory (DIT), which frames adoption as a non-linear process shaped by individual innovativeness and perceptions of an innovation's relative advantage, complexity, and compatibility (Rogers, 2003). Recent research confirms that more innovative pre-service teachers integrate AI tools in pedagogically richer ways, while less innovative users adopt them cautiously or instrumentally (Uzumcu & Acilmis, 2024). Building on this foundation, the present study investigates how targeted GAI training influences adoption patterns, pedagogical perceptions, and ethical awareness among future Science and Mathematics teachers.

Objectives

The main aim of this study was to analyse the impact of targeted training in the use of GAI among pre-service Science and Mathematics teachers. Grounded in the DIT, the study sought to understand how structured training experiences influenced GAI adoption patterns, pedagogical perceptions, and ethical awareness in initial teacher education.

Specifically, the study aimed to:

1. Identify distinct profiles of GAI use among pre-service teachers, based on adoption patterns informed by the DIT;
2. Examine changes in participants' pedagogical and ethical perceptions of GAI before and after participation in practical workshops;
3. Analyse the relationship between GAI adoption profiles and the evolution of pedagogical and ethical perceptions following the training experience.

Methodology

We conducted a quantitative study to examine how targeted GAI training affects pre-service Science and Mathematics teachers. The sample comprised 183 pre-service teachers (aged 19–21) enrolled in initial teacher education programs.

Between April and May 2024, participants attended practical workshops designed to integrate GAI tools—particularly ChatGPT—into Science and Mathematics teaching. Workshop activities included lesson planning with AI assistance, applying Polya's method for problem-solving, and discussing ethical implications of AI use in education.

We collected data through pre- and post-workshop questionnaires. The pre-workshop instrument assessed participants' baseline GAI use patterns: frequency, diversity of applications, and initial perceptions of pedagogical potential. The post-workshop questionnaire measured changes in pedagogical and ethical perceptions, as well as confidence in using GAI for instruction.

We analysed the data using descriptive statistics and cluster analysis. The cluster analysis, informed by DIT categories, identified distinct GAI adoption profiles among participants. This approach allowed us to examine how adoption patterns relate to changes in pedagogical and ethical perceptions after training.

Results

Two main findings emerged from the analysis. First, pre-service teachers showed heterogeneous patterns of GAI adoption. Second, the training produced statistically significant changes across all perception dimensions ($p < .001$), though the magnitude varied considerably.

GAI Adoption Profiles And Openness To Innovation

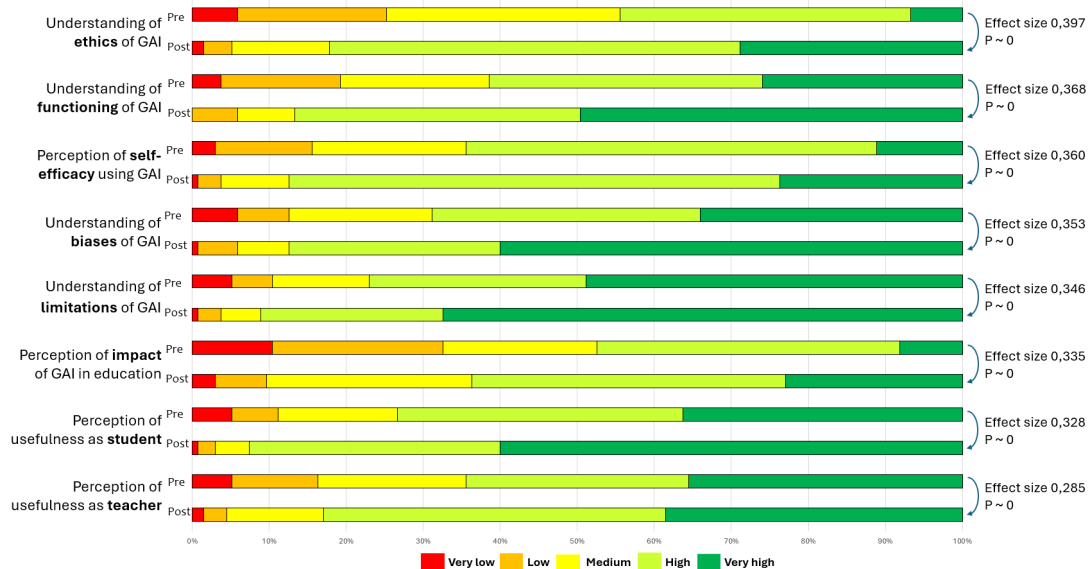
Cluster analysis identified six distinct profiles of GAI adoption: Innovative Adopters (13.97%), Moderate Adopters (33.52%), Conservative Users (22.35%), and three additional profiles (30.16%) showing varying degrees of engagement. These profiles reflect substantial variability in participants' experience, frequency and diversity of use, engagement with extensions and communities, and practices of updating knowledge about GAI tools. When interpreted through the DIT, the profiles can be broadly situated along a continuum ranging from more innovative adopters to more conservative users.

Notably, Innovative Adopters and Moderate Adopters together comprised the majority of the sample (47.49%), suggesting a favourable context for GAI adoption—likely reflecting participants' academic environment and digital familiarity. However, the substantial presence of Conservative Users (22.35%) demonstrates that openness to innovation varies even within digitally oriented cohorts. This heterogeneity validates DIT as an analytical lens for understanding GAI uptake in teacher education.

Differential Effects Of The Training On GAI-Related Perceptions

Wilcoxon signed-rank tests confirmed statistically significant pre-post differences across all perception variables ($p < .001$). As Figure 1 shows, the training consistently improved participants' GAI perceptions, but effect sizes varied substantially across dimensions.

Figure 1. Pre- and post-workshop changes in pre-service teachers' perceptions of Generative Artificial Intelligence.



The largest effect sizes emerged for ethical awareness ($r = 0.397$), understanding of GAI functioning ($r = 0.368$), and self-efficacy ($r = 0.360$). These results indicate that the training most effectively strengthened participants' conceptual clarity about how GAI operates, their confidence in using such tools, and their sensitivity to ethical issues including biases, limitations, and responsible use. In DIT terms, these gains reflect reduced perceived complexity and increased perceived relative advantage—two attributes that facilitate innovation adoption.

Moderate effect sizes characterized perceptions of biases, limitations, and GAI's broader impact on knowledge, learning, and problem-solving. These findings suggest participants moved beyond

purely instrumental views of GAI toward a more critical, reflective stance—increasingly recognizing its epistemic and ethical implications for educational contexts.

In contrast, perceived usefulness for future teaching showed smaller gains ($r = 0.285$), though still statistically significant. This pattern reveals a critical gap: technical understanding and ethical awareness do not automatically translate into a professional vision of GAI as an integral teaching tool. The distinction between intention and anticipated practice suggests that deeper pedagogical integration may require sustained, practice-based training experiences.

Discussion

This study examined how targeted GAI training affects pre-service teachers' adoption patterns and perceptions, using Rogers' DIT as an analytical framework. The core finding is clear: structured training positively influences how future teachers understand and evaluate GAI. Yet the results also reveal an important boundary—openness to innovation does not automatically translate into pedagogical consolidation.

Through a DIT lens, the strongest changes occurred in ethical awareness ($r = 0.397$), understanding of GAI functioning ($r = 0.368$), and self-efficacy ($r = 0.360$). These gains map directly onto two attributes Rogers (2003) identifies as central to innovation adoption: reduced perceived complexity and increased perceived relative advantage. For GAI—where uncertainty about system functioning and ethical risks often inhibits adoption—such improvements are particularly consequential for fostering informed, responsible engagement.

Moderate effects in awareness of biases, limitations, and GAI's broader impact on knowledge and learning indicate an important conceptual shift. Participants moved from viewing GAI as merely a productivity tool toward recognizing its epistemic and ethical implications—a crucial step in adopting complex educational technologies thoughtfully rather than instrumentally.

In contrast, perceived usefulness of GAI for future teaching practice exhibited comparatively smaller gains (effect size = 0.285), despite remaining statistically significant. This finding highlights a crucial distinction between increased confidence and understanding, on the one hand, and the integration of GAI into teachers' professional identity and anticipated classroom practices, on the other. Consistent with previous research, this result suggests that intentions to use GAI do not necessarily translate into sustained pedagogical practices, particularly when training experiences are short-term or not embedded in authentic teaching contexts.

The identification of heterogeneous adoption profiles further reinforces the relevance of DIT for interpreting these findings. The coexistence of innovative, early adopter, and more conservative profiles within the same cohort suggests that openness to GAI is uneven, even among digitally oriented pre-service teachers. This heterogeneity implies that uniform training approaches may be insufficient to address diverse levels of readiness, motivation, and concern. Differentiated training strategies that consider distinct adoption profiles may therefore be more effective in supporting equitable and sustained integration of GAI in teacher education.

Finally, some limitations of this study should be acknowledged. The reliance on self-reported data captures perceptions and intentions rather than observed teaching practices, and the relatively short duration of the intervention limits inferences about long-term adoption. These limitations are consistent with exploratory studies on emerging technologies and point to the need for longitudinal and practice-based research. Future studies could investigate how repeated and contextually embedded training experiences influence the consolidation of GAI use in classroom practice over time.

Final Considerations

This study demonstrates that targeted training can meaningfully shape how pre-service teachers perceive and approach GAI. The training enhanced ethical awareness, deepened understanding of AI functioning, and strengthened self-efficacy—all prerequisites for thoughtful technology integration. Yet the more modest gains in perceived usefulness for teaching reveal a crucial insight: knowledge and positive attitudes alone do not guarantee pedagogical transformation.

The results suggest that training experiences are particularly effective in addressing dimensions related to ethical awareness, understanding of GAI functioning, and self-efficacy, which are closely associated with reducing perceived complexity and enhancing the relative advantage of the innovation. At the same time, the more modest gains observed in perceived usefulness for future teaching practice highlight that adoption processes are gradual and do not automatically result from increased technical knowledge or positive attitudes alone.

By identifying heterogeneous adoption profiles among pre-service teachers, this study reinforces the relevance of the DIT for understanding how GAI is appropriated in initial teacher education. The coexistence of innovative and more conservative profiles underscores the need for differentiated and sustained training strategies that account for varying levels of readiness, experience, and concern.

This study contributes empirical evidence on how targeted training shapes GAI perceptions and adoption patterns among future teachers. The practical implication is clear: teacher education programs should incorporate structured GAI training that addresses not just technical skills, but ethical awareness and pedagogical integration. Future research should employ longitudinal designs to track how these early experiences translate into sustained classroom practice.

Acknowledgement

We acknowledge the financial support receive by CAPES and FAPERGS/RS (Process 25/2551-0000798-4)

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

- Avraamidou, L. (2024). Can we disrupt the momentum of the AI colonization of science education? *Journal of Research in Science Teaching*, 1(1), 1–5. <https://doi.org/10.1002/tea.21961>
- García Sánchez, O. V. (2023). Uso y percepción de ChatGPT en la educación superior [Use and Perception of ChatGPT in Higher Education]. *RITI Journal*, 11(23), 98–107. <https://doi.org/10.36825/RITI.11.23.009>
- Koehler, M. J., & Mishra, P. (2008). Introducing technological pedagogical content knowledge. In AACTE Committee on Innovation and Technology (Eds.), *Handbook of Technological Pedagogical Content Knowledge (TPCK) for Educators* (pp. 3–29). Routledge.
- Montesano de Talavera, M., Reis, M. L., Prieto, A., & Zorzal, R. (2023). Competencias digitales de los profesores de educación básica: Una mirada reciente desde una formación en línea [Digital Competencies of Primary Education Teachers: A Recent Look from Online Training]. *Publicaciones*, 53(1), 49–64. <https://doi.org/10.30827/publicaciones.v53i1.27985>
- Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.
- San Andrés, E. M., Rodríguez, M. C., Pazmiño, M. F., & Mero, K. M. (2022). Tecnologías Web 2.0 en el proceso de formación universitaria: Programa de capacitación para favorecer el conocimiento y habilidades de los docentes [Web 2.0 Technologies in University Education: Training Program to Enhance Teachers' Knowledge and Skills]. *Formación Universitaria*, 15(1), 127–134. <http://dx.doi.org/10.4067/S0718-50062022000100127>
- Sok, S., & Heng, K. (2024). Opportunities, challenges, and strategies for using ChatGPT in higher education: A literature review. *Journal of Digital Educational Technology*, 4(1), ep2401. <https://doi.org/10.30935/jdet/14027>

Uzumcu, O., & Acilmis, H. (2024). Do innovative teachers use AI-powered tools more interactively? A study in the context of diffusion of innovation theory. *Technology, Knowledge and Learning*, 29, 1109–1128. <https://doi.org/10.1007/s10758-023-09687-1>

Design-Based Research To Create A Collaborative Game To Foster Spatial Skills: The Role Of Teachers

Alain Sénécaïl, Eric Sanchez, Mireille Bétrancourt, Gaëlle Molinari and Fatou-Maty Diouf

TECFA-FPSE – University of Geneva, Switzerland

This research explores the role of teachers in the co-design process of a hybrid collaborative game aimed at enhancing spatial skills among students in primary school. Spatial competencies, such as visuospatial perspective-taking and frame-of-reference shifting, are an important part of learning in STEM subjects but often are not explicitly taught. Using a Design-Based Research (DBR) approach, the project integrates immersive virtual environments (VR) with tangible elements to create an equitable and engaging learning experience. Teachers are involved in the design and development phases, providing input that shaped gameplay mechanics to meet curriculum intentions and classroom realities. The game, which combines VR, digital maps, and physical boards, encourages students to solve spatial problems collaboratively while navigating a narrative-driven environment. Preliminary findings reveal how teacher engagement might overcome challenges of equitable engagement, adaptive scaffolding, and integration into the curriculum. The teachers pointed out that the need for visual cues, verbalization of spatial reasoning, and adaptive difficulty levels underlined the support for students with different spatial abilities. Other main orchestration strategies include balancing guidance and student autonomy and transitioning between digital and physical tasks smoothly. These recommendations really point to the importance of collaboration between researchers and teachers in the design of pedagogically relevant and feasible game-based learning environments. This will be further implemented in-class, studying how teachers adapt and scaffold learning within this hybrid environment. This research contributes to understanding how teacher expertise and DBR methodologies can enhance the design and integration of game-based tools for spatial skill development in primary education.)

Keywords: Game-Based Education; Spatial Cognition; Collaborative Learning

Context and Conceptual Network

Spatial Abilities At Stake

Spatial thinking is significant in science education. It enables students to mentally represent, manipulate, and reason about spatial relationships, which are essential for understanding key scientific concepts and solving problems in disciplines such as mathematics, earth sciences, physics, chemistry and engineering (Newcombe et al., 2023; Tversky, 2005, Uttal, Miller & Newcombe, 2013). Despite its importance, spatial thinking is rarely taught explicitly in primary education (Newcombe et al., 2023). However, research shows that these skills can be trained and effectively transferred to academic tasks (He et al., 2022). Digital technologies, particularly immersive virtual environments (VE), provide promising opportunities to support this development, especially for students with lower spatial abilities (Makransky & Petersen, 2023).

The present contribution is part of a four-year project aimed at designing and evaluating collaborative game-based learning environments to enhance spatial thinking in primary school students.

Two key spatial competencies are at the core of the project: visuospatial perspective-taking and frame-of-reference. Visuospatial perspective-taking is designed as “the ability to infer the visual or spatial properties of a scene in relation to another person or position” (Samuel et al., 2024).

Frame of Reference Shifting is the ability to switch between different spatial reference frames to process and interpret spatial information. There are three main types (Figure 1.). The egocentric frame is defined relative to one's own position (e.g., the dog is on my left). The allocentric frame is based on external landmarks and fixed coordinates (e.g., the dog is south of the school). Finally, the intrinsic frame is determined by the objects themselves (e.g., the school entrance is in front of the dog's tail).

Figure 1. Three Frames of Reference.



Hybrid Environment: Equity And Teacher's Orchestration

By integrating game-based interactions with immersive virtual environments, the project aims to foster spatial competencies by ensuring an equitable learning experience for every student. Indeed, there exist differences regarding gender, socioeconomic background, and prior exposure (Newcombe et al., 2023). The lack of consideration for inclusion in the design process of digital environments may aggravate these gaps (Gottschalk & Weise, 2023). In our project, we will adopt a hybrid approach that combines virtual reality and tangible elements to maximize inclusion by fostering the engagement of all students. More specifically, the learning environment will integrate virtual interactions with board-style physical elements. In this context, the challenge for teachers will be to effectively orchestrate the gaming sessions to ensure that all students can develop spatial skills in an equivalent manner.

In game-based learning, teachers must integrate games that align with academic objectives and learning goals (Arnab et al., 2012). Moreover, real-time feedback is crucial for tracking progress and enabling timely intervention, helping to maintain student engagement (Holstein et al., 2019). However, implementing these games effectively requires balancing student autonomy with structured guidance, which calls for dynamic orchestration strategies (Dillenbourg & Jermann, 2010). Ultimately, the success of game-based learning depends on teachers' involvement in both the design process and classroom integration (Sanchez, 2022). Therefore, teacher's involvement during the design process is an interesting avenue.

Teachers' Role In Co-Design Of The Game

The project uses a Design-Based Research (DBR) approach to design, develop and implement the collaborative game-based learning environment (Sandoval & Bell, 2004). DBR is a participatory, iterative methodology that involves two interconnected cycles of design, implementation, analysis, and refinement. This approach actively involves teachers at every stage, ensuring that their needs and perspectives shape the game design. As a result, the game is both research-informed and practice-oriented, increasing its relevance and likelihood of

integration into classroom practices (Sanchez & Monod-Ansaldi, 2015).

In particular, teachers are actively involved in two key moments: (a) the design and development phases, and (b) the implementation and evaluation phases. During the design and development phases, teachers collaborate with researchers and developers to create the game. In the implementation and evaluation phases, their role shifts to orchestrating students' actions and social interactions within the game environment.

This study examines the role of teachers in the design and development phases of the game. Specifically, it focuses on three key dimensions: how teachers draw on their teaching experience to inform the design process, how they anticipate orchestrating the game in the classroom, and how this projected orchestration influences the design and development of the game prototype. The aim is to understand how teachers' active involvement in these phases contributes to creating a game that is both pedagogically relevant and aligned with classroom realities.

In this context, we analyse the role of teachers in two aspects:

- Teachers as Co-Designers: game design sessions were conducted, bringing together four researchers, one developer, three primary school teachers, and one external expert. It shaped a narrative-driven collaborative game where students explore, manipulate, and co-construct spatial representations through multimodal interactions (VR headset, physical board, digital maps). The game mechanics were developed to support spatial frame of reference shifts and visuospatial perspective-taking, fostering equitable collaboration and knowledge-sharing among students (Olsen et al., 2021).
- Expected Teacher Orchestration in the Classroom: Teachers must strike a balance between student autonomy, scaffolding, and instructional direction to ensure equitable involvement (Koehler & Mishra, 2009). Focus is placed on teacher orchestration techniques, such as managing the shift between digital and physical activities, providing real-time student support, and implementing adaptive interventions.

Preliminary Insights

During first sessions, the first version of the game prototype was co-designed for the stimulation of collaboration, spatial exploration, and problem-solving. It integrated physical board, digital tablet, and VR headset.

Players take on the role of explorers in this intricately created universe, tasked with reestablishing harmony in the cursed middle-age city of Luméria. Through cooperative navigation and spatial problems in an interesting, dynamic environment, students acquire important spatial competencies.

Integrating Spatial Skills Into Game Mechanics

During the game design sessions, the teachers and the researchers embedded spatial reasoning tasks into the game, jointly ensuring that it will support perspective taking, frame of reference shifts, and cognitive mapping.

