Part 7 / Strand 7

Discourse And Argumentation In Science Education

Co-editors: *Andri Christodoulou & Metin Sardag*

Part 7 / Strand 7 Discourse And Argumentation In Science Education

Understanding, supporting, and promoting the use of evidence and argumentation discourse in science education. Includes scientific practices related to knowledge evaluation and communication, supporting the development of critical thinking, discourse analysis, talking and writing science in the classroom, and meaning making in science classrooms.

Sub-themes:

1) Analysis of Argumentation and Student Discourse in Science Education
2) Developing Critical Thinking through Science Argumentation
3) The Role of Talking and Writing in Science Knowledge Construction
4) Tools and Techniques for Enhancing Scientific Discourse
5) Meaning-making in Science Classrooms

Contents

Strand 7: Discourse And Argumentation In Science Education

Metin Şardağ[1] *and Andri Christodoulou*[2]
[1]Van Yuzuncu Yil University, Turkey
[2] University of Southampton, UK

Argumentation and discourse studies in science education can be defined as a broad research field that focuses on how individuals evaluate, justify, explain, discuss, and meaningfully use knowledge in scientific and social contexts on the basis of evidence (Kelly, Brown & Jiménez-Aleixandre, 2023). This field examines how scientific knowledge is spoken about, written, represented, defended, and questioned. Therefore, research on argumentation and discourse constitutes a strong area of intersection among scientific literacy, critical thinking, media literacy, science communication, and socio-scientific issues.

One of the main reasons why this field has become more visible today is that information circulates through media and digital environments at an irresistible speed and reaches large audiences within a short period of time. Misinformation, distorted scientific claims, sources of uncertain reliability, and anti-science discourses spreading through social media make it necessary to question the source of information, its evidence base, the context in which it is produced, and the way it is presented (Osborne & Pimentel, 2023).

The studies included in this section make this need visible through different age groups, subject areas, and methodological approaches. Kelles et al. (2026) examine students' relationships with media content and information sources in biology lessons and reveal the importance of teacher questions in questioning the origin, reliability, and possible conflicts of interest of sources. Matsuyama et al. (2026), through the Argu-made 2.0 system developed in the context of the bioaccumulation of pufferfish toxins, support sixth-grade students' processes of selecting appropriate evidence and constructing claim-evidence-reasoning structures. Dachauer and Lembens (2026) demonstrate that the Toulmin Argumentation Model can be used to improve the structural quality of high school students' scientific arguments in the context of ocean acidification. Santos Coutinho and Sasseron (2026), in a writing activity on climate change using the En-ROADS Climate Solutions simulator, show that students partially activated their critical thinking skills; however, they also needed further pedagogical support for strong justification and the evaluation of alternative perspectives. These studies indicate that selecting evidence, constructing reasoning, and questioning the reliability of information are learning processes that need to be explicitly designed and taught in science lessons.

The discourse dimension of the section includes studies on how teacher-student interactions shape learning opportunities. Cortés Morales and Marzabal (2026) analyse teachers' discursive moves in modeling-oriented instruction that make students' ideas visible, prompt them to question these ideas, and guide them toward a shared scientific model. Loidl et al. (2026), by positioning students who produce podcasts on electromagnetic radiation as science communicators, emphasize that content knowledge alone is not sufficient for producing high-quality explanations; explanation and communication skills also need to be explicitly supported. Kang and Adofo (2026), by combining automatic speech recognition and social network analysis, show that participation in science group work should be understood not only in terms of the amount of talk, but also together with variables such as students' relational positions, gender, science identity, and extraversion. Thus, the section addresses discourse both in terms of teachers' moment-to-moment pedagogical decisions and students' practices of explanation, participation, and meaning-making.

The main trends in these studies appear to concentrate on argumentation, source evaluation, teacher discourse, the production of scientific explanations, learning supported by digital tools, and equitable participation. Another noteworthy trend is that students are positioned as active meaning-makers who construct arguments, write explanations, produce podcasts, use simulators, question sources, and participate in group interactions.