A key objective is to help students adopt multiple viewpoints in a 3D environment. To do so, the game follows a progressive difficulty structure, starting with egocentric tasks (e.g., "describe what you see") before moving to allocentric tasks (e.g., "navigate using a landmark"). Teachers agreed on the value of visual clues such as maps, icons, and directional indicators to support frame transitions and to orient the students and facilitate the interpretation of spatial information

across perspectives.

The hybrid game model of VR, physical board, and digital maps supports students in creating and updating their cognitive maps. Teachers emphasized the importance of students verbalizing their spatial reasoning; therefore, collaborative reflection questions were added to have them explain their navigation choices and strategies to their peers.

Equity In Spatial Skill Development

A preoccupation of the research is that pupils with poorer spatial skills might find it difficult to participate in activities that require intricate perspective changes. Teachers underlined how having to switch between frames of reference could cause these pupils to become disinterested or frustrated.

To handle this, the game design is planned to incorporate adaptive scaffolding mechanisms. To ensure that they gain spatial competencies at their own rate, students are given customized and guided navigation prompts.

Alignment With Curriculum & Classroom Feasibility

Making sure the game fits with STEM learning objectives without adding to teachers' workload in terms of class preparation and planning was another important consideration. Teachers emphasized that rather than requiring separate instructional time, spatial learning should be smoothly integrated into current curricula. By integrating spatial tasks into subjects like geography (map reading, landmark), mathematics (spatial transformations, coordinate systems), French (vocabulary, verbalization), biology, and physics (motion, reference frames), the game design aims to address this problem by fostering cross-disciplinary connections.

Challenges & Affordances In Classroom Implementation

Teachers projected several pedagogical constraints and agreed with researchers on opportunities in integrating game-based spatial learning into primary education (Tab. 1).

Table 1. Opportunities and Challenges in Orchestration.

	Affordances (Opportunities)	Challenges (Constraints)
Time Management	Encourages active, multimodal engagement, helping students bridge virtual and real-world spatial reasoning.	Managing time constraints while integrating the game into existing STEM lessons.
Equitable Collaboration	Collaborative problem-solving fosters spatial reasoning.	Ensuring that low spatial ability students do not disengage.
Guidance	Teachers can use scaffolding strategies (e.g. hints) to gradually fade support.	Feedback: Balancing teacher guidance vs. student autonomy in open-ended navigation tasks.
Diversity of Media	Using hybrid tools (VR, digital maps, artifacts) ensures multiple entry points for learning.	Classroom logistics for handling VR equipment and transitions between digital/physical tasks.
Game Universe	Keep in mind the necessity of a narrative and game universe that are age-appropriate and engaging for primary school students.	

Preliminary co-design results reveal interesting tensions in integrating game-based virtual environments. Teachers highlight challenges in time management, participatory equity, and orchestration load, while their perceptions of students' spatial skills and engagement shape

curriculum integration. The next phase will examine classroom implementation, focusing on how teachers adapt, scaffold, and co-regulate learning in this hybrid environment.

Acknowledgement

The authors gratefully acknowledge the support of the DEEP Consortium (<https://deep-consortium.ch/en>) and the Jacobs Foundation (<https://jacobsfoundation.org>).

References

- Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In M. S. Khine & I. M. Saleh (Eds.), *New Science of Learning* (pp. 525–552). Springer.
- Downs, R. M., & Stea, D. (Eds.). (2017). *Image and environment: Cognitive mapping and spatial behaviour*. Transaction Publishers.
- Gottschalk, F., & Weise, C. (2023). Digital equity and inclusion in education: An overview of practice and policy in OECD countries. OECD Working Paper No. 299.
- He, C., Chrastil, E. R., & Hegarty, M. (2022). A new psychometric task measuring spatial perspective taking in ambulatory virtual reality. *Frontiers in Virtual Reality*, 3, 971502.
- Holstein, K., McLaren, B. M., Alevan, V. (2019). Designing for teacher–AI complementarity in classroom orchestration. *Proceedings of the 20th International Conference on Artificial Intelligence in Education (AIED)*, 157–171.
- Makransky, G., & Petersen, G. B. (2023). The Theory of Immersive Collaborative Learning (TICOL). *Educational Psychology Review*, 35(4), 103.
- Newcombe, N. S., Hegarty, M., & Uttal, D. (2023). Building a cognitive science of human variation: Individual differences in spatial navigation. *Topics in Cognitive Science*, 15, 6-14.
- Olsen, J. K., Rummel, N., & Alevan, V. (2021). Designing for the co-Orchestration of social transitions between individual, small-group, and whole-class learning in the classroom. *International Journal of Artificial Intelligence in Education*, 31, 24–56.
- Sanchez, E. (2022). *Le paradoxe du marionnettiste. Jeu et apprentissage*. Toulouse: Octares.
- Sanchez, E., & Monod-Ansaldi, R. (2015). Recherche collaborative orientée par la conception. Un paradigme méthodologique pour prendre en compte la complexité des situations d'enseignement-apprentissage. *Éducation & Didactique*, 9(2), 73-94.
- Sandoval, W. & Bell, P. (2004), *Design-Based Research Methods for Studying Learning in Context: Introduction*, *Educational Psychologist*, Vol. 39, No. 4: pages 199-201.
- Tversky, B. (2005). Visuospatial reasoning. In *Handbook of reasoning* (pp. 209-249). Cambridge University Press.
- Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics. *Current Directions in Psychological Science*, 22(5), 367–373.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835. <https://doi.org/10.1037/a0016127>

Designing Digital STEM Teaching-Learning Sequences By Primary Pre-Service Teachers

Nikos Kapelonis and Dimitris Stavrou
University of Crete, Greece

Digital technologies offer opportunities to enhance teaching by providing interactive and immersive learning experiences. The research on the connection between digital learning objects in learning management systems to implement teaching learning sequences (TLSs) is rather limited. To this end, Digital Learning Objects (DLOs) and Learning Activity Management Systems (LAMS) are useful digital tools. DLOs are defined as digital resources which correspond to learning activity types. Learning activity types are distinguished in three categories: a) Knowledge-building, b) Convergent Knowledge Expression and c) Divergent Knowledge Expression. In addition, using LAMS tools TLSs could be structured. Therefore, LAMS platforms and DLOs can play a central role in developing TLSs in STEM education. In the present paper we will present a study investigating the way primary pre-service teachers design digital STEM TLSs using the LAMS platform. In the study participated 17 pre-service teachers in third and fourth year of their bachelor's degree in primary education. The participants were randomly divided into six subgroups (five groups of three and one group of two) and developed digital TLSs focused on the topic of Climate Change. For the analysis qualitative methods of content analysis were used based on the digital TLS developed by pre-service teachers in the LAMS platform. Our results provide evidence of the connection between the types of learning activities and the possibilities offered by the digital tools and indicate that using Digital Learning Objects enhance the Digital Learning Sequences in STEM education and pre-service teachers are capable of designing coherent, linear digital STEM sequences that prioritize content exploration and guided inquiry.

Keywords: Digital Learning, Digital Resources, e-Learning

Introduction

Digital technologies offer unique opportunities to enhance teaching by providing interactive, inquiry-based, and immersive learning experiences (Hsu, 2012). The use of digital technologies in education, especially after the COVID-19 pandemic, has increased considerably. Since teachers are responsible for designing and structuring lessons, it is crucial to effectively integrate Technology into them. The TPACK framework (Mishra & Koehler, 2006) supports the effective integration of technology into teaching. Effective lesson planning is essential in science education, as it significantly enhances teaching quality and improves student outcomes (Großmann et al., 2024). A key approach integrating digital tools in education involves the development and use of Digital Learning Objects (DLOs)—digital resources that facilitate specific learning objectives and promote student engagement (Kay & Knaack, 2008). Based on the above, developing Teaching-Learning Sequences (TLSs) (Psillos & Kariotoglou, 2016), provide coherent and inquiry-driven frameworks for addressing complex STEM topics like climate change.

Learning Management Systems (LMS) are platforms that provide a range of digital tools, like chat, forum, document repository etc within a course. One of the many options of LMS platforms is the Learning Activity Management System (LAMS) platform. While LAMS provides tools similar to standard LMS platforms, it uniquely facilitates the structuring of digital TLSs by sequencing learning activities. DLOs can be part of the content of the various platform's tools.

Since LAMS platform support DLOs positively (Levy et al., 2009), pre-service teachers (PSTs), acting as TLS designers, can play a central role in STEM education. A group of LAMS tools form learning blocks. The learning blocks correspond to learning activity types. Group of blocks may form larger learning blocks, supporting learning activities in an inquiry-based framework. According to the categorization of Harris et al. (2009), learning activity types may belong to the following categories:

- a) Knowledge-building (e.g., Read text),
- b) Convergent Knowledge Expression (e.g., answer questions), and
- c) Divergent Knowledge Expression (e.g., do a presentation).

In our study, the development of TLSs was structured in the LAMS platform. The platform provided tools categorized as:

- a) Flow (Gate, Grouping etc.),
- b) Evaluation tools (Assessment, like Multiple Choice and Reflective, like Question & Answer),
- c) Collaboration (Chat, Forum etc.), and
- d) Content (Noticeboard etc.).

LAMS tools can be used to organize TLSs according to the flow of the designed sequence. For example, the gate tool can be used when the designer wants participants to reach a specific point in the TLS before all the participants move forward. In the same line, the grouping tool can be used in order to create groups dealing with different activities.

This study examines the following research question: How do primary pre-service teachers design digital STEM Teaching-Learning Sequences? In particular we investigate the learning activity type integrated by pre-service teachers in the design of a digital TLS about climate change in the LAMS environment.

Research Context And Implementation

The research was conducted under the European Union Erasmus+ project STEM-DIGITALIS. The aim of the project was to develop digital tools and learning objects for advanced STEM topics, like climate change, renewable energy (RE) sources, etc. Within this framework, four (4) Digital Learning Objects (DLO) were developed to facilitate inquiry-based learning and provide immersive, interactive experiences for students. The four DLOs included two interactive graph representations, an interactive video, a computer game, and an augmented reality (AR) application. These TLSs were structured according to the Model of Educational Reconstruction (Duit et al., 2012), which emphasizes the integration of scientific content, student perspectives, and effective teaching strategies. Below it is a short description of the three units:

The first TLS focuses on the correlation mechanism between global temperature and CO₂ concentration. It is based on a serious game, designed with Unity 3D platform, where participants navigate through the last four decades, searching for evidence on the correlation between CO₂ concentration and temperature. The game is assisted with hotspots and players access additional information, such as interactive graph representations, an interactive video, answer multiple-choice questions and explore online dashboards for visualizing trends and drawing conclusions. The game is designed to foster critical thinking and data analysis skills, enabling participants to explore the correlation between rising global temperature and increasing CO₂ concentration.

The second TLS shifts focus to the anthropogenic activities that contribute to climate change and

the mitigation measures individuals can take. This TLS utilizes interactive maps and a mobile application to engage participants. The maps display data on CO₂ emissions per country and per activity (e.g., transportation, energy use), providing a global perspective on CO₂ emissions. The application allows the users to calculate their personal CO₂ footprint based on their daily activities. Users have choices on two daily personal activities: transportation and electrical device use. The results are visualized through daily pie charts, offering an understanding of their personal environmental impact, encouraging reflection on mitigation strategies.

The third TLS explores the socio-scientific dimensions of RE, particularly the challenges and opportunities associated with transitioning to climate neutrality. This TLS employs an augmented reality (AR) application and a “treasure hunt” activity to engage participants in role-playing and collaborative problem-solving. Participants are divided into groups and assigned roles (e.g., politicians, activists, researchers, etc.) to represent different stakeholder views in the debate. Using the AR application, participants scan images to access information, solve riddles, and examine the perspectives of stakeholders. This activity enhances critical thinking, negotiation skills, and the understanding of the complex social implications of RE policies.

The TLSs were introduced in a seminar entitled “Digital Tools in Science Education”, offered to 17 PSTs, in third and fourth year of their bachelors in primary education. The participants had attended one compulsory foundational course on the concepts of physics and one on the principles of science education. The participants were randomly divided into six subgroups (five groups of three and one group of two) to encourage collaboration and peer learning.

The seminar was structured into three phases:

- **Introductory Phase:** It introduced participants to various digital tools used in STEM education, including interactive videos, graphs, infographics, games, mobile applications, AR, and digital labs. The participants focused on the attributes, benefits and drawbacks of each tool. The LAMS platform was presented, highlighting its features for designing and managing TLSs.
- **Familiarization Phase:** Participants engaged with the three pre-designed TLSs on climate change (CC), utilizing different DLOs. They experienced TLSs as learners and reflected on their potential application in future teaching practices.
- **Design Phase:** Participants designed their own digital TLSs using the introduced DLOs, enriching them with other activities and tools. Each PST group developed a unique TLS on a CC-related topic, applying the knowledge and skills acquired during the seminar.

Data Collection & Analysis

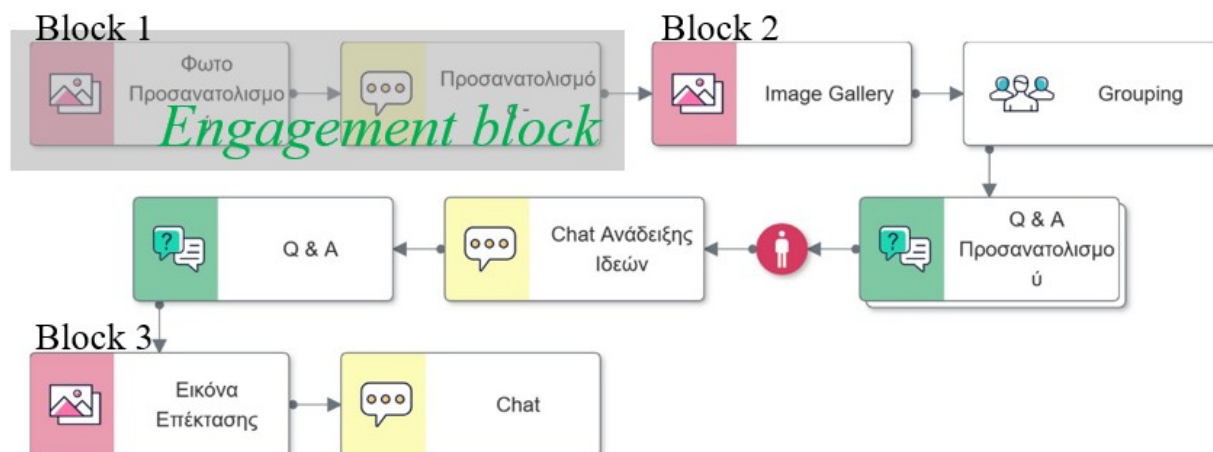
The analysis focused on how the PSTs integrated the DLOs and other types of learning activities within each digital TLS.

The topics of the PST-designed TLSs included:

1. Climate Change and Electric Transportation,
2. Human Factors Affecting Climate Change,
3. Transportation and Climate Change,
4. The CO₂ footprint of Food,
5. Human Activities and Sustainable Development, and
6. The Personal CO₂ Transportation Footprint.

A variety of technologies were used to support the digital TLSs. Scaffolding activities created larger blocks guiding students through the inquiry process in developing TPACK and Technology-assisted, inquiry-based TLSs. A sample TLS is shown on Image 1. The first two blocks formed the engagement activities of the TLS (Image 1 – Engagement Block).

Image 1 Sample TLS (Group1 - Climate change and electric transportation).



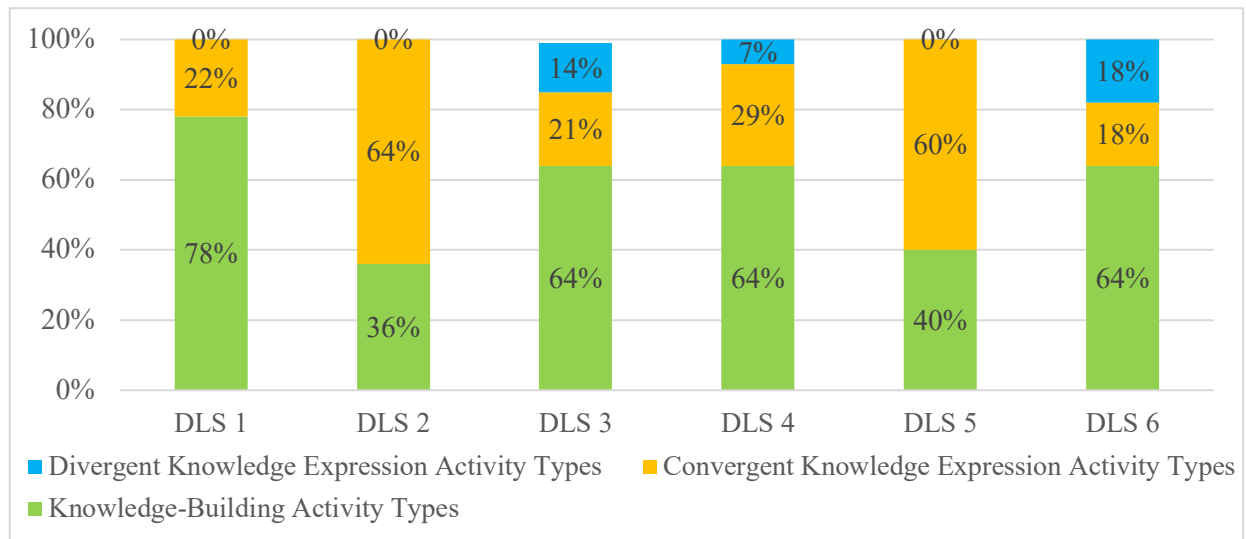
To analyse the data, we used qualitative method of content analysis (Mayring, 2015). As Harris et al. (2009) suggest, the possible categories of learning activity types follow the schema shown in Table 1:

Table 1 Data Analysis Categories

CATEGORIES	CRITERIA	Example
Knowledge-building	Provide information and promote the understanding for the students for the necessary content-related information-building	Read text, view presentation, view images, etc.
Convergent Knowledge Expression	Handle types of tools that “structure representations of prior knowledge building”.	Answer questions, take a test, etc.
Divergent Knowledge Expression	“Extend the students” content-related understanding by communicating the content to a wider audience.	Create an illustrated map, do a presentation, etc.

Results

There is diversity on the learning activity type categories. Every sequence used different LAMS tools/activities. The total number of the LAMS tools used in the six TLSs is seventy-four (74). The percentage distribution of the three learning activity type categories across the six digital learning sequences, based on the schema presented above, is shown in Figure 1. Summative results indicate that all the TLSs used activity-type tools for Knowledge-building and Convergent Knowledge Expression. Knowledge-Building activity types prevail among the three categories examined and the Convergent activity types come second out of all. The three of six sequences integrate all three learning activity type categories. Four sequences use Knowledge-building activities in a rather high rate in contrast to the other two learning activity type categories.

Figure 1 Distribution of the four activity types in TLSs

As we delve deeper into each of the three categories, the distribution of learning activity types used with respect to Knowledge-building category is shown in Figure 2. A total of thirty-eight (38) LAMS tools were identified within this category. Debate and Simulation activities featured prominently. In contrast, the Complete Charts/Tables, Research and Devise a story activities barely register.

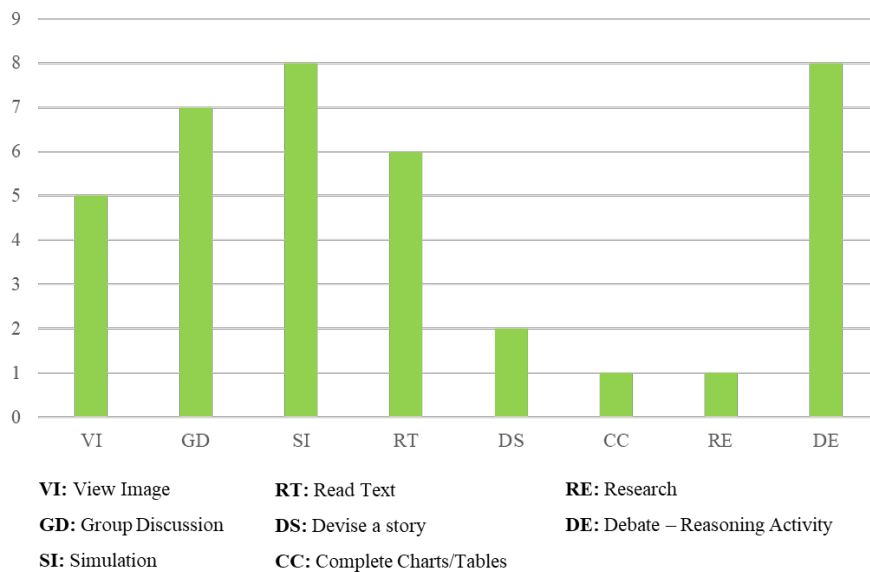
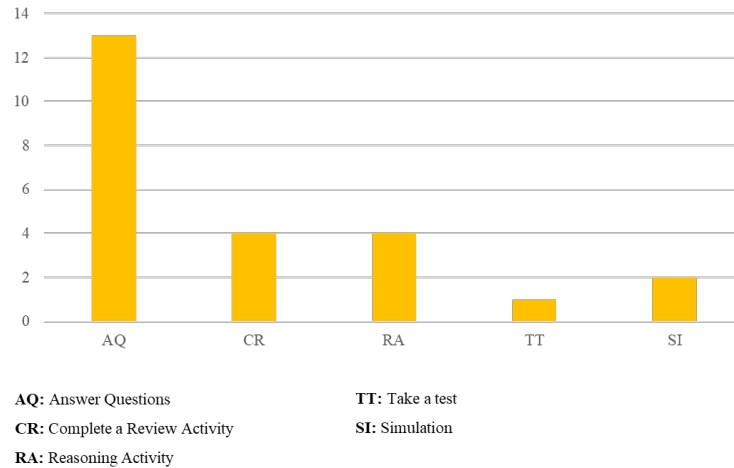
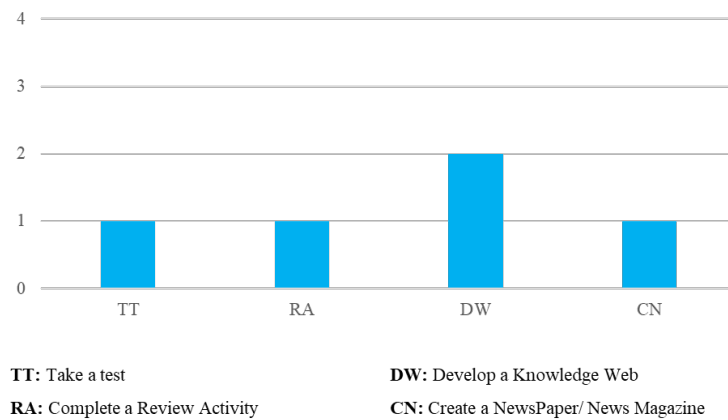
Figure 2 Knowledge-building activity types of the six DLSs.

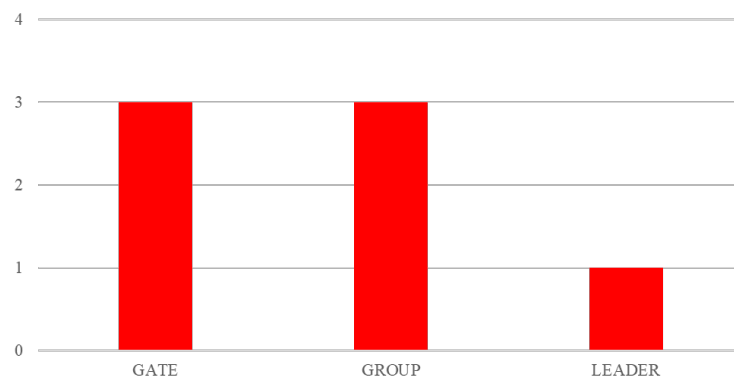
Figure 3 shows the distribution of Convergent Knowledge-building activity types among the six TLSs. This category comprises twenty-four (24) LAMS tools. The distribution reveals a pronounced dominance of the Answer Questions category. The remaining categories constitute the remaining proportion of the total.

Figure 3 Convergent Knowledge building activity types of the six DLSSs.

As shown in Figure 4, the Divergent Knowledge-building representation is the least populated category. Specifically, only five (5) LAMS tools were identified across three out of the six TLSSs.

Figure 4 Divergent Knowledge building activity types of the six DLSSs.

The flow tools category composes an important role at the sequences. Although the flow tools have rather technical attributes, than pedagogical or content specific attributes, they can contribute on a well-structured and well-designed digital sequence. These tools can split students into groups, create different learning paths for the students and other kind of structural options. In our findings, seven (7) LAMS flow tools used in three of six sequences.

Figure 5 Flow tools used into the six DLSSs.

Conclusion/Discussion

Our research examined how primary pre-service teachers (PSTs) design digital STEM Teaching-

Learning Sequences (TLSs) within the LAMS platform, with a specific focus on the learning activity types they integrate into their digital sequences. The analysis of the six PST-designed TLSs reveals the design priorities of digital sequences.

The data confirms a preference for Knowledge-building and Convergent Knowledge Expression learning activity types, within the LAMS environment. The high frequency of structured activities like debates, simulations, and answering questions indicates that PSTs effectively leverage digital tools to deliver content and scaffold understanding through structured, guided inquiry. These findings highlight the PSTs' preference in choosing digital tools that correspond with specific objectives regarding content exploration and formative evaluation.

While PSTs' digital designs show familiarity in activities belonging in Knowledge-building and Convergent Knowledge Expression learning activity types, there is a notable lack of use Divergent Knowledge Expression activity types. The tools intended for use as activities in Divergent Knowledge Expression category suggest a more inspired and open-inquiry driven lesson structure. Therefore, the PSTs' digital designs show a tendency toward the design of clearly guided inquiry driven digital sequences.

Moreover, the limited use of flow tools (like gate, grouping and branching) indicates a strong preference among PSTs for linear digital learning sequences. In addition, Digital Learning Objects were only used in Knowledge-building and Convergent Knowledge Expression learning activity types. The use of DLOs was limited to those characterized by simplicity with direct functional value, like a mobile application for calculating personal carbon dioxide footprint. As a conclusion, a deeper understanding of digital tools' attributes and properties equips PSTs with the familiarity needed to utilize them more skilfully and flexibly.

A limitation of our research is its focus on the design phase by the PSTs. The effectiveness of the sequences in actual settings, in-class or remotely, is to be evaluated. Future research should therefore implement these sequences in real teaching contexts to assess their impact on student learning outcomes and engagement.

In summary, our results indicate that using Digital Learning Objects enhance the Digital Learning Sequences in STEM education and pre-service teachers are capable of designing coherent, digital STEM sequences that prioritize content exploration and guided inquiry.

Acknowledgement

This work was supported by the Erasmus+ Programme STEM-DIGITALIS (<https://stemdigitalis-project.eu/>), under grant agreement No 2020-1-EL01-KA226-HE-094691

References

- Duit, R., Gropengießer, H., Kattmann, U., Komorek, M., & Parchmann, I. (2012). The Model of Educational Reconstruction – a Framework for Improving Teaching and Learning Science. In *Science Education Research and Practice in Europe* (pp. 13–37). SensePublishers. https://doi.org/10.1007/978-94-6091-900-8_2
- Großmann, L., Koberstein-Schwarz, M., Scholl, D., Krüger, D., & Meisert, A. (2024). Establishing common ground in empirical research on science teachers' lesson planning competence: a scoping review. In *Studies in Science Education*. Routledge. <https://doi.org/10.1080/03057267.2024.2415246>
- Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' Technological Pedagogical Content Knowledge and Learning Activity Types. *Journal of Research on Technology in Education*, 41(4), 393–416. <https://doi.org/10.1080/15391523.2009.10782536>
- Hsu, H. (2012). The Acceptance of Moodle: An Empirical Study Based on UTAUT. *Creative Education*, 03(08), 44–46. <https://doi.org/10.4236/ce.2012.38B010>

- Kay, R. H., & Knaack, L. (2008). A formative analysis of individual differences in the effectiveness of learning objects in secondary school. *Computers & Education, 51*(3), 1304–1320. <https://doi.org/10.1016/j.compedu.2008.01.001>
- Levy, P., Aiyegbayo, O., & Little, S. (2009). *Designing for inquiry-based learning with the Learning Activity Management System*. <https://doi.org/10.1111/j.1365-2729.2008.00309.x>
- Mayring, P. (2015). *Qualitative Content Analysis: Theoretical Background and Procedures* (pp. 365–380). https://doi.org/10.1007/978-94-017-9181-6_13
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education, 108*(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Psillos, D., & Kariotoglou, P. (2016). Theoretical Issues Related to Designing and Developing Teaching-Learning Sequences. In *Iterative Design of Teaching-Learning Sequences* (pp. 11–34). Springer Netherlands. https://doi.org/10.1007/978-94-007-7808-5_2

Digital Experimental Protocols And Literacy: Interventions For More Learner Support In Science Education

Philipp Behrens, Jürgen Menthe and Anke Meisert
University of Hildesheim, Germany

Adapting experimental protocols from professional scientific formats is a traditional but challenging writing task in science education. These formats often provide limited guidance and involve specific difficulties such as disjointed sections, impersonal language, and unclear expectations regarding the audience. Language barriers further complicate matters for second language learners and native speakers with developing literacy skills, making protocols a daunting prospect for students. Despite these challenges, however, experimental protocols provide valuable opportunities for students to develop systematic documentation and data analysis skills, thereby enhancing their understanding of science. To facilitate the writing process, a student-centred approach has been proposed that uses redesigned protocols to divide the text into boxes containing sentences, connected by arrows across the results and discussion sections. Targeted writing scaffolds also support sentence construction.

This pilot study investigates whether a fragmented digital protocol format and the use of procedural language support increase students' perceived cognitive load compared to more traditional formats. A 2×2 design was implemented in four-year 8 physics classes (n = 78) during a four-lesson unit on motion and inertia, with two documentation formats (traditional vs. fragmented) and two types of language support (classical vs. procedural). Cognitive load was measured after the intervention using a nine-item questionnaire. The results show no significant differences in perceived cognitive load between the different documentation formats and types of language support. Therefore, despite their structural novelty, neither the fragmented protocol structure nor the procedural writing scaffolds led to increased cognitive load. These findings suggest that structural and linguistic design modifications can be introduced without overloading students cognitively. Ongoing analyses will examine how these design features influence the quality and structure of students' experimental documentation.

Keywords: cognitive load, laboratory reports, scaffolding

Theoretical background

Writing Experimental Protocols As A Linguistic And Cognitive Challenge

Experimental protocols are a traditional, yet challenging, writing task in science education. In particular, the linear text format—into which different types of statements (e.g., observations/results, and interpretations) must be integrated—requires a high level of linguistic competence (Bayrak et al., 2015; Boubakri et al., 2015). In conventional lab reports, clearly separated sections (e.g., observation vs. interpretation) support structure, but the purely linear production of text often makes it difficult to construct explicit argumentative links between results and their interpretation. This linkage, however, is central to scientific knowledge generation (Köhler & Meisert, 2025). For learners of German as a second language, as well as for first-language learners with weaker academic language proficiency, it is particularly challenging to navigate complex subject content and demanding linguistic structures simultaneously (Leisen, 2010; Childs et al., 2015) and even proficient speakers face difficulties due cognitive and structural complexity (Müllner et al., 2024; Heinicke & Friedrich, 2023).

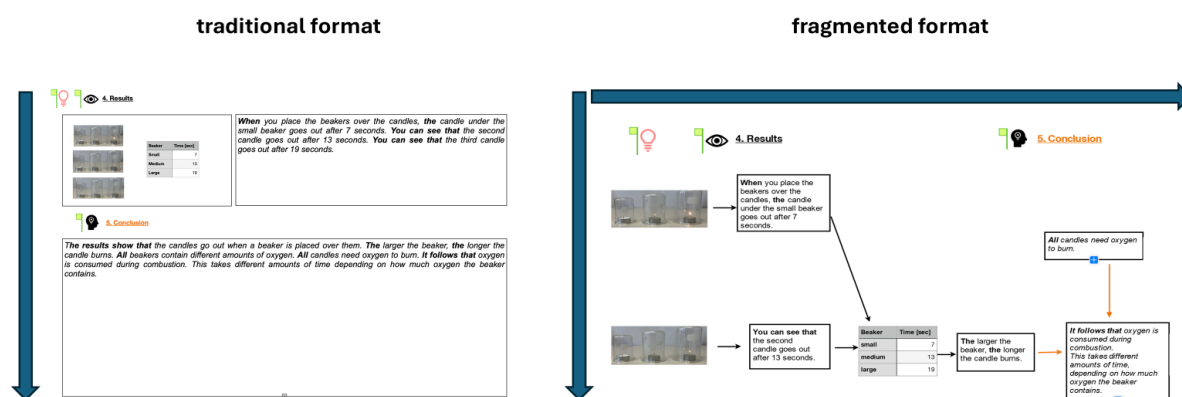
To meet these demands, support measures are needed that combine disciplinary language, visual, and structural elements (Groß & Reiners, 2012; Leisen, 2010, 2022; Schmölzer-Eibinger et al., 2018). Digital writing environments offer additional potential in this respect, as they reduce the constraints of strictly linear text production and allow for flexible revision, rearrangement, and

linking of text elements (Clark & Mayer, 2016; Heinicke & Friedrich, 2023).

Structural Scaffolding Through Text Fragmentation And Visual Linking

One promising approach is the fragmentation of experimental protocols into smaller, functionally defined text units that can be visually linked to each other using arrows or other symbolic connectors, for example. Rather than listing observations first and interpreting them in a separate section later, this design enables students to link results and interpretations directly and step by step (see figure 1).

Figure 1. Traditional linear protocol format (left) vs. fragmented protocol format (right).



The fragmented protocol design makes it possible to link results and interpretations directly and to visualize their relations by means of arrows. In contrast to the classical linear protocol—in which observations and interpretations are strictly separated into their own sections—relationships can thus be reconstructed in smaller steps and with greater transparency (Groß & Reiners, 2012). The increased density of logical and linguistic connections between observations, data, and interpretation also facilitates the construction of argumentative or descriptive relations (Stein & Stumpf, 2019), improves clarity, and creates flexible options for additions and corrections during the writing process (Clark & Mayer, 2016; Heinicke & Friedrich, 2023). The decomposition into individual components additionally directs attention to subaspects and linguistic markers such as procedural expressions (Feilke & Rezat, 2020). Control questions, action prompts, and procedural language patterns make the requirements of individual protocol segments explicit and support the planning of the writing process (Feilke & Rezat, 2020; Krabbe, 2023).

Linguistic Scaffolding: From Classical To Procedural Language Aids

Students need more than just structural scaffolding; they also need linguistic support to cope with the genre-specific requirements of experimental protocols. In science education, so-called classical language aids are widely used for this purpose. These include word fields, sentence starters, writing frames and partially pre-formulated text elements (Leisen, 2010). Their primary function is to facilitate text production and reduce linguistic barriers at the formulation stage. Unlike these approaches, procedural language aids are based on the concept of text procedures (Feilke & Rezat, 2020). As well as supporting formulation, they explicitly address the function and epistemic purpose of each text segment. This is typically achieved through structured prompts, control questions and procedure-specific linguistic patterns that guide students through the processes of planning, monitoring and justifying their writing.

Figure 2. Classical language aids according to Leisen (2010) – word fields. (left) vs. procedural language aids (Krabbe, 2023; Feilke & Rezat 2020; Feilke & Rezat, 2021; Krabbe, 2023) – structured prompts (left), control questions (right).

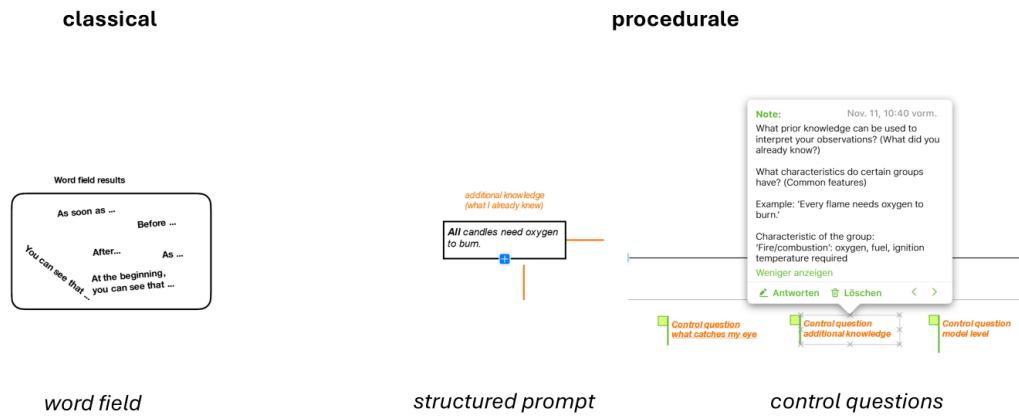


Figure 2 illustrates the conceptual difference between classical language aids, such as word fields and sentence starters, and procedural language aids, which are based on structured prompts and guiding questions (Krabbe, 2015, 2023; Feilke & Rezat, 2020). While the former primarily provide guidance on phrasing, the latter focus on the type of epistemic action to be performed in a given text segment (e.g. describing, justifying or interpreting). Procedural language aids therefore not only aim to reduce linguistic load, but also to foster a more explicit understanding of the writing task itself and the functional architecture of experimental protocols. In this sense, they can be understood as scaffolds that address the linguistic and epistemic dimensions of scientific writing.

The Tension Between Scaffolding And Cognitive Load

Although both fragmented protocol structures and procedural language aids offer significant support for students' scientific writing, they introduce new representational formats and increased orientation requirements. According to cognitive load theory, such design changes may initially increase extraneous or intrinsic cognitive load, particularly when learners encounter unfamiliar structures or multiple representational systems (Krieglstein et al., 2023). This creates a fundamental design tension: On the one hand, scaffolds should make relations, functions, and expectations more explicit, thereby reducing unnecessary cognitive demands. However, every additional structure, symbol system or prompt also constitutes an element that must be processed, understood and coordinated.

Therefore, the central design challenge is not to maximise scaffolding, but to calibrate support so that it reorganises and redistributes cognitive demands rather than simply adding new ones. Whether fragmented protocol structures and procedural language aids achieve this balance is therefore an empirical question rather than a purely theoretical one.

Research Questions

Against this background, the present study addresses the following research questions:

RQ1: Does a text-fragmented and visually structured experimental protocol lead to higher perceived cognitive load than a traditional linear protocol format?