From a methodological perspective, these studies indicate that classroom interactions should be examined not only through outcome indicators but also through process data. Video recordings, students' written products, simulator outputs, podcasts, automatic speech recognition data, and social network structures show that research on argumentation and discourse relies on multilayered, multi-modal data sources. In this way, the field opens up both the effects of pedagogical interventions and students' positioning in classroom interactions to more detailed discussion.

The themes of sustainability and digital developments also strengthen the contemporary importance of this field. Climate change, ocean acidification, global warming news, food safety, and sustainability-themed group work enable students to relate scientific knowledge to social and environmental responsibilities. Digital argumentation systems, simulators, podcasts, automatic speech recognition, and social network analysis both support students' thinking and communication processes and provide researchers with opportunities to examine classroom discourse in a more detailed and scalable way.

References

Kelly, G. J., Brown, B., & Jiménez-Aleixandre, M. P. (2023). Discourse practices in science learning. In *Handbook of research on science education* (pp. 413-446). Routledge.

Sources Of Information And Media In Biology Lessons: Relationships In A First-Grade High School Classroom

Ludmila Kelles[1], Blanca Puig[2] and Luiz Franco[1]
[1]Universidade Federal de Minas Gerais, Brazil
[2]Universidade de Santiago de Compostela, Spain

In this paper, we present an ethnographic study in which we observed Biology lessons over an academic year in a Brazilian 1st-year High School classroom. Based on video records of the lessons, we select events in which misinformation was a central part of discussions. These events were described in event maps and transcribed into message units. Analyses were guided by notions of reflection/refraction of language (Bakhtin circle) in order to answer two questions: i) For these students, how does the media reflect and/or refract science when disseminating diverse socioscientific issues? ii) How teachers' questions influenced students' evaluation of information sources throughout the year? As results, the students showed two roles that media could assume: as refracting scientific information, by: i) deceiving; ii) hiding; iii) distorting; iv) using retracted research data; v) using outdated information or manipulating news; or as reflecting scientific information, when: i) being used as checking tools; ii) recognizing mistakes and correcting itself. The teacher's questions were relevant in bringing about changes in students' evaluation of information sources throughout the year. His questions challenged arguments based on authority and gave visibility to key aspects of source analysis, such as explaining a source's origin, assessing the credibility of cited sources and considering possible conflicts of interest. These results bring empirical evidence that pedagogical activities articulated with the aim of promoting autonomous and well-informed positioning can generate learning opportunities beyond superficial criticality.

Keywords: Media literacy, Sources of information, Biology lessons.

Context And Relevance To Science Education

In this study we investigate how High School students discursively constructed relationships between science and media over Biology lessons in a public school in Brazil. Allchin (2022) argues that the role of the media is not secondary in science education. Students should have opportunities to understand how science is mediated, how it is communicated to society and what are the science images disseminated for different media.

Such demands are linked to discussions on how post-truth scenarios has affected schooling and, more specifically, science education in the contemporary world (Allchin, Bergstrom & Osborne, 2024). The media (traditional and new media) are one of the main foci of such issues. Traditional media lose support and trust (Chinn; Barzilai & Duncan, 2021), while new media produce echo chambers and increase the dissemination of fake news (Barzilai & Chinn, 2020).

The role of science education in this scenario goes beyond to promote critical thinking for distinguish between what is "true" and "false", to take critical action (Martínez Pena et al., 2024), what is coherent with the "criticality" notion presented by Davies and Barnett (2015). Other discussions have been demanded at school, such as: who invests in misinformation, what are the interests in propagating it, what is behind false information, what mechanisms are linked to beliefs in misinformation, what is the role of sources (e.g., Barzilai & Chinn, 2020). We seek to contribute to these discussions through a perspective of social interactions in the classroom as linguistic processes in which critical thinking can emerge and take action.

Theoretical Framework And Research Questions

People, through discourse, continually negotiate the meanings of texts, events and contexts in the flow of their social life. In the post-truth era a very large part of critical thinking is evaluating the truth – or at least plausibility – of the premises (Lyons & Ward, 2024) or claims that are spread by social media.