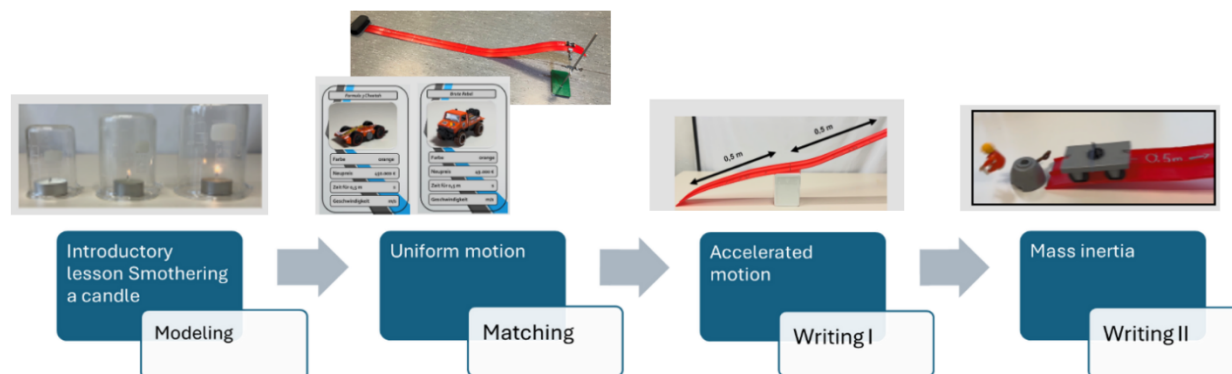
RQ2: Do procedural, function-oriented language aids lead to higher perceived cognitive load than classical formulation-oriented language aids?

Research method and design

This study used a 2×2 design comparing two documentation formats (traditional vs. fragmented) and two types of language scaffolding (classical vs. procedural). The design resulted in four experimental conditions with post-measuring of perceived cognitive load as dependent variable.

A four-lesson physics unit on motion and inertia was completed by four seventh-grade classes ($n=78$). Depending on their assigned condition, students progressed from guided modelling (Lesson 1) over scaffolded practice (Lesson 2) to unsupported use (Lessons 3–4).

Figure 3. 4-lesson sequence on motion and inertia, showing the materials (above) and the language aids with decreasing level of instruction and support (below) used during the lessons.



Cognitive load was measured using a 9-item Likert scale (1–5), which was adapted from Krieglstein et al. (2023). These items addressed students' perceived mental effort, structural clarity and ease of orientation when working with the digital experimental protocol (e.g. 'The explanations of the protocol sections were difficult to understand'). The questionnaire was administered as a post-test after the fourth lesson and referred to students' experiences with the documentation format and language aids used during the unit.

As the data violated normality assumptions and sample sizes were limited, non-parametric statistical procedures were employed for the primary analyses. Group differences in cognitive load were examined using Mann–Whitney U tests. Additionally, Welch's t-tests were employed as robust parametric comparisons tolerant of unequal variances and group sizes. Effect sizes are reported as r for non-parametric tests and as η^2 for Welch tests (Bortz, 2005).

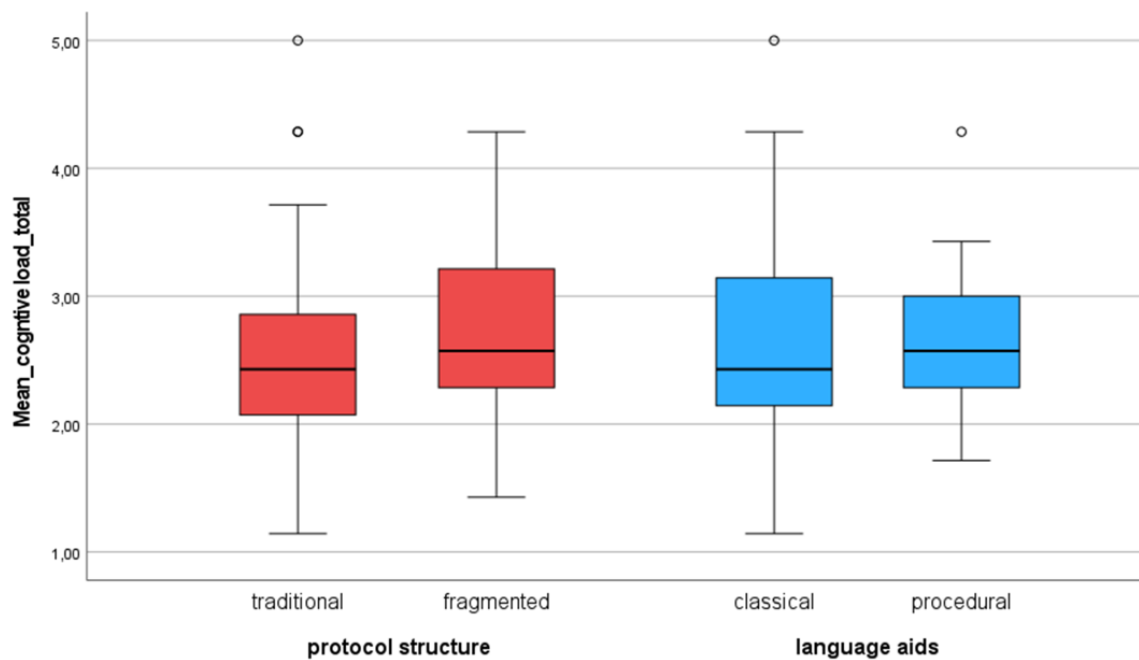
Results

The used cognitive load scale demonstrates good internal consistency (Cronbach's $\alpha = 0.816$).

Valid cognitive load data is available for group comparisons from $n = 78$ students (traditional format: $n = 43$; fragmented format: $n = 35$). Descriptively, mean cognitive load values are very similar across conditions. Students using the traditional format achieve a mean score of 2.53 (SD = 0.77; median = 2.33), whereas those using the fragmented format achieve a mean score of 2.56 (SD = 0.62; median = 2.56) (see figure 4, next page).

No statistically significant difference in cognitive load is found between the two protocol formats (Mann–Whitney $U = 672.50$, $Z = -0.805$, $p = .421$, $r \approx .09$). The corresponding Welch test also shows no significant effect (Welch's $F(1, 75.99) = 0.032$, $p = .858$, $\eta^2 < .01$). Tests of variance homogeneity is non-significant (all $p > .40$).

Figure 4. Paired box plot of the cognitive load of the protocol structure (left) and language aids (right).



Discussion

The present results show that the fragmented protocol format does not lead to a higher perceived cognitive load than traditional, linear documentation. This is a notable finding, given that the fragmented format introduces an unfamiliar structure to the documentation process and could therefore be expected to increase the demands of orientation and navigation (Krieglstein et al., 2023).

Two factors may help to explain this result. Firstly, the introduction of the fragmented format followed a gradual fading strategy with guided examples, which likely reduced the effects of initial disorientation. Secondly, the functional design of the format itself may compensate for its novelty. While the fragmented format introduces a new representational structure, it simultaneously reduces demands relating to strictly linear text production, linguistic planning and integrating different types of statements into a single, continuous text (Clark & Mayer, 2016; Heinicke & Friedrich, 2023). These reductions may offset the additional demands of navigating a segmented representation, resulting in overall cognitive load levels comparable to those of traditional formats. A central design intention of the fragmented format is to strengthen the argumentative link between observations, data, and interpretations. This addresses a well-known weakness of experimental protocols (Groß & Reiners, 2012) (Ref.?), whereby interpretative statements are often only weakly connected to the underlying observations and results. The visual structuring and explicit linking of these elements is intended to help students construct more coherent evidential relations. Ongoing qualitative analyses of the protocols will examine whether this functional promise is actually realised in students' documentation.

A similar pattern emerges when procedural writing scaffolds are used. While conventional language aids enable students to produce text quickly with minimal reflection, procedural scaffolds require them to engage more explicitly with the function and structure of each protocol section and its specific requirements. In principle, this could increase cognitive load. However, the absence of such an increase suggests that greater task transparency and clearer orientation

may stabilise or even reduce perceived cognitive demands rather than amplify them (Feilke & Rezat, 2020; Krabbe, 2023). In this sense, the fragmented structure and procedural scaffolds may free up cognitive resources for dealing with the task's content rather than surface-level text production.

Overall, the findings suggest that structural and linguistic design changes do not automatically result in a higher cognitive load. Instead, they appear to reorganise cognitive demands in a way that maintains overall mental effort at a comparable level (Krieglstein et al., 2023). From an instructional perspective, this implies that reworked documentation formats with a clearer structure and language can be incorporated into regular classroom practice without overwhelming students cognitively. However, the present data only address perceived overall cognitive load. Ongoing and future analyses will therefore examine students' protocol structures and their use of language scaffolds in more detail. Particular focus will be given to the separation of observations and interpretations, the linking of data and interpretations, and the development of data-based arguments.

References

- Bayrak, C., Hoffmann, L., & Ralle, B. (2015). Sprachliches und fachliches Lernen im Experimentalunterricht. *Der mathematische und naturwissenschaftliche Unterricht*, 68(3), 177–182.
- Boubakri, C., Krabbe, H., & Fischer, H. E. (2015). Sprachkompetenz im Versuchsprotokoll. Erste Ergebnisse aus der Pilotierung im Projekt SchriFT. In C. Maurer (Hrsg.), *Authentizität und Lernen – Das Fach in der Fachdidaktik* (S. 205–207). Universität Regensburg.
- Bortz, J. (2015). *Statistik für Human- und Sozialwissenschaftler* (6., überarb. Aufl.). Springer.
- Childs, P. E., Markic, S., & Ryan, M. (2015). The role of language in the teaching and learning of chemistry. In J. García-Martínez, E. Serrano-Torregrosa, & P. W. Atkins (Hrsg.), *Chemistry education: Best practices, opportunities and trends*. Wiley-VCH.
- Clark, R. C., & Mayer, R. E. (2016). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). Wiley. <https://doi.org/10.1002/9781119239086>
- Feilke, H., & Rezat, S. (2020). Textprozeduren. *Werkzeuge für Schreiben und Lesen. Praxis Deutsch*, 48(281), 4–13.
- Groß, K., & Reiners, C. S. (2012). Experimente alternativ dokumentieren. Ein Beitrag zur Möglichkeit der Differenzierung und Diagnose im Chemieunterricht. *Chemie konkret*, 19(1), 13–20.
- Heinicke, S., & Friedrich, S. (2023). Sammelhefte goes digital: Formen, Chancen und Herausforderungen einer digitalen Heftführung. *Naturwissenschaften im Unterricht Physik*, 34(195/196), 64–67.
- Krabbe, H. (2023). Schrittweise bessere Versuchsprotokolle verfassen. *Unterricht Physik*, 2023(195/196), 32–36.
- Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review*, 35(1), Article 9. <https://doi.org/10.1007/s10648-023-09738-0>
- Köhler, K.-H., & Meisert, A. (2026). Welche Erkenntnismethoden sind für den Biologieunterricht relevant? Cornelsen.
- Leisen, J. (2010). *Handbuch Sprachförderung im Fach: Sprachsensibler Fachunterricht in der Praxis*. Varus.
- Leisen, J. (2022). *Handbuch Fortbildung Sprachförderung im Fach: Sprachsensibler Fachunterricht in der Praxis*. Klett.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12. Aufl.). Beltz.
- Müllner, B., Heidinger, C., Hammerschmid, L., Scheuch, M., & Möller, A. (2024). Kognitive Prozesse beim Schreiben naturwissenschaftlicher Versuchsprotokolle: Eine explorative Studie zum sprachsensiblen Fachunterricht. *ZISU – Zeitschrift für interpretative Schul- und Unterrichtsforschung*, 13(1), Article 1. <https://www.budrich-journals.de/index.php/zisu/article/view/44288>
- Schmölzer-Eibinger, S., Dorner, M., Helten-Pacher, M.-R., & Langer, E. (2018). *Sprachförderung im Fachunterricht in sprachlich heterogenen Klassen*. Klett.
- Stein, S., & Stumpf, S. (2019). *Muster in Sprache und Kommunikation: Eine Einführung in Konzepte sprachlicher Vorgeformtheit (Grundlagen der Germanistik 63)*. Erich Schmidt.

Potential Of ChatGPT As A Pedagogical Tool In Biology Education: Enhancing Student Engagement And Critical Thinking

*Arianne Redigolo¹, Murillo Beloti², Julia Silva¹ Tainá Lopes¹, Victor Romanato¹,
Magda Pechliye¹*

¹Mackenzie Presbyterian University, Brazil

²Federal University of São Paulo, Brazil

The integration of Artificial Intelligence (AI) in education has transformed learning processes, with ChatGPT emerging as a valuable pedagogical tool. The objective of this study was to analyse how ChatGPT could enrich the educational experience, promoting engagement, reflection and development of critical skills among students. Conducted in a state school in Diadema, São Paulo, Brazil, the research involved a first-year high school class (N=20) engaging in structured activities using ChatGPT. Students formulated questions, interacted with the AI, and completed a questionnaire assessing their perceptions. Results indicate that initial AI usage was superficial, with students seeking direct answers rather than engaging in critical inquiry. However, as the activity progressed, students developed more complex and reflective questions, demonstrating deeper engagement with the subject matter. Post-activity questionnaire responses highlighted a shift in AI usage, with students recognizing ChatGPT as a research tool rather than a means for effortless answers. Despite ChatGPT's benefits, the study underscores the irreplaceable role of teachers in guiding students toward meaningful learning. Teacher mediation was identified as essential for fostering critical thinking, refining questioning skills, and contextualizing AI-generated content. The findings reinforce the importance of balancing AI integration with human interaction, advocating for pedagogical strategies that leverage technology while maintaining teacher-led inquiry-based learning. Ultimately, this research highlights AI's potential to enhance education while emphasizing the necessity of teacher involvement in knowledge construction.

Keywords: Artificial Intelligence, Critical Thinking, Science Education.

Introduction

The integration of technology in classrooms has gained significant space in recent years, with emphasis on the use of Artificial Intelligence (AI) tools. Among these, ChatGPT (Generative Pre-trained Transformer), a language model developed by OpenAI, emerges as an ally in the teaching and learning process. Launched in 2019, ChatGPT is based on a large language model technology and uses Reinforcement Learning with Human Feedback (RLHF) to interact naturally with users, offering personalized responses and immediate feedback on tasks (Cano; Venuti; Martinez, 2023). However, its use requires caution, since meaningful learning is based on the principle of constructing one's own and personal meaning in relation to knowledge and learning situations, which demands understanding, analysis and critical reflection (Coll; Sole, 2006; Gomes; Frank; Rocha, 2020).

In addition, there is a risk that students might become excessively dependent on the tool, compromising the development of critical and emotional skills. Božić (2023) highlights that AIs can reduce face-to-face interactions and negatively affect students' emotional and social development. There is also a need to verify the accuracy of the information provided by the AI, as errors can occur. That said, even with the advantages of AI, it's evident that the role of the teacher remains essential. In addition to teaching academic content, the teacher collaborates with

the development of students' social, emotional and cultural skills.

Barreira, Moura-Silva, and Gonçalves (2024) emphasize that the use of AI should be complementary, helping teachers to deepen learning processes and students' socio-emotional development. The teacher also plays a crucial role in teaching students to ask questions and develop opinions, promoting autonomy and critical thinking. According to Farace (2024), it is up to the teacher to establish the connection between the content provided by AI and the individual experiences of students, promoting more integrated learning.

From a pedagogical perspective, the use of generative AI tools can be situated within the framework of active methodologies mediated by technology, which emphasize student protagonism, inquiry, and intentional teacher mediation (Araújo et al., 2025). In this view, digital technologies do not inherently promote meaningful learning; rather, it is the teacher's critical mediation that organizes, guides, and gives pedagogical meaning to students' investigative processes. This understanding aligns with research on interest development, which highlights that learners' engagement can evolve from situational interest toward more enduring forms of individual interest when students are encouraged to explore self-selected questions within a shared learning context, involving both affective and cognitive dimensions (Hidi & Renninger, 2006). Together, these perspectives help frame inquiry-oriented activities supported by generative AI as learning environments in which meaningful engagement may emerge through the combination of personal interest, structured inquiry, and teacher mediation.

The main aim of this study was to explore how upper secondary students interact with a generative AI tool (ChatGPT) during an inquiry-oriented Biology activity, and how teacher mediation influences the depth and nature of this interaction. Rather than evaluating learning gains or content mastery, the study focused on understanding students' inquiry behaviours, including the formulation and refinement of questions, the use of AI-generated responses, and students' perceptions of the role of the teacher in this process.

More specifically, the study sought to examine whether students' initial use of ChatGPT, often characterized by direct answer-seeking, could evolve toward more reflective and exploratory forms of questioning when embedded in a mediated classroom activity.

Based on these aims, the study was guided by the following research questions:

1. How do students formulate and refine questions when interacting with ChatGPT during an inquiry-based Biology activity?
2. What changes can be observed in the nature of students' interactions with ChatGPT throughout the activity?
3. How do students perceive the role of teacher mediation in supporting meaningful use of generative AI in the learning process?

These questions were intentionally framed in an exploratory manner, consistent with the short duration of the intervention and its implementation in a single classroom session. Accordingly, the study does not aim to establish causal relationships or generalizable outcomes, but to generate situated insights that can inform future research and pedagogical practices involving the integration of generative AI in science education.

Methodology

Activity Design And Procedures

The study was conducted in a public state school located in Diadema, São Paulo, Brazil, with a first-year high school class composed of approximately 20 students, aged between 15 and 16

years old. This educational stage was selected because the topic Fungi is formally introduced in the biology curriculum at this level. The theme involves abstract biological concepts, such as structure, reproduction, ecological roles and symbiotic relationships, that often require a deeper conceptual understanding to support knowledge construction. Consequently, students require pedagogical support to actively relate new information to prior knowledge, establish conceptual and personal connections and engage in explanatory reasoning to construct scientifically grounded understandings. In this context, the use of Artificial Intelligence was included as a mediating digital tool to support and stimulate the access to scientific information, aligned with the teacher guidance in fostering reflective and meaningful learning processes.

Rather than testing predefined learning outcomes, the proposal sought to understand changes in student's questioning behaviour, use of ChatGPT, and perceptions of teacher mediation. The activities took place in the school's computer laboratory. It is important to highlight that, prior to the start of activities, a survey of the student's prior knowledge was conducted through a word cloud built based on what the students associated with the topic of fungi. This word cloud was organized on the blackboard. The central activity was structured in three main stages, with students organized in pairs for the first and second stages.

A) Stage 1: Question formulation (20 minutes)

Students were asked to formulate general and specific questions related to the theme Fungi, starting from the guiding question "What are fungi?". There were no explicit instructions on how to ask those questions at this point.

B) Stage 2: Guided interaction with ChatGPT (20 minutes)

Students used the ChatGPT to seek answers to their questions and deepen their knowledge. The teacher acted as a mediator, encouraging students and emphasizing investigation, exploration and reflection rather than the search for direct answers.

C) Stage 3: Questionnaire (individual)

At the end of the activity, the students completed a questionnaire specifically designed to identify their perceptions of: before using ChatGPT, their interactions with the AI tool and the difference before and after the activity, the teacher's role during the activity, and their own expectations regarding the future use of ChatGPT for learning purposes.

Data Collection And Analysis

Data were collected from two main sources, the questions formulated by students during ChatGPT interaction and the questionnaire. The collected data were analysed using a qualitative approach. Qualitative data consisted of the questions formulated by students during the second stage of interaction. These questions were initially read and grouped according to their thematic focus, based on the biological content, that were organized in three categories:

1. Structure and Reproduction of Fungi;
2. Ecological Interactions and Health Impact;
3. Applications and Uses of Fungi.

The categorization made it possible to identify how the students searched for different dimensions of the subject during their guided interaction with ChatGPT, navigating from structural and reproductive aspects to ecological roles and practical applications.

Quantitative data from the questionnaire were analysed descriptively using absolute frequencies.

Qualitative and quantitative results were interpreted together to identify consistencies between

student's reported perceptions and their observed questioning behaviour.

Results And Discussion

In the first stage of the activity (question formulation), students initiated their interaction with ChatGPT by asking the question "What are fungi?". As expected, the AI-generated responses were very similar, given the uniformity of the initial prompt. For this reason, these responses were not transcribed.

Prior to the intervention with teacher's mediation, students tended to use the tool superficially, obtaining quick answers or through "copy-and-paste" practices. This was noticed in the first stage, as students asked direct questions and didn't engage in follow up questioning, and directly address to research question 1, regarding how students initially formulate and refine questions when interacting with ChatGPT.

In the second stage (guided interaction with ChatGPT), students worked in pairs to formulate their own questions so they could deepen their understanding of fungi. The questions generated during this stage were grouped into three categories (Table 1).

Table 1. Questions asked to AI during the second stage of the research, later categorized.

Category	Questions
1. Structure and Reproduction of Fungi	What are hyphae and mycelium?
	How is each fungal species formed, and what characterizes them?
	What does sexual and asexual reproduction mean?
	Do fungi create genetically similar clones?
2. Ecological Interactions and Health Impact	Why are fungi so ecologically important?
	What are beneficial symbioses?
	Where can parasitic fungi be found, and how can they be identified?
	Why are some fungi pathogenic?
	Why are some diseases are caused by fungi?
3. Applications and Uses of Fungi	How are fungi used in medicine?
	How can fungi be used in food production?
	So, is the bread we eat made by a fungus?

After the intervention, student's demonstrated a more reflective and in-depth use of ChatGPT as an ally to the learning processes, characterized by the elaboration of more complex questions and detailed research, with a more sustained inquiry process, that suggests the increased curiosity, engagement and interest of the students.

In the first two categories above, although the questions were not very reflective and still focused on conceptual understanding, some went beyond basic concept definition. For example: Do fungi create genetically similar clones? (category 1) and Where can parasitic fungi be found, and how can they be identified? (category 2). These can indicate the effort to explore mechanisms, processes and applications of Fungi. In the third category, several questions were connected with student's everyday experiences, providing a more reflective perspective, such as food

consumption and health. Here we observe that as the guiding teacher encourages more reflective questioning, students will have the opportunity to learn how to ask new questions, and enhance their inquiry skills (Farace, 2024).

Four of the ten questions involved in the third stage (questionnaire) were selected for these discussions. Of the 20 students, 17 responded to the questionnaire. In the first question ("Had you ever used Chat GPT before this activity?"), 16 students reported prior use of ChatGPT, which was expected given the frequent exposure to digital technologies nowadays. Even though current generations were born in contact with technology, this familiarity does not necessarily indicate effective or critical use of those digital tools such as ChatGPT.

The second question addressed the main differences students noticed in using Chat GPT before and after this activity. Three representative literal responses were selected for analysis: "*Now I've stopped cheating and started using it as a research tool*"; "*I stopped getting answers and started asking*"; and "*Before, I only got the answer to the activities; today I was able to study through it to learn and not just copy what is written.*" These responses suggest that students began to understand the importance of moving beyond the mere reproduction of the contents toward seeking more connections, and constructing understanding, align with the meaningful learning processes proposed by Coll and Solé (2006).

In the third question ("Do you think that the teacher can be removed from this activity, so you (students) would do the activity alone?") of the 17 answers obtained, 16 considered the teacher's presence essential, while one student reported not understanding the question. Two highlighted literal answers were: "No, because the teacher is there to guide and teach" and "I think it is good for the teacher to advise the students in the activity". These answers indicate that the students realize the importance of the teacher's role as a mediator, who supports active learning, and not as the source of answers, which aligns with Barreira, Moura-Silva, and Gonçalves (2024), that emphasizes the teacher's role in the knowledge construction fostering critical questioning, although, unfortunately, teacher education programs often do not prepare them adequately to carry out this mediation role (Nóvoa, 2017).

In the fourth question ("Do you think your way of interacting with ChatGPT will change from now on?"), 13 students said that they believed their interaction would change, while four disagreed and said that wouldn't change. Statements such as "Use for research and not for cheating" and "I will dive deeper and create more questions instead of just getting simple answers." indicate that students understood the purpose of the activity as an opportunity to investigate and learn, rather than merely reproduce information. Some responses highlight the relevance of teacher mediation supporting high-quality questioning and knowledge construction such as: "More detailed and specific question", "I would change to try to understand the lessons, for example, ask questions about the doubts I have about the activities" and "Now I know how to ask the right way".

Although ChatGPT does not always respond with the adequate concept and knowledge, it is possible to reduce the chance of errors when formulating good questions. As for good questions, we consider that the more the teacher encourages the student to be reflective and critical about a given subject, the more these questions will be improved in terms of quality (Farace, 2024). These responses relate to research questions 2 and 3, particularly regarding changes in interaction patterns and student's perceptions of teacher mediation. In some way, ChatGPT's clarity and accessibility contributed to more meaningful learning, reinforcing its usefulness as a complementary tool. However, the study also underscores the importance of teacher mediation, as highlighted by Barreira, Moura-Silva, and Gonçalves (2024), who point out the irreplaceable

role of educators in guiding the learning process. These findings can be interpreted considering interest-driven learning perspectives, which emphasize the role of student's personal interests in fostering deeper engagement and inquiry (Hidi & Renninger, 2006).

This study presents limitations that should be considered. First, the intervention was conducted with a single class, involving a limited number of participants, which restricts the possibility of statistical generalization. Second, the activity took place during one class session, which constrains the observation of longer-term learning processes and the consolidation of inquiry practices supported by generative AI. In addition, the study was carried out in a specific public-school context, in Brazil, characterized by particular curricular, institutional, and temporal conditions. As a result, these findings should be understood as context-dependent and illustrative of possibilities rather than as representative of all educational settings or student populations.

Therefore, the results do not aim to establish any causal relationships or universal claims about the impact of ChatGPT on learning. Instead, they indicate tendencies and patterns of interaction that emerge when generative AI is integrated into an inquiry-oriented Biology activity with explicit teacher mediation. These limitations are consistent with the exploratory nature of the study and support a cautious interpretation of its scope and applicability.

Conclusion

This study reinforces that ChatGPT can be a powerful educational support tool, as long as used in a critical and mediated manner. Its capacity to personalize content and provide clear answers makes it a valuable resource for enriching the educational process. However, teacher mediation remains essential to ensure that interactions are meaningful and aligned with learning objectives. In this perspective, the teacher is understood as a mediator and collaborator, not a transmitter of knowledge. However, to assume this mediating role, it is necessary to disengage from the traditional conception of teaching and focus on the process of knowledge construction. This highlights the need for teachers to reflect on their conceptions of science, learning and education, which should be addressed in teacher training courses.

The integration of artificial intelligence in education also raises ethical and cultural challenges that demand continuous reflection and adaptation. Moreover, the new learning culture requires that teachers and students to develop critical thinking, soft and digital skills, fostering a more inclusive and meaningful education. Within the context of this study, the use of ChatGPT appears to have contributed not only to development of critical and investigative skills, but also to increase student engagement and interest, expanding the possibilities of integrating AI in various disciplinary contexts. This work emphasizes the importance of balancing technology and human interaction, pointing out ways for a more innovative and inclusive education.

From a methodological standpoint, reflective and critical reading and writing activities related to the topic of fungi could have been carried out prior to the use of the AI tool in this specific context. Such an approach might have provided students with conceptual and argumentative references that would enable them, under teacher mediation, to engage in a more critical, conscious, and problematizing dialogue with the artificial intelligence.

In this perspective, it's interesting to consider for further research these reflective and critical reading and writing practices as pedagogical actions that might precede the use of generative AI, that's because it highlights the central role of intentional teacher mediation in shaping students' interactions with such tools. This understanding supports the relevance of comparative designs that explicitly examine the presence or absence of teacher mediation, as well as investigations into teachers' pedagogical beliefs and professional development. Together, these lines of inquiry

may contribute to the development of evidence-based guidelines for the responsible, critical, and pedagogically grounded use of generative AI in science education.

Acknowledgement

The authors gratefully acknowledge the participating school and the students involved in this study for their collaboration and engagement, which were essential to the development of the research. We also thank the ESERA participants, researchers, and members who read, questioned, commented on, and offered suggestions for our work, with special appreciation to Jonathan Osborne, whose contributions and feedback during the presentation provided important inspiration for the final version of this study. Finally, we thank our professors, colleagues, and institutions that, directly or indirectly, contributed to the completion of this research.

References

- Araújo, L. A. de; Damacena, R.; Santos, Z. R. dos; Silva, D. A. da; Carnio, R. G. Metodologias ativas mediadas por tecnologia: análise crítica do papel do professor na aprendizagem colaborativa. *Interference Journal*, v. 11, n. 2, p. 2553–2573, 2025. DOI: 10.36557/2009-3578.2025v11n2p2553-2573.
- Barreira, J. , Moura-Silva, M. , & Gonçalves, T. (2024). Desafios e oportunidades para o uso do Chat GPT na pesquisa em educação matemática. *RIAAEE – Revista Ibero-Americana de Estudos em Educação*, 19(esp. 2), e024084.
- Božić, V. (2024). Chat GPT and education. General Hospital Koprivnica. https://www.researchgate.net/profile/Velibor-Bozic-2/publication/369926506_Chat-GPT-and-education/links/64350844ad9b6d17dc4d3a79/Chat-GPT-and-education.pdf.
- Cano, Y. M. (2024). ChatGPT and AI Text Generators: Should academia adapt or resist?. Harvard Business Publishing Education. <https://www.hbsp.harvard.edu/inspiring-minds/chatgpt-and-ai-text-generators-should-academia-adapt-or-resist>.
- Coll, C., & Solé, I. (2006). Os professores e a concepção construtivista. In C. Coll, E. Martín, T. Mauri, M. Miras, J. Onrubia, I. Solé, & A. Zabala (Eds.), *O construtivismo em sala de aula* (pp. 9-28). Ática.
- Farace, D.(2024). Inteligência artificial para aprendizagem: Chat GPT. Universidade Federal de Minas Gerais. <https://repositorio.ufmg.br/bitstream/1843/68473/1/Trabalho%20final%20-%20com%20todas%20as%20retifica%C3%A7%C3%B5es%20e%20ratifica%C3%A7%C3%B5es%20-%20Davi%20Dumont%20Farace.pdf>.
- Gomes, É., Franco, X., & Da Rocha, A. (2020). Aprendizagem significativa em David Ausubel. In *Uso de simuladores para potencializar a aprendizagem no ensino de Física* (pp. 18-26). EDUFT.
- Hidi, S., & Renninger, K. A. (2006). *The four-phase model of interest development*. *Educational Psychologist*, 41(2), 111–127. DOI:10.1207/s15326985ep4102_4. Disponível em: The Four-Phase Model of Interest Development.
- Nóvoa, A. (2017). Afirmar a posição como professor, afirmar a profissão docente. *Cadernos de Pesquisa*, 47(166), 1106–1133. <https://doi.org/10.1590/198053144843>

How To Teach Science To 3-6-Year-Old Pupils: Pre-Service Teachers' Perceptions Of Immersive Training With 360° Classroom Videos

Estelle Blanquet¹ and Alain Sénécaïl²

¹LACES University of Bordeaux - INSPE, France

²TECFA-FPSE University of Geneva, Switzerland

*The paper explores how immersive training using 360° classroom videos can enhance science education for children aged 3-6. In France, future primary teachers, often with limited scientific backgrounds, receive minimal science education during their Master's-level training, which poses challenges for teaching science effectively. To address this, the study investigates the perceptions of pre-service and early-career teachers regarding immersive 360° video training as a tool for improving their pedagogical approaches. The training involved two phases: a 2D workshop presenting a lesson on buoyancy and an immersive 360° session using a Quest headset. The 360° session featured augmented content and a sequence based on the children's book *Plouf* to explain the concept of pulleys. Ninety-four participants, primarily inexperienced teachers, evaluated the training through pre- and post-session questionnaires. Findings reveal that 73.3% found the 360° videos immersive, with key elements like spatial exploration and audio ambiance enhancing engagement. Over 74% felt the training significantly improved their understanding of scientific criteria, while 66.7% reported increased motivation to teach science through inquiry-based methods. However, 68% of participants experienced physical discomfort during 360° sessions, indicating room for improvement. The study highlights the potential of immersive tools such as 360° videos to bridge the gap between theory and practice in teacher training. Future research will compare the effect of the order of the two session modalities offered to students on their appropriation of the targeted elements of scientificity and perception of the 360° modality.*

Keywords: Immersive Teacher Education; Science Education; 360° video

Introduction

This paper explores how immersive training using 360° classroom videos can support science education in early childhood (ages 3–6). In France, future primary school teachers, many of whom have limited scientific backgrounds, receive relatively little training in science education during their Master's programs. This situation is particularly problematic given that international assessments consistently highlight significant difficulties in science learning among primary school pupils (IAEEA-TIMSS, 2024; OECD-PISA, 2025). Furthermore, opportunities for trainee teachers to observe science teaching practices in kindergarten classrooms are rare, and existing practices are often misaligned with curricular expectations (Cavaillès et al., 2023).

In response to these challenges, this study examines the potential of immersive technologies to enhance teacher education in science. More specifically, it focuses on the perceptions of pre-service and early-career teachers regarding the use of 360° classroom videos as a professional training tool, and on how such environments may help them better appropriate key elements of the scientific approach in early childhood education.

The training program was designed around two complementary modalities. The first consisted of a trainer-led workshop in which excerpts from a kindergarten science lesson on buoyancy, filmed using 360° video, were projected onto a conventional 2D screen. The second modality involved an immersive 360° session using a Quest headset, combining classroom environments with augmented pedagogical resources. This 360° sequence was based on the picture book *Plouf*

(Philippe Corentin, 1991) and aimed to introduce the principle of the pulley (Blanquet, 2010; MEN, 2023). Across both modalities, the training emphasized the explicit formulation and analysis of scientific criteria, such as the primacy of experimentation, reproducibility, and the robustness of experimental setups (Blanquet & Picholle, 2016).

The design of this training sequence draws on research suggesting that immersive and contextualized learning environments can support professional learning by fostering situated understanding, engagement, and transfer to practice (Degeuse & Sénécaïl, 2025; van den Beemt & den Brok, 2019). Studies have also shown that immersive technologies can enhance learner motivation and knowledge acquisition (Bower et al., 2020). In the field of teacher education, 360° video has been identified as a particularly relevant tool for providing authentic classroom experiences and supporting professional vision (Qian, Shang & Qin, 2023). However, despite growing interest in immersive video for teacher training, relatively few studies focus on specific subject areas (Roche, Rolland & Cunningham, 2023), and even fewer address science education in early childhood contexts (Tan, Wiebrands, O'Halloran & Wignell, 2020).

Building on this gap, the present study investigates how pre-service and early-career teachers perceive immersive 360° video-based training for teaching science in kindergarten.

Context and Theoretical Framework

In France, future primary school teachers are trained to teach pupils aged 3 to 11, and are recruited on the basis of a regional competitive examination at Master's level. Training institutes within universities offer Master's-level courses dedicated to the training of these teachers, the content of which is defined by the Ministry of Education at national level. Even though most of the students in these Master's programs do not come from scientific backgrounds, and the results of international tests show that primary school pupils have major difficulties in science, their training in science teaching is very limited (32 hours in all over the course of the Master's program). The situation is such that it's also difficult for them to observe teachers doing science particularly at Kindergarten level, as the latter are rare and their practices are rarely in line with expectations.

This paper addresses the critical need to enhance science education in early childhood (3–6-year-old) by empowering teachers with innovative teaching approaches, specifically utilizing 360° videos to foster engaging learning environments that stimulate curiosity and critical thinking through the analysis of in situ situations filmed in classrooms with teachers. Indeed, the teaching plan that we designed aims to equip future and in service teachers with the necessary skills to implement scientific criteria in their daily practices.

Researchers suggest that immersive technologies, provided with authentic and contextualized learning experiences, can enhance training (Milgram et al., 2020; 1995). The benefits of virtual immersion in education are varied: it offers variety to teaching methods, encourages situated learning, facilitates transfers to professional environments... (Dunleavy & Dede, 2014; Mellet-d'Huart, 2021, Walshe & Driver, 2019; Theelen, van den Beemt & den Brok, 2019). Moreover, research shows that interacting with these technologies stimulates student engagement and motivation, and improves knowledge acquisition (Slater, 2017; Bower et al., 2020). Regarding 360° video, it is considered an effective tool for assessing professional engagement, providing future teachers with immersive experiences closer to reality (Roche & Rolland, 2020; Qian, Shang & Qin, 2023). The potential of interactive video as a learning tool seems to be a real enabler, emphasizing its role in actively engaging students and improving cognition and attention (Altinpulluk, 2019; Merkt & Schwan, 2014). However, only few studies focus on a specific subject-matter (Roche, Rolland & Cunningham, 2023), and fewer on science education (Tan,

Wiebrands, O'Halloran & Wignell, 2020).

Research Questions

While 360° video immersive learning environments have been explored in terms of engagement, motivation, and professional development, little research has examined their perceived benefits and challenges for teaching science in preschool education. Although our study encompasses a broader analysis of the impact of 360° training, this paper specifically focuses on how teacher trainees perceive its effectiveness in supporting science instruction in early childhood settings.

Method and Design

We designed a two-phase training program using two different modalities:

- A trainer-led workshop based on extracts from a learning sequence on the concept of buoyancy filmed via 360° and projected onto a conventional 2D screen.
- A virtual reality training session (Quest headset) featuring hybrid resources: interactive environments with augmented documentation and a detailed presentation of a learning sequence, based on the picture book *Plouf* (Philippe Corentin, 1991), to explore the principle of the pulley.

Each phase was designed to include the presentation of the science sequence in kindergarten, with the emphasis on the teacher's explicit formulation of elements of the scientific approach, while providing a definition of these elements (testing reproducibility, the robustness of an experiment, the primacy of the experiment, etc.).

To assess the impact of the training sequence on perception and knowledge acquisition, students followed one of two pathways: either starting with the 2D workshop before moving on to the 360° training or reversing the order by beginning with virtual reality before engaging with the 2D workshop. The training was conducted six times with master's level students, totaling 94 participants (Table 1.), about half of whom (51) were very-early-career teachers.

Table 1. Participants.

		Preservice Teachers	Early Career Teachers
Age Average		24	35,7
Population		43	51
Gender	Male	11	8
	Female	32	42
	Other	-	1
Average teaching experience alone		Less than a week	Approximately three months part-time

Among those who had already taught, their teaching experience was very limited (only a few months at most), and very few had worked in preschool settings.

When it comes to their exposure to early childhood education, 80.9% reported having observed preschool classrooms, but only 42% had ever seen science being taught in these settings. When science activities were observed, they were primarily focused on biology, with an emphasis on hands-on activities, plant-related topics, and questioning strategies.

Questionnaires were distributed before the training sequence and after each phase (2D and 360°). The first questionnaire aimed to assess students' familiarity with immersive technologies as well as their initial conceptions of scientific criteria. The two subsequent questionnaires focused on gathering their perception of the training they had just completed, evaluating their sense of self-efficacy, and measuring their progress in understanding the definitions of scientific criteria. For this paper, we choose to focus exclusively on the question of pre-service and in-service teachers' perception of the 360° training.

Findings

Perception Of The 360° Experience And Setup

Overall Satisfaction

Figure 1. Global satisfaction concerning the immersive training.