Thus, meanings of certain textual and contextual elements of language can reflect or refract other texts, events and contexts (Bloome et al., 2022). This conception is particularly relevant to our study. When analysing how students relate media and science, it is important to consider that the meanings, interpretations and relationships proposed by them are intrinsically related to how these students reflect and/or refract histories, cultures, references, knowledge, values and norms.

The notions of reflection and refraction used in this paper are based on scholars from the so called Bakhtin circle. Language can be used to reflect or refract aspects of a culture, a context, a history. When language reflects something, it is maintaining certain historical, social and cultural meanings, without changes or questions, that is, it reproduces what is communicated. When language refracts something, in turn, it is being used to modify the original, historical meanings, considered standard for that particular culture, resulting in changes in meanings (Bakhtin, 1981).

To contribute to discussions in science education field, we investigated discursive interactions in a 1st year high school classroom (15-year-old students in Brazil) over the course of an academic year. The following questions guided the study: i) For these students, how does the media reflect and/or refract science when disseminating diverse socioscientific issues? ii) How teachers' questions influenced students' evaluation of information sources throughout the year?

Research Method And Design

The study took place in a public school in a big city in Brazil. Our research team conducted participant observation and recorded written field notes (Spradley, 1980). Data were gathered from videotaped records of Biology lessons and students' artifacts to better understand forms of participation in the classroom (e.g., Castanheira et al., 2001).

We followed a 1st year high school classroom (9 boys and 24 girls) over the course of the academic year in 2021. As a result of the COVID-19 pandemic, the school adopted an emergency remote teaching system, in which teachers used platforms such as Google Classroom® to communicate with the class. For online synchronous lessons, Google Meet® was used.

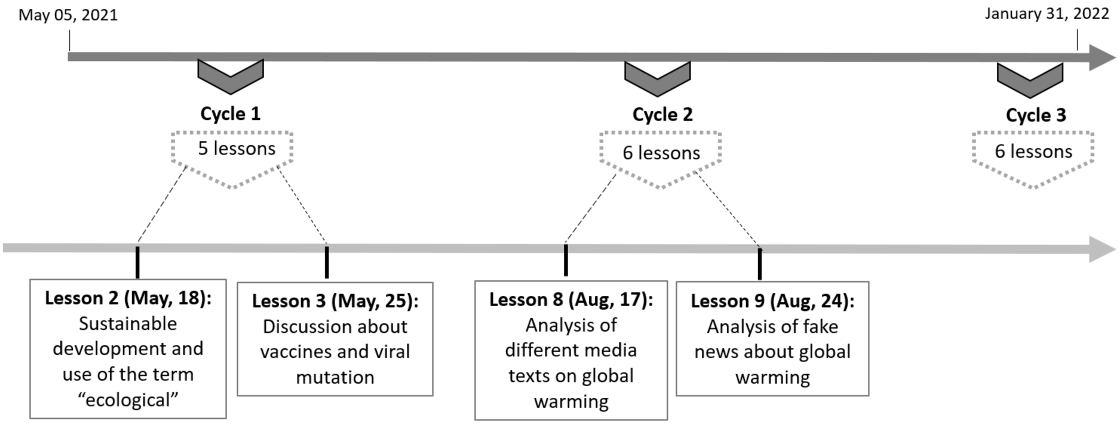
The lessons were organized in three cycles of activities (Figure 1) in which students participated in scientific inquiry units (e.g., Pedaste et al., 2015) and activities guided by the use of socio-scientific issues (e.g., Sadler, 2009).

We adopted elements of an ethnographic perspective on education to develop our analysis. We built macroscopic representations, as charts of lessons and event maps (Castanheira et al., 2001) that allowed a broader view of the history of the group in different levels of detail. These representations made it possible to establish whole-part relationships, situating events in the group's history. From this panoramic view, we selected four lessons for analysis of interactions (Figure 1).