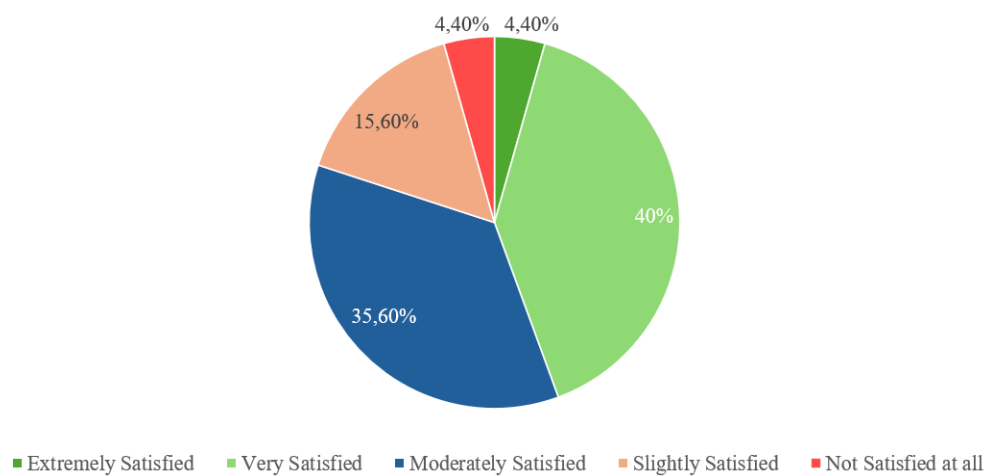
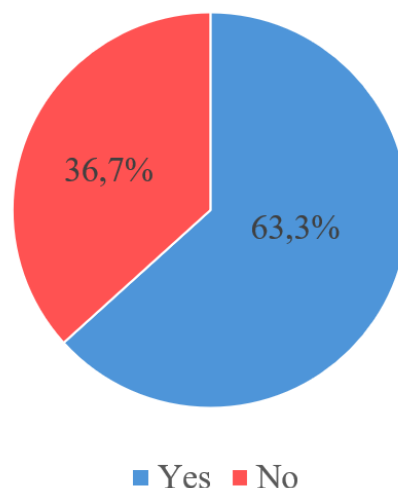


Figure 2. Participants' willingness to recommend the training to pairs.

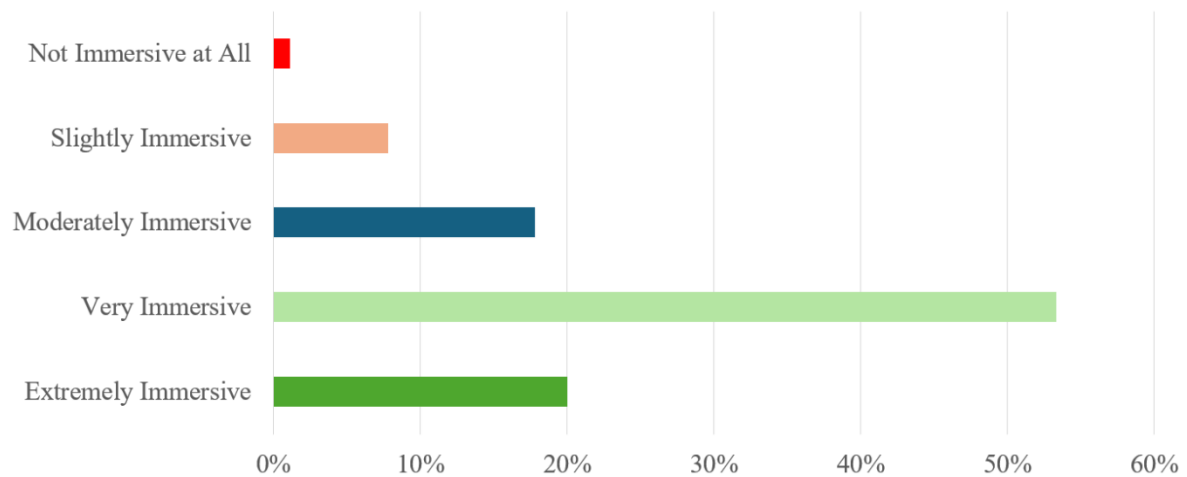


The results (Figures 1 and 2) show that 44.4% of the participants were extremely or very satisfied with the immersive training, while 35.6% rated it as moderately satisfying, which is an engagement with the experience. On the other hand, 20% expressed lower satisfaction, which is an opportunity to further improve the training in order to meet the expectations of the participants.

Similarly, 63.3% of participants would recommend this training to others, reflecting overall positive feedback. While some did not feel inclined to recommend it, this provides valuable insights for potential improvements to make the experience even more impactful.

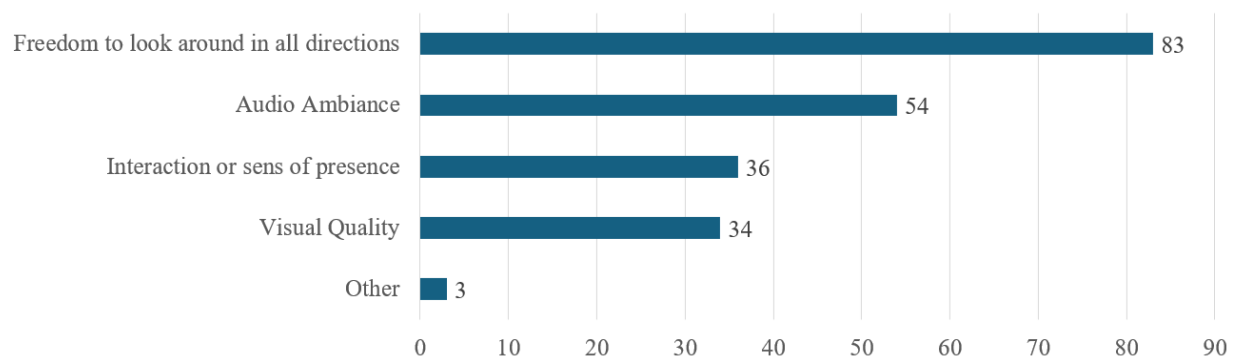
Immersion

Figure 3. Perceived level of immersion in the 360° video.



Most participants (73.3%) found the video very or extremely immersive (Figure 3), confirming its strong impact. With 17.8% rating it as moderately immersive and only 8.9% as slightly or not immersive, the experience was generally engaging.

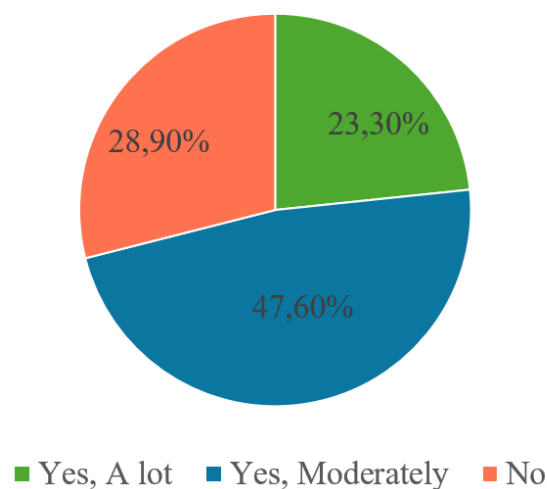
Figure 4. Aspects contributing to participants' sense of immersion.



The key factor enhancing immersion was the freedom to look around (83 responses), followed by audio ambiance (54) and visual quality (34). Interaction and sense of presence (36) also contributed, while only 3 participants mentioned other factors (Figure 4). This highlights the importance of spatial exploration and sound in creating an immersive 360° experience.

Furthermore, when asked whether the use of 360° video led to better engagement, the majority (71,2%) responded yes (Figure 5).

Figure 5. Perceived impact of the immersive format on participants' engagement with the content.



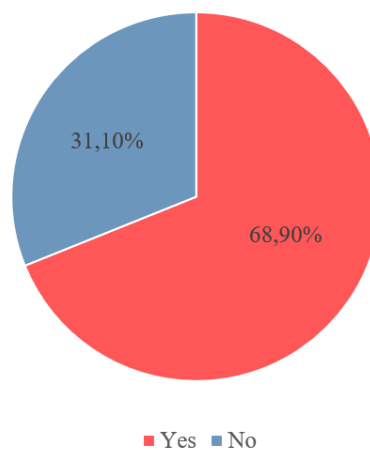
When asked about the reasons for their increased engagement (Table 2), participants emphasized the importance of immersion, particularly through a sense of involvement (10) and being in a science classroom (9). They also highlighted better concept understanding (6), stepping into the teacher's role and perspective (10), and connecting theory to practice (5). Technological affordance and freedom of navigation (4) further reinforced the immersive experience, making it a key factor in engagement.

Table 2. Categories related to participants' perceived reasons for increased engagement.

Category	Count
Sense of involvement/engagement	10
Immersion in a science classroom	9
Better understanding of the concepts at play	6
Projecting oneself into the role of a science teacher	5
Stepping into the teacher's perspective	5
Connecting theoretical content to real-world practice	5
Putting oneself in the place of a student	4
Technological affordance	4
Freedom of navigation and choice / autonomy	4

Nevertheless, physical discomfort emerged as a major obstacle (68.9%) to students' well-being during the 360° session (Figure 6)

Figure 6. Participants' reported physical discomfort.



Perceived Efficacy

The vast majority of students (96.6%) said they were able to identify the elements of the scientific approach targeted in the videos (360° session) and 74.4% that the session contributed enormously or a lot to their understanding of these elements. 66.7% considered that the 360° session enhanced their motivation to teach using the scientific inquiry approach.

Figure 7. Perceived contribution of the videos to understanding scientific criteria.

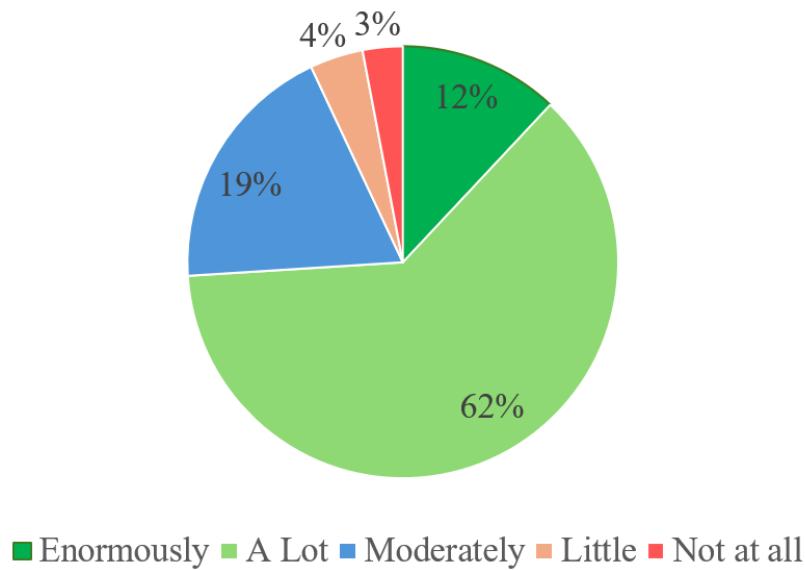
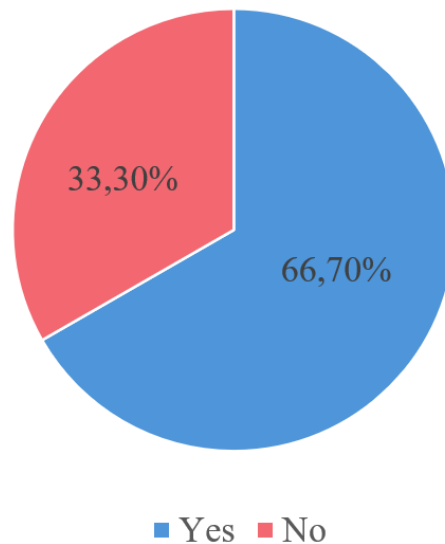


Figure 8. Perceived impact of the immersive experience on participants' motivation to teach the scientific method.



Discussion and Implications

Such results reinforce the idea that 360° videos (at least according to 74.4% of respondents) could enhance the learning of scientific criteria. It corresponds to the previous findings regarding the degree of engagement or immersion affected such aspects as the liberty of exploration in the environment (83 responses) and audio ambiance (54 responses), which presumably influenced an increase in conceptual understanding. Only 6.7% said they said there was little or no benefit.

On the other hand, the fact that 18.9% of the participants gave a moderated rating signals that the training, although efficacious, may not suffice for all learners and probably demands further assistance or additional materials for the understanding of the element of scientificity. This is where the instructional design could be improved, for example, by involving more interactive components or guided reflection exercises.

One extension for this study will be a comparison of the two cohorts, those who first have the 360° workshop and those afterwards, to identify and analyse differences between the two. For instance, it could appear that the students who did the 360° session second might be more likely to state that the 360° format adds nothing more than 2D session except for immersion.

The research is ongoing, and data is being processed. A comparison of pre- and post-tests will

also assess how students' knowledge and perception of scientific criteria evolve. This will help measure the impact of 360° training on their understanding and conceptual gains, as well as differences between training groups. Findings will be shared at the conference.

Acknowledgement

The authors would like to thank INSPE Bordeaux (University of Bordeaux), the Immersions 360 en Nouvelle-Aquitaine project, and the POLE PI (Pôle Innovation, INSPE Bordeaux) for their support in developing immersive 360° video resources and pedagogical innovations that enrich teacher education and training practices.

References

- Altinpulluk, H. (2019). “Determining the trends of using augmented reality in education between 2006-2016”. *Education & Information Technologies*, 24(2), 1089-1114.
- Blanquet, E. (2010). *Sciences à l'école, côté jardin : Le guide de l'enseignant*. Nice, France: Somnium.
- Blanquet, E., & Picholle, É. (2016). Two attainable skills in kindergarten: Testing reproducibility and robustness in an experiment. *Proceedings of the ESERA Conference 2015 (Helsinki, Finlande, 31 août–4 septembre)*. <https://hal.archives-ouvertes.fr/hal-01352266>
- Bower, M., DeWitt, D., Lai, J.W.M., (2020), Reasons associated with preservice teachers' intention to use immersive virtual reality in education, *British Journal of Educational Technology*, 51, 6, pp. 2214-2232.
- Cavaillès, A., Julien, S., Minault, B., Moreau-Fauvarque, C., Paitel, E., & Picaronny, C. (2023, 16 mai). La sensibilisation et la formation à la démarche scientifique (Rapport n° 21-22-099A). Inspection générale de l'Éducation, du Sport et de la Recherche ; Ministère de l'Enseignement supérieur, de la Recherche et de l'Espace. <https://www.enseignementsup-recherche.gouv.fr/fr/la-sensibilisation-et-la-formation-la-demarche-scientifique-90824>
- Degeuse, L., & Sénécaïl, A. (2025). Développer sa professionnalité grâce à la vidéo 360° : exemple d'un dispositif pour la formation des professeurs documentalistes. *Médiations et médiatisations*, (20), 138–151. <https://doi.org/10.52358/mm.vi20.435>
- Dunleavy, M., Dede, C., (2014), “Augmented Reality Teaching and Learning”, in Spector, J.M, et al.(eds.), *Handbook of Research on Educational Communication and Technology*, Springer Science+ Business Media, New York, pp.735-745.
- International Association for the Evaluation of Educational Achievement. (2024). *TIMSS 2023 international results in mathematics and science*. IEA. <https://www.iea.nl/studies/iea/timss/timss2023>
- Mellet-d'Huart, D. (2021). Learning in Virtual Environments: What About Paradigms and Metamodels? An Illustration Through Enaction and Trinologic. In G. Akcayir & C. Demmans Epp (Eds.), *Designing, Deploying, and Evaluating Virtual and Augmented Reality in Education* (pp. 1-19). IGI Global. <https://doi.org/10.4018/978-1-7998-5043-4.ch001>
- Ministère de l'Éducation nationale et de la Jeunesse. (2023, janvier). *Plouf ! ou la poulie* (PDF). Éduscol. <https://eduscol.education.fr/document/46399/download>
- Merkt M. & Schwan S., (2014), How does interactivity in videos affect task performance?, *Computers in Human Behaviour*, 31, p. 172-181.
- Milgram, P., Takemura, H., Utsumi, A. & Kishino, F. (1995). Augmented reality : a class of displays on the reality-virtuality continuum. *SPIE Proceedings*. <https://doi.org/10.1117/12.197321>
- Organisation for Economic Co-operation and Development. (2025). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Qian, J., Shang, J. and Qin, L. (2023). A systematic scoping review of 360-degree videos in teacher education, *Journal of Research in Innovative Teaching & Learning*, <https://doi.org/10.1108/JRIT-03-2023-0029>
- Roche, L., & Rolland, C. (2020). Scaffolding professional learning with 360° video for pre-service teachers. Dans E. Langran (dir.), *Proceedings of SITE Interactive 2020 Online Conference* (pp. 569-576). Online: Association for the Advancement of Computing in Education (AACE). <https://hal.science/hal-03092206>
- Roche, L., Rolland, C., & Cunningham, I. (2023). L'utilisation de la vidéo 360° dans la formation professionnelle des enseignants : une synthèse des connaissances. *Médiations Et médiatisations*, (15), 33–49. <https://doi.org/10.52358/mm.vi15.360>
- Slater, M., (2017). Implicit Learning Through Embodiment in Immersive Virtual Reality, In *Virtual, Augmented, and Mixed Realities in Education*, pp. 19-33.

- Tan, S., Wiebrands, M., O'Halloran, K., & Wignell, P. (2020) Analysing student engagement with 360-degree videos through multimodal data analytics and user annotations. *Technology, Pedagogy and Education*, 29(5), 593-612, <https://doi.org/10.1080/1475939X.2020.1835708>
- Theelen, H., van den Beemt, A., & den Brok, P. (2019). Using 360-degree videos in teacher education to improve preservice teachers' professional vision. *Journal of Computer Assisted Learning*, 35, 582–594. <https://doi.org/10.1111/jcal.12361>
- Walshe, N., & Driver, P. (2019). Developing reflective trainee teacher practice with 360-degree video. *Teaching and Teacher Education: An International Journal of Research and Studies*, 78(1), 97-105. <https://doi.org/10.1016/j.tate.2018.11.009>

Learning Through Comics And Students' Stress In Physics: An EDA Approach

Eman Sharaf and Martin Hopf

University of Vienna, Centre for Teacher Education, Austrian Educational Competence Centre
Physics, Austria

Physics is one of the most challenging sciences to study, and many students find it highly stressful. The significant increase in stress during learning negatively affects the outcomes of the educational process and students' motivation and achievement in physics. The aim of this paper is to explore the potential of educational comics as an instructional approach for reducing students' stress in physics learning. In this paper, we designed and developed two comics on energy topics, presented them to upper and lower secondary school students, and measured their stress levels using modern EDA technology via Moodmetric smart rings. The results showed that using comics in physics instruction significantly reduced students' stress compared to the traditional method. This result is an important indicator for those working in the field of physics education of the importance of providing students with educational scientific comics during learning.

Keywords: Comics, Stress, Electrodermal Activity

Introduction

A new trend in Physics Education Research (PER) is exploring ways to reduce negative emotions, such as anxiety and stress, among students. Comics can be visually stimulating and help alleviate pressures associated with learning complex scientific concepts. The objective is to advance the field by measuring students' stress levels using the Electrodermal Activity (EDA) technique. Prior research is limited in PER. This research aims to contribute to knowledge accumulation in this field.