When analysing the 17 lessons of the year, we selected 4: lessons 2, 3, 8 and 9. The selected events were chosen because misinformation was a central part of each SSIs discussion in these lessons, thus opportunities for critical evaluation. The events were transcribed word by word and the analysis was guided by proposals on reflection/refraction of language (Bahktin, 1981).

In lesson 2, students analyse three media that use the word "ecological" in different situations: 1) ecological leather for clothing production; 2) water reuse in towel washing; 3) ecological bags produced from organic matter. In lesson 3, the teacher showed a map with the distribution of dengue cases around the world. Then, the students discussed possible relationships between the development of a dengue vaccine and the ecological data on the map. In lesson 8, students analyse three reposts about global warming. Each repost addressed different aspects of the phenomenon, including false information. Finally, in lesson 9, the teacher showed fake news about global warming that has become common in WhatsApp® groups. The students discussed how they would react to that information if they received it in a group of their families.

Figure 1. Biology lessons selected for analysis in this paper.



Findings

Throughout the selected lessons (Lessons 2, 3, 8 and 9), we identified different relationships between media and science constructed by the students. We also observed changes over the lessons regarding the role of sources. As the complete discursive interactions are too long for this paper, we opted for showing excerpts from the participants' discourse in order to illustrate the analyses (Figure 2).

Regarding our first research question, the analysis indicates different interpretations by students about the relationship between media and science. We found events in which the media was interpreted as an agent capable of refracting scientific information, when it acted by: i) deceiving the reader (Lesson 2 and 9); ii) hiding data (Lessons 8 and 9); iii) distorting them (Lessons 3 and 8); iv) using retracted research data (Lesson 8); v) using outdated information (Lesson 8); or manipulating news (Lessons 2, 8 and 9).

Despite this, the students also indicated the media as an agent capable of reflecting scientific information. For example, when the media: i) consistently uses scientific knowledge (Lessons, 2, 3 and 8); ii) disseminates actions corresponding to scientific practice (Lesson 2); iii) promotes actions aligned with science official discourse (Lesson 2); iv) recognizes mistakes and corrects itself (Lesson 9). In addition, the students indicated that the media could also act in this way by functioning as a tool for checking information (Lessons 2 and 8).

Regarding our second research question, we mapped teacher issues onto student performances when evaluating information sources. Table 1 shows some of these questions.

In the first lessons of the school year, the students started from more restricted analyses, relying on arguments from authority. The teacher's questions challenged authority arguments, which appeared, for example, in the speeches of Gleice and Gisele (Lesson 2). In addition, the questions gave visibility to relevant aspects of source analysis, such as the need to explain the origin of the source and, even, to send them to colleagues, as seen, when discussing with Isabel (Lesson 3). In lessons 8 and 9, we observed more complex verifications that included assessing the credibility of the cited sources and considering possible conflicts of interest (e.g., Gisele's remarks in Lesson 8 and Lúcia's in Lesson 9). Questions made by the teacher were also important in this process.

Figure 2. Illustrative excerpts from the analyses. All excerpts are student speeches over the lessons.

	Lesson 2	Lesson 3	Lesson 8	Lesson 9
How does the media refract scientific information?	“it was a lot of marketing by these companies to put the word ecological” “all three are useless, because it doesn't help those who consume think they are leaving of consuming animal leather since it is not true” “it's more of an excuse to say you did something”	“I confused some data, it's been more than three months that I had been watched”	“in this text, he used the fallacy of authority a lot” “not to mention that the research is from 1958 and his text is from 2017, and notions of global warming have already changed a lot” “in the text itself, he uses very complicated words and graphics to try to somehow fit the research done” “He also uses information in a distorted way to convince people”	“Many of his studies have montages in the images. He makes a mistake and that's why his information was wrong” “Teacher, what you said is true that there is manipulation of the news even from important newspapers” “Of course, this newspaper will stop criticizing climate change. Not necessarily spreading misinformation but choosing what they're going to talk about.”
How does the media reflect scientific information?	“there was a survey there in the United States that even made a documentary in which the result was that even if everyone reduced the water bill by forty percent, it would not be enough to not run out of water in the near future why the problem it's the industries” “I logged into the hotel's Instagram while we were discussing and I didn't see anything about tree planting”	“I saw a video from a YouTube channel. It's something like that. They brought the data and then the vast majority of people got dengue haemorrhagic fever. They brought the research data” “I saw reports talking about studies of two new viruses that cause symptoms similar to those of dengue or Zika that were being researched, but it is not yet known if these are dengue virus”		
How do students deal with sources of information?	“it is a biologist who creates, because I think it conveys more credibility” “the last one was the only one that we thought the term ecological fits and also the only one [media] that has the real intention of helping the environment so much so that he was a biologist, right?”	“The information is from a biomedical”. “Googling more, I just found an infectologist talking about the 5 dengue viruses”.	“NASA only posts on its website, the research it uses as a basis” “The sources he uses are unreliable and he uses this political scope a lot”	“the sources he uses are not very reliable” “he also uses an author a lot in this report. A man called Nils-Axel Mörner, who was a former president of the IPCC. When we went to research, we saw that he is not a source that has credibility”

We identified some challenges over this process. Even in the lessons we observed more complex analysis of sources, students still gave authority to some figures without going into detail (e.g. the authority that Izabel gave to the teacher and the researcher who had written the text analysed in classroom 8). So, the changes in analysing sources of information did not happen linearly but happened throughout the events in a variable way and correlated to teacher’s questions over the discussions.

Table 1. Excerpts of the teacher's questions over the lessons and students' reactions.

Lesson	Teacher questions	Student	Excerpts from students' performances in evaluating sources
2 2	So, I wanted to put a reflection here for you to think about. Is there a biologist who does bad things for the environment?	Gleice	(…) the third one is interesting (..) that says is a biologist who creates, because, yes, I think it gives more credibility.
	Does every biologist have this environmental concern because he is a biologist? So, for you to think about (…)	Gisele	Yes, the real intention of helping the environment so much so that he was a biologist, right?
3 3	Based on what are you saying this? On a report you saw in a newspaper?	Izabel	I saw a video from (…) from a YouTube channel (…) they show the data from research.
	Because when we make a statement like that, it's good to bring the source, the data that we are relying on (…) Can you send it to us?	Izabel	Yes, I'll send it (…) I'll send it.
8 8	You commented on partisan political use of information, remember?	Gisele	(…) the text (…) it is radical to one side. The side that does not consider what (…) what the human being affects.
	But you have to doubt and even criticise what I say (…) Got it? You must be critical of your own teacher	Izabel	Wow teacher (…) I said (…) he is making fun of us (…) seriously, he didn't (…) he didn't put this here (…) for real
9	You should check what's in the source, who is that source (…) Who's talking about this?	Lúcia	There is manipulation of news (…) this newspaper not necessarily (…) will spread misinformation but choose what they will say.

Conclusions

The results of this research bring empirical evidence that pedagogical activities articulated with the aim of promoting autonomous and well-informed positioning can generate learning opportunities beyond a shallow criticality. We align ourselves with the proposals of Barzilai and Chinn (2020), when they suggest that science education should go beyond just identifying deficiencies in people's knowledge and skills to critically deal with information. Science teaching must be able to deal with cognitive biases, assessing disagreements about "how to know", recognizing and coordinating multiple epistemologies.

The analyses indicate challenges in this process. Naive views about science and difficulties in engaging in argumentative practices are examples of such challenges (Nehring, 2020) and were identified in our data. In this context, the teacher has a role in the organization and conduction of didactic processes. Our analyses reinforce the importance of the school as a fundamental institution in social organization, as well as the need to value actions and curricula for autonomous and critical students.

Acknowledgement

We would like to thank National Council for Scientific and Technological Development - CNPq Brazil (process: 402839/2024-2), for their support.