Stress In Physics Education

According to the theory of emotions proposed by Russel and Mehrabian (1977), the emotional response can be delineated by three factors or dimensions: valency/degree of happiness, arousal/degree of intensity or stress, and dominance/degree of control (Russel & Mehrabian, 1977). However, Bodin and Winberg (2012) recommended the researchers to focus on the arousal or stress dimension. This research will investigate this dimension of the emotion. Citing the seminal Yerkes-Dodson law (1908) stipulates the relationship between stress and performance. This law posits that low levels of arousal lead to diminished performance due to a lack of motivation and concentration. Conversely, high stress can be detrimental to performance due to the onset of anxiety. Moderate arousal is regarded as optimal for peak performance. Molin et al. (2021) have indicated that physics students who experience anxiety or stress are more likely to achieve lower grades than their peers. For instance, students experiencing anxiety may encounter difficulties in regulating their emotions, consequently affecting their performance in physics classes. This may, in turn, result in learning difficulties and underachievement.

Learning Through Comics

The use of comics in science education is not a novel concept. The utilisation of comics is regarded as an efficient medium for conveying scientific concepts and enhancing student engagement. It has been argued that today's students encounter difficulties with extensive reading (Trnova, Trna, & Vacek, 2013), prompting physics educators and experts to devise alternative

methods for the exposition of physics concepts without the need for extensive reading. The use of well-crafted comics has been shown to enhance student interest and comprehension, particularly in physics (Yulianti, Khanafiyah, & Sulistyorini, 2016). Several PER researchers have explored the efficacy of comics in storytelling with university students, leading to a positive impact on both cognitive and emotional domains (Ozdemir, 2017). In research on the use of comics as a medium and content to reduce stress levels and enhance mental well-being, for example, a subsequent study was done by Brand et al. (2019). Also, other studies have explored the efficacy of comics as a pedagogical tool for secondary school students with the objective of enhancing comprehension of physics concepts and fostering greater motivation in the students towards their physics learning (Badeo & Ong Kian Koc, 2021). Research shows that comics positively impact students, but there's a lack of studies on physics education that address their use as stress-reducing tools.

Electrodermal Activity (EDA) In Education

EDA measurements facilitate the assessment of emotional arousal levels, i.e., the intensity of stress. EDA comprises two components: Skin Conductance Level (SCL), which measures long-term emotion and Skin Conductance Response (SCR), which measures short-term emotion (Boucsein., 2012). EDA sensors have been developed to measure stress and reflect an individual's emotional state (Geršak G. & Drnovše J., 2020). The use of EDA sensors is recommended for several reasons. This includes their relatively low cost, a simple and fast physiological response, and easy connection to raw data via Bluetooth, Wi-Fi, the Cloud, or cable (Braithwaite et al., 2013). EDA's applications extend to both laboratory and non-laboratory settings (Picard et al., 2015). However, further efforts are required within the educational sector to leverage the potential of EDA technology to its full potential.

Research Questions

RQ: How do students' average stress levels differ between comic-based and traditional text-based instruction in physics energy topics?

Our study explores the impact of a digital comic on students' energy-related topics, using Moodmetric smart rings to measure stress levels.

Methodology

Participants

The study involved 43 secondary school students (10 females, 33 males) in grades 9-12, who were either learning in English schools or had B2 proficiency, and was conducted in Alexandria, Egypt, in summer 2023.

Research Design

The study used a control-group experimental design. As shown in Figure 1, it has two experimental groups: one got the full version of comics (explanation modules and question modules are in comics), called 'comics-comics' C-C; the other experimental group got a version of comics and texts (explanation modules in comics and question modules are in text), termed 'comics-text C-T'; the control group received the full text version (explanation modules and question modules are in text), termed 'text-text' T-T, for testing.

Measurement Tools

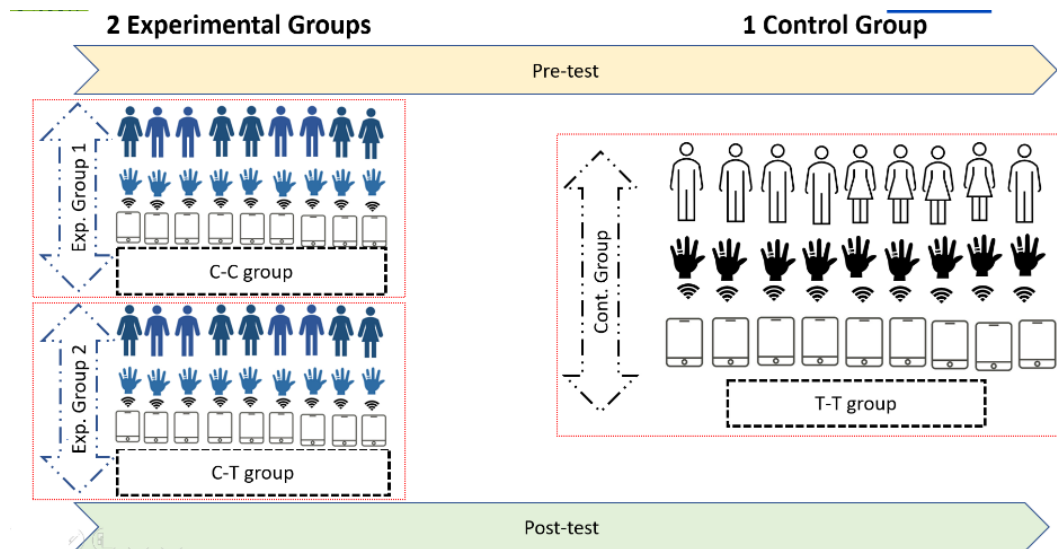
In this research, we used Moodmetric smart rings that measures student stress levels through EDA technique, including EDA/SCR and EDA/SCL per minute. They also measure Moodmetric

MM level, which combines SCR and SCL. Users can monitor stress levels through the app, which displays tension and calm times. These are expressed digitally as an index from 0 to 100, indicating the level of stress (e.g., MM levels 30-50 = moderate stress, <30 = relaxed, 51-55 = elevated stress, ≥ 56 = significantly elevated stress).

Procedure

We conducted an initial interview with 58 students, and they completed a pre-test on energy. 14 students left the research unfinished, and we had incomplete sensor data for another student. Then, a proportional random sampling was done to get the three groups (one control group and two experimental groups). Three digital learning models were developed: T-T, C-C and C-T. These focused on the same energy topics. All students studied the same content, either in comic or text format, interacted through question modules, and concluded with a post-test. The digital learning and the test were facilitated via a tablet using a dedicated website and occurred in laboratory settings. Students were given a moodmetric smart ring connected to a tablet to record stress which is shown in Figure 1.

Figure 1. Experimental design flowchart.

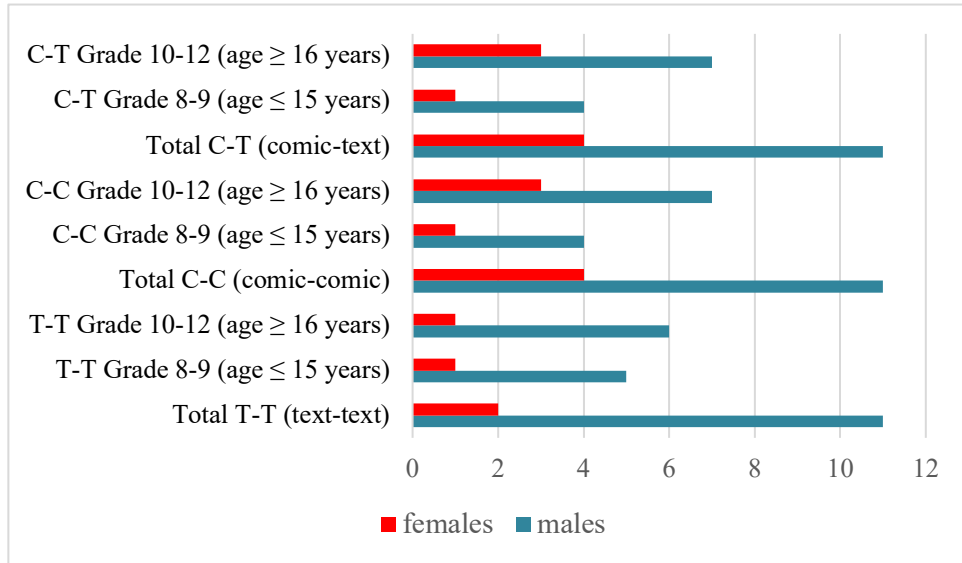


Result and Analysis

Sample Analysis

Figure 2 represents the analysis of the sample. The ratio of students in lower grades (8-9) to those in higher grades (10-12) is lower (12% vs. 23%), though similar across the two experimental groups C-C and C-T. However, in the control group T-T, the percentage of lower graders to higher graders is 14% and 16%. The female-to-male ratio is imbalanced (23% females to 77% males) because most of the volunteer students were learning in boys' schools. Another reason is the conservative Egyptian society, which sometimes stands still. This sometimes prevents girls from participating in research and international projects; however, we shall mention that none of the girls failed to attend and complete all steps of the pre-interview, learning session and post-test, indicating interest and motivation.

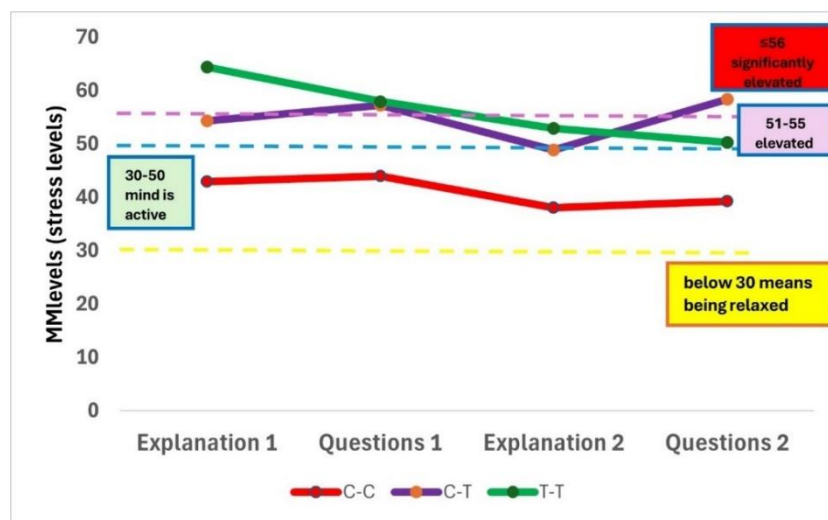
Figure 2: Sample analysis within the three groups according to gender.



Stress Analysis

General descriptive analysis has been done in order to answer the research question. Figure 3 illustrates the mean MM levels, which indicate students’ stress levels across the four modules in the three groups. C-C is represented in red, C-T is represented in purple, and T-T is represented in green. Figure 3 shows that the C-C group is below the C-T and T-T groups in the range between 30-50 MM levels, with a moderate stress area for all four modules’ explanations and questions. However, the MM levels of the C-T group fluctuate up and down; in the comics explanation modules, the MM levels are lower than in the text question modules, and MM levels go up to reach the T-T group in the first question module or exceed it in the second question module. Furthermore, the T-T group started with high MM levels in the first explanation module and then decreased in the following modules until it reached a moderate MM level in the last module.

Figure 3. Moodmetric MM levels of stress for three groups comics-comics C-C, comics-text C-T, text-text T-T during the four modules.



To test a null hypothesis of “There is no statistically significant difference in the mean stress levels between participants who received the comics-based instructional format and participants who received the text-based instructional format”, we applied One Way ANOVA for the three groups, a significant difference $F=3.70$ ($p=.033$) was found. A post hoc Tukey test determined a

significant difference in favour of the C-C group in comparison with T-T $F=19.4$ ($p=.026$). Therefore, we can reject this null hypothesis.

Discussion

This paper aims to answer the research question of how comics impact students' stress levels in physics. The stress line diagram and the observed patterns for the three groups (C-C, C-T and T-T) suggest that the C-C group experienced lower stress during explanations and questions. We can figure out from this study paper that using comics as a tool to present energy content significantly reduces stress levels of students among the C-C group during the learning process, compared with traditional text-based methods. The line graph shows that the T-T group experienced relatively stable high stress levels across all modules, reflected by an almost linear pattern. This may be because the text does not help to reduce students' stress levels, either in explanations or questions, resulting in a straight line with constant flow.

In contrast, stress levels declined in explanation modules and increased in question modules for the C-C and C-T groups. This may be due to the structure of the content itself, rather than the explanations or questions. This can be seen in the C-C group, where the narrative flow of the comics is disrupted by questions. This makes it difficult for students to immerse themselves in the story and escape the tension and stress. Accordingly, stress levels increased in the question modules and decreased in the comic explanation modules. The same pattern is evident in the C-T group, where stress decreases in the explanation modules and increases in the question modules.

The inferential analysis done to test the null hypothesis revealed a significant difference in mean stress levels between the comics-based groups and the text-based group, leading to rejection of the null hypothesis. In other words, the students who received the comics models reported significantly lower stress levels compared to the students who received the text models.

Conclusion

This paper investigated differences in students' stress levels when learning physics through comic-based and text-based instructional formats. Three groups were compared: a comic-comic (C-C) experimental group, a comic-text (C-T) experimental group, and a text-text (T-T) control group. We recognise the reduction in stress when using comics rather than text and when using a complex format combining comics and text.

Acknowledgement

This research project is a recipient of a DOC Fellowship of the Austrian Academy of Sciences (ÖAW) at the Institute of Austrian Educational Competence Centre AECC, Physics Department, from 2021-2023.

References

- Badeo, J. S., & Ong Kian Koc, B. (2021). Use of a comic-based learning module in physics in enhancing students' achievement and motivation. *Science Education International*, 32(2), 131–136.
- Bodin, M., & Winberg, M. (2012). Role of beliefs and emotions in numerical problem solving in university physics education. *Physical Review Special Topics – Physics Education Research*, 8(1), 010108. <https://doi.org/10.1103/PhysRevSTPER.8.010108>
- Brand, A., Gao, L., Hamann, A., Crayen, C., Brand, H., Squi, S., Stangl, K., Kendel, F., & Stangl, V. (2019). Medical graphic narratives to improve patient comprehension and periprocedural anxiety before coronary angiography and percutaneous coronary intervention: A randomized trial. *Annals of Internal Medicine*, 170(8), 579–587. <https://doi.org/10.7326/M18-2976>
- Braithwaite, J., Watson, D., Jones, R., & Rowe, M. (2015). *A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments* (2nd ed.).

University of Birmingham.

Molin, F., Cabus, S., Haelermans, C., & Groot, W. (2021). Toward reducing anxiety and increasing performance in physics education: Evidence from a randomized experiment. *Research in Science Education, 51*(Suppl. 1), 233–249. <https://doi.org/10.1007/s11165-019-9845-9>

Ozdemir, E. (2017). Comics in modern physics: Learning blackbody radiation through quasihistory. *Studies in Educational Research and Development, 1*(2), 21–59.

Picard, R. W., Fedor, S., & Ayzenberg, Y. (2015). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review, 8*(1), 62–75. <https://doi.org/10.1177/1754073914565517>

Rockett, D. (2020, February 27). Comics aren't just for fun—They can help prevent teen depression too. *Chicago Tribune*. <https://www.chicagotribune.com/2020/02/27/comics-arent-just-for-fun-they-can-help-prevent-teen-depression-too/>

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11*(3), 273–294. [https://doi.org/10.1016/0092-6566\(77\)90037-X](https://doi.org/10.1016/0092-6566(77)90037-X)

Trnova, E., Trna, J., & Vacek, V. (2013). The roles of cartoons and comics in science education. In *Proceedings of the 10th International Conference on Hands-on Science* (pp. 239–242).

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459–482. <https://doi.org/10.1002/cne.920180503>

Yulianti, D., Khanafiyah, S., & Sulistyorini, S. (2016). Inquiry-based science comic physics series integrated with character education. *Jurnal Pendidikan IPA Indonesia, 5*(1), 38–44.

Building Bridges: AI Custom Chatbots As Mediators Between Mathematics And Physics

Julia Lademann¹, Jannik Henze² and Sebastian Becker-Genschow¹

¹University of Cologne, Faculty of Mathematics and Natural Sciences, Digital Education Research, 50931 Cologne, Germany,

²University of Cologne, Faculty of Mathematics and Natural Sciences, Institute for Physics Education, 50931 Cologne, Germany

This study examines the use of AI custom chatbots in educational settings, with a particular focus on the interface between mathematics and physics. Given the growing adoption of AI tools like ChatGPT in learning environments, their potential as personalized tutoring systems is analysed. The study evaluates the impact of AI-generated learning materials on the learning experiences and performance of sixth-grade students, focusing on proportional relationships in mathematical and physical contexts. In a randomized controlled study (N = 214), traditional textbook materials were compared with explanations generated by an AI custom chatbot. The findings indicate that, while the AI-generated materials had an indefinite impact on learning performance, they significantly enhanced positive-activating emotions, situational interest, and self-efficacy expectations, while simultaneously reducing intrinsic and extrinsic cognitive load. These results underscore the potential of AI tools to transform educational practices and to enrich learning experiences. Further research is required to clarify the effects on learning performance and long-term learning outcomes. Overall, the present study highlights the importance of a careful integration and customization of AI tools in order to fully leverage their potential for physics education and related domains.

Keywords: Artificial Intelligence, K-12 Education, STEM Education

Introduction And Theoretical Background

Within current educational research, increasing attention has been directed toward the potential of AI chatbots in educational contexts (Chen et al., 2020). For instance, they can function as intelligent tutoring systems, support research and writing processes, provide individualized feedback, and address specific learning difficulties (Kasneci et al., 2023; Neumann et al., 2024). In STEM fields, chatbots can provide feedback to exercise-related and topic-specific questions (Avila et al., 2024; Liang et al., 2023), opening up new opportunities for the design of learning environments in specialized fields such as physics (Küchemann et al., 2024). While AI-based chatbots offer the potential to provide individualized support across diverse learning situations, empirical evidence regarding their effects on students' learning performance and learning-related experiences remains unclear (Neumann et al., 2024), making it beneficial to assess the potential influence within the context of physics education.

Students' comprehension difficulties in physics lessons are multifaceted – often related to insufficient mathematical skills (Christensen & Thompson, 2012; Leinhardt et al., 1990; Pollock et al., 2007) as well as to challenges in transferring mathematical knowledge to physical contexts (Redish & Kuo, 2015). The causes of these difficulties are not uniform across all students and may arise from misconceptions regarding the physical meaning of ratios, products, or functions (Uhden, 2016). In addition, comprehension problems may arise from difficulties in interpreting and linking multiple representations, such as formulas, graphs, diagrams, and tables (Treagust, 2018). As these representations encode important information, the ability to interpret and

transition between them is considered beneficial for learners (Hansen & Richland, 2020; Opfermann et al., 2017; Treagust, 2018). Targeted assistance in fostering this competence has been shown to enhance students' conceptual understanding (Hansen & Richland, 2020).

Vision-capable AI chatbots, such as ChatGPT, may offer a promising approach in this regard, as they are able to analyse images of representations (Kasneci et al., 2023; Neumann et al., 2024). In this manner, they can provide individualized feedback to support learners in interpreting and linking multiple representations as well as in extracting the encoded information.

Despite the availability of a multitude of chatbots, their effective use in educational settings often requires adaption to specific instructional purposes (Kieser et al., 2023). AI custom chatbots based on existing large language models can be configured to meet such requirements by specifying desired response behaviours, such as the use of situation-specific language or controlled answer length. Moreover, subject-specific knowledge can be incorporated through the integration of relevant instructional materials. Given the current limitations of research in this area, it remains an open question whether chatbots can be used to design suitable learning materials and generate explanations that positively influence students' learning processes. Against this background, the present study investigates the impact of explanations generated by an AI custom chatbot on students' learning experiences and learning performance in physics.