References

Allchin, D. (2022). Ten competencies for the science misinformation crisis. Science Education, 107(2), 261–274. https://doi.org/10.1002/sce.21746

Allchin, D., Bergstrom, C. T., & Osborne, J. (2024). Transforming Science Education in an Age of Misinformation. Journal of College Science Teaching, 53(1), 40–43. https://doi.org/10.1080/0047231X.2023.2292409

Bakhtin, M. M. (1981). The dialogic imagination: Four essays. Vol. 1. University of Texas Press.

Barzilai, S., & Chinn, C. A. (2020). A review of educational responses to the "post-truth" condition: Four lenses on "post-truth" problems. Educational Psychologist. 55(3), 107-119, https://doi.org/10.1080/00461520.2020.1786388.

Bloome, D., Power-Carter, S., Baker, W., Castanheira, M., Kim, M., & Rowe, L., W. (2022). Discourse Analysis Of Languaging And Literacy Events In Educational Settings: A Microethnographic

Perspective. (1st ed.). Routledge.
Castanheira, M. L., Crawford, T., Dixon, C., & Green, J. (2001). Interactional Ethnography: An Approach to Studying the Social Construction of Literate Practices. Linguistics an Education, 11(4), 353-400. https://doi.org/10.1016/S0898-5898(00)00032-2.
Chinn, C., Barzilai, S., & Duncan R. G. (2021). Education for a "Post-Truth" World: New Directions for Research and Practice. Educational Researcher. 50(1), 51–60.
Davies, M., & Barnett, R. (2015). The Palgrave handbook of critical thinking in higher education. New York: Palgrave Macmillan US.
Lyons, J., & Ward, B. (2024). The New Critical Thinking: An Empirically Informed Introduction (2nd ed.). Routledge. https://doi.org/10.4324/9781003311027
Martínez-Pena I., Puig, B. & Uskola, A. (2024). One Health education for criticality on vaccination in teacher training. Front. Public Health. 12:1408965. https://doi.org/10.3389/fpubh.2024.1408965
Nehring, A. (2020). Naïve and informed views on the nature of scientific inquiry in large-scale assessments: Two sides of the same coin or different currencies? J Res Sci Teach. 57, 510–535. https://doi.org/10.1002/tea.21598.
Pedaste, M., Mäeots, M., Siiman, L. A., Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. Educational Research Review, 14, 47-61. https://doi.org/10.1016/j.edurev.2015.02.003.
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.
Spradley, J. P. (1980). Participant Observation. New York: Holt, Rinehart and Winston.

EUROPEAN SCIENCE EDUCATION RESEARCH ASSOCIATION ESERA 2025 Copenhagen

Exploring High School Students' Mobilization Of Critical Thinking Skills Through Argumentation About Climate Change

João Pedro Santos Coutinho[1], Lúcia Helena Sasseron[2]
[1]Interunit Graduate Program in Science Teaching of the University of São Paulo, Brazil
[2]School of Education of the University of São Paulo, Brazil

The purpose of this study was to investigate how Brazilian high school students mobilized critical thinking through a writing activity in the Chemistry class in which the students had to investigate and analyse a simulator related to the impact of anthropogenic variables on climate change, entitled En-ROADS Climate Solutions Simulator. The activity was developed with 57 first-year high school students at a Brazilian public school in the state of São Paulo. In order to analyse the information collected, we used the Critical Thinking Analytic Rubric (CTAR), composed of 6 categories (interpretation, analysis, evaluation, inference, explanation and disposition) evaluated with scores from 1 to 6. The results showed that students mobilized half of what would be expected to be the ideal level of high cognitive demand skills related to critical thinking, since most of the scores were at the midway point. This highlights the need to promote more activities in science teaching that help students develop critical thinking, but it also reveals limitations both in the wording of the activity questions analysed in this study and in the ability of the rubrics to assess actions related to critical thinking. Therefore, it is interesting that further research on the topic is conducted in order to address these issues.

Keywords: Critical Thinking. Scientific Argumentation. Classroom Research.

Introduction

In recent years, the spread of misinformation and disinformation involving scientific topics through digital means has reached alarming proportions and, even though young people are usually expected to have skills necessary to use today's technology