Research Questions

The study is designed as a randomized controlled field study, with students as the target group and is addressing the following key research questions:

How does learning with AI-generated explanations using an AI custom chatbot affect

RQ1: ... emotional aspects?

RQ2: ... situational interest?

RQ3: ... cognitive load?

RQ4: ... self-efficacy expectations?

RQ5: ... learning performance in a mathematical and physical learning context, particularly in terms of the necessary transfer performance to the physical context?

Methodology

Chatbot Design

The AI custom chatbot used in this study was developed on the basis of the model GPT-4.0 and adapted to a specific learning context and target group. The chatbot uses engaging and age-appropriate language, making it suitable for sixth and seventh grade students. Its focus lies on “proportional relationships” as an example of learning with multiple representations, including graphs, tables, and formulas. Students must develop the ability to interpret these representations and transition between them from both mathematical and physical perspectives (Bing & Redish, 2009; Nilsen et al., 2013). To this end, the AI custom chatbot was equipped with subject-specific knowledge and designed to analyse and interpret the different forms of representations in order to provide targeted and individualized feedback.

Nevertheless, it remains open to question whether chatbots are capable of designing suitable learning materials and environments, as well as generating text-based explanations that have a positive impact on the aforementioned constructs. In the present study, the chatbot was employed to provide an AI-generated, text-based explanation of the topic “proportional relationships”, based on a textbook overview (Körner, 2020). The explanation was selected according to the following criteria: the accuracy of the content and language, as well as the correct and coherent

use of the three forms of representation – graph, table, and formula.

Study Design

A total of $N = 214$ sixth-grade students (146 female, 66 male, 2 n.a.) with an average age of 11.7 years ($SD = 0.51$) at German secondary schools were randomly assigned to an experimental (EG) or control group (CG). Both groups were provided with learning materials on “proportional relationships” which had not been covered in class before. The CG received conventional textbook material, including a topic overview and accompanying explanatory examples. The EG was provided with the same textbook overview but had the examples replaced by a text-based explanation of the topic generated by the custom chatbot.

Immediately following the learning phase, various variables were measured using a four-point Likert scale: the participants' emotions (Pekrun et al., 2011) and situational interest (Linnenbrink-Garcia et al., 2010), the intrinsic and extrinsic cognitive load (ICL and ECL) experienced during learning (Leppink et al., 2013) and the self-efficacy expectation with regard to solving a topic-related task (Jerusalem & Satow, 1999) were assessed. Subsequently, the students completed a self-developed performance test on the “spring scale”, which served as an example of the physical concept of proportional relationships.

Data Analysis

The data analysis was conducted using the statistical software R (version 4.4.0). First, the scales used were tested for internal consistency by calculating Cronbach's alpha. The results were acceptable to excellent, and no significant improvements were identified when individual items were removed from the scales. Therefore, the scales were retained for further data analysis. The data was then tested for normal distribution. The Shapiro-Wilk test with a significance level $\alpha = 0.05$ was used to determine the distribution of the data. The results indicated that the data was not normally distributed. The non-parametric Mann-Whitney U test with a significance level $\alpha = 0.05$ was used in the subsequent analysis to identify significant differences between the EG and CG. Cohen's d was calculated to estimate the respective effect size.

Results

The results of the study are summarized below. The complete methodology and results of the study can be found in the publication released in May 2025 (Lademann et al., 2025).

Concerning emotional aspects (RQ1), the EG reported a significantly higher incidence of positive-activating emotions (pleasure and satisfaction) than the CG ($p = 0.00093$, $d = 0.48$, small effect size). Regarding negative-deactivating emotions (boredom, frustration and uncertainty), only a trend could be identified, showing that the CG tended to experience more negative emotions. In sum, the results indicate that learning with AI-generated explanations can have a positive impact on the emotional state of the learners (Figure 1). Regarding the influence on situational interest (RQ2), learning with the custom chatbot's explanation triggered significantly higher situational interest ($p = 0.00223$, $d = 0.45$, small effect size) compared to learning with textbook materials (Figure 2).

Figure 1. Perceived emotions with standard error and level of significance (Lademann et al., 2025).

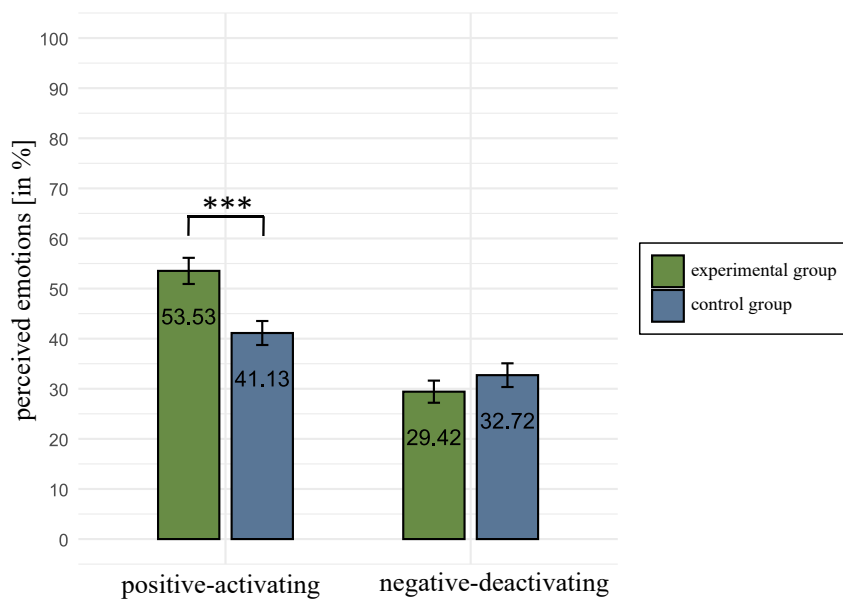
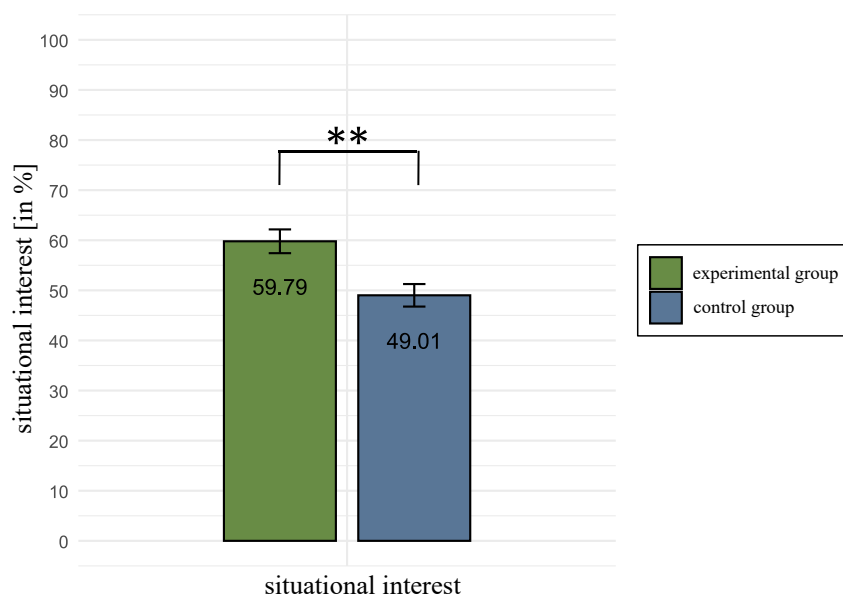


Figure 2. Situational interest with standard error and level of significance (Lademann et al., 2025).



In contrast, ICL ($p = 0.00060$, $d = 0.47$, small effect size) and ECL ($p = 0.0001$, $d = 0.59$, medium effect size) were significantly reduced in the EG (Figure 3). These findings suggest that learning with AI-generated materials diminished cognitive load compared to textbook-only materials (RQ3). Additionally, the EG had significantly higher self-efficacy expectations ($p = 0.00001$, $d = 0.63$, medium effect size) than the CG (RQ4), when it comes to solving a topic-related task (Figure 4).

Concerning the influence on learning performance in a mathematical and physical context (RQ5), no clear trends were identified in the overall result of the performance test or the subtasks. The available data prevents drawing definitive conclusions about the impact of AI-generated materials on learning outcomes compared to traditional textbook materials. After thorough analysis, it was determined that there were no significant differences between the EG and CG in terms of learning performance.

Figure 3. Intrinsic cognitive load (ICL) left and Extrinsic cognitive load (ECL) right, both with standard error and level of significance (Lademann et al., 2025).

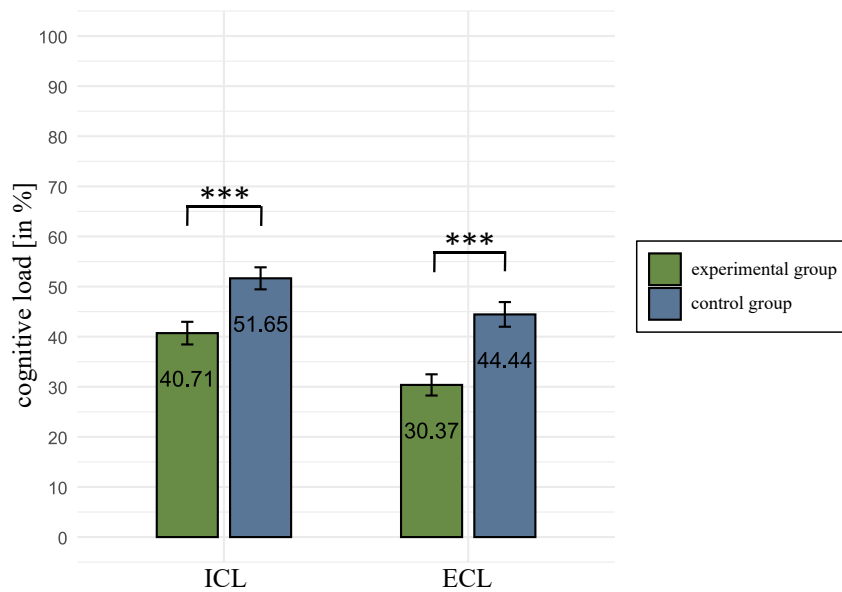
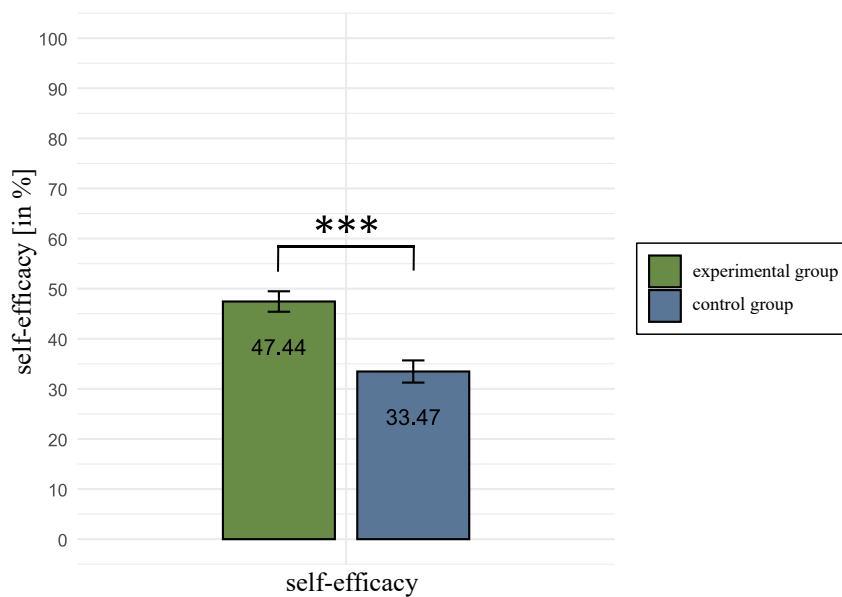


Figure 4. Self-efficacy with standard error and level of significance (Lademann et al., 2025).



Discussion And Conclusion

The results of the present study indicate that learning with materials generated by the AI custom chatbot has a positive effect on the learning experience of students.

The findings indicate that learning with AI custom chatbot-generated explanations, compared to traditional textbook materials, significantly enhances learners' positive-activating emotions and situational interest. The configuration of the chatbot with respect to communicating in a manner intended to encourage and motivate the learner might provide a potential explanation for these outcomes. The motivating tone may have exerted a positive influence on emotions and situational interest.

Furthermore, in comparison to traditional learning with textbook materials, cognitive load was reduced. This could be due to the chatbot's use of accessible language and its conversational and human-like tone. Additionally, the chatbot is adapted to the students' level of knowledge and uses language that is comprehensible to them. Moreover, the EG showed significantly higher self-efficacy, indicating that the explanation of the AI chatbot can foster students' confidence in their

ability to complete tasks. A reason for this may be the encouraging tone of the AI-generated text or the reduced cognitive load, or a combination of both. With regard to learning performance, the data revealed no statistically significant differences. A study comprising a greater number of participants over an extended period of time, in conjunction with individual utilization of the chatbot, has the potential to yield more profound insights.

Several factors limit the generalizability of the present findings, including the relatively small size of the sample, the short duration of the intervention, and the focus on a single topic tailored to a specific target group. Additionally, the learning materials were provided in a centralized manner, which resulted in the lack of consideration of individual differences in material use and the interaction with the chatbot. Consequently, the potential effects of personalized and adaptive feedback could not be examined.

Future research should therefore employ longer-term studies that cover a broader range of topics, and include more diverse target groups in order to validate and extend the findings and assumptions presented here. Moreover, learners should be given the opportunity to interact with the AI chatbot independently to systematically investigate the effects of individualized feedback.

Building on the present results, a subsequent study will conduct a more comprehensive examination of learning outcomes. An optimized AI custom chatbot will be developed, including revisions, refinements, and an adaptation to the didactic framework of the follow-up study. Within a pre-post design, the independent use of the chatbot by students will be examined with regard to its effects on learning performance and the learning experience (emotions, situational interest, cognitive load, self-efficacy expectations), and the technological acceptance of the AI chatbot.

Acknowledgement

We would like to express our sincere thanks to all participating schools, teachers and students for contributing to our study.

References

- Avila, K., Steinert, S., Ruzika, S., Kuhn, J., & Küchemann, S. (2024). Using ChatGPT for teaching physics. *The Physics Teacher*, 62. <https://doi.org/10.1119/5.0227132>
- Bing, T. J., & Redish, E. F. (2009). Analysing problem solving using math in physics: Epistemological framing via warrants. *Physical Review Special Topics - Physics Education Research*, 5(2), 020108. <https://doi.org/10.1103/PhysRevSTPER.5.020108>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Christensen, W. M., & Thompson, J. R. (2012). Investigating graphical representations of slope and derivative without a physics context. *Physical Review Special Topics - Physics Education Research*, 8(2), 023101. <https://doi.org/10.1103/PhysRevSTPER.8.023101>
- Hansen, J., & Richland, L. E. (2020). Teaching and Learning Science through Multiple Representations: Intuitions and Executive Functions. *CBE—Life Sciences Education*, 19(4), ar61. <https://doi.org/10.1187/cbe.19-11-0253>
- Jerusalem, M., & Satow, L. (1999). Schulbezogene Selbstwirksamkeitserwartung [School-related self-efficacy expectation]. In *Skalen zur Erfassung von Lehrer- und Schülermerkmalen* (pp. 15–16). Institut für Psychologie, Freie Universität Berlin.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kieser, F., Wulff, P., Kuhn, J., & Küchemann, S. (2023). Educational data augmentation in physics education research using ChatGPT. *Physical Review Physics Education Research*, 19(2), 020150. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020150>

Körner, H. (Ed.). (2020). *Mathematik—Neue Wege 7 [Mathematics - new approaches 7]* (Für Gymnasien [G9, Nordrhein-Westfalen, Schleswig-Holstein], Druck A). Schroedel.

Küchemann, S., Steinert, S., Kuhn, J., Avila, K., & Ruzika, S. (2024). Large language models—Valuable tools that require a sensitive integration into teaching and learning physics. *The Physics Teacher*, 62, 400–402. <https://doi.org/10.1119/5.0212374>

Lademann, J., Henze, J., & Becker-Genschow, S. (2025). Augmenting learning environments using AI custom chatbots: Effects on learning performance, cognitive load, and affective variables. *Physical Review Physics Education Research*, 21(1), 010147. <https://doi.org/10.1103/PhysRevPhysEducRes.21.010147>

Leinhardt, G., Stein, M. K., & Zaslavsky, O. (1990). Functions, Graphs, and Graphing: Tasks, Learning, and Teaching. *Review of Educational Research*, 60(1), 1–64. <https://doi.org/10.3102/00346543060001001>

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behaviour Research Methods*, 45(4), 1058–1072. <https://doi.org/10.3758/s13428-013-0334-1>

Liang, Y., Zou, D., Xie, H., & Wang, F. L. (2023). Exploring the potential of using ChatGPT in physics education. *Smart Learning Environments*, 10(1), 52. <https://doi.org/10.1186/s40561-023-00273-7>

Linnenbrink-Garcia, L., Durik, A., Conley, A., Barron, K., Tauer, J., Karabenick, S., & Harackiewicz, J. (2010). Measuring Situational Interest in Academic Domains. *Educational and Psychological Measurement - EDUC PSYCHOL MEAS*, 70, 647–671. <https://doi.org/10.1177/0013164409355699>

Neumann, K., Kuhn, J., & Drachsler, H. (2024). Generative Künstliche Intelligenz in Unterricht und Unterrichtsforschung – Chancen und Herausforderungen [Generative artificial intelligence in teaching and education research - opportunities and challenges]. *Unterrichtswissenschaft*, 52(2), 227–237. <https://doi.org/10.1007/s42010-024-00212-6>

Nilsen, T., Angell, C., & Grønmo, L. S. (2013). Mathematical competencies and the role of mathematics in physics education: A trend analysis of TIMSS Advanced 1995 and 2008. *Acta Didactica Norge*, 7(1), (Art. 6-21 sider). <https://doi.org/10.5617/adno.1113>

Opfermann, M., Schmeck, A., & Fischer, H. E. (2017). Multiple Representations in Physics and Science Education – Why Should We Use Them? In D. F. Treagust, R. Duit, & H. E. Fischer (Eds.), *Multiple Representations in Physics Education* (Vol. 10, pp. 1–22). Springer International Publishing. https://doi.org/10.1007/978-3-319-58914-5_1

Pekrun, R., Götz, T., Frenzel, A., Barchfeld, P., & Perry, R. P. (2011). *Measuring emotions in students' learning performance: The achievement emotions questionnaire (AEQ)*. <http://kops.uni-konstanz.de/handle/123456789/13732>

Pollock, E. B., Thompson, J. R., & Mountcastle, D. B. (2007). Student Understanding Of The Physics And Mathematics Of Process Variables In P-V Diagrams. *AIP Conference Proceedings*, 951(1), 168–171. <https://doi.org/10.1063/1.2820924>

Redish, E. F., & Kuo, E. (2015). Language of Physics, Language of Math: Disciplinary Culture and Dynamic Epistemology. *Science & Education*, 24(5), 561–590. <https://doi.org/10.1007/s11191-015-9749-7>

Treagust, D. F. (2018). The Importance of Multiple Representations for Teaching and Learning Science. In *Education Research Highlights in Mathematics, Science and Technology 2018* (pp. 215–223). ISRES.

Uhden, O. (2016). Verständnisprobleme von Schülerinnen und Schülern beim Verbinden von Physik und Mathematik [Students' comprehension problems when combining physics and mathematics]. *Zeitschrift für Didaktik der Naturwissenschaften*, 22(1), 13–24. <https://doi.org/10.1007/s40573-015-0038-4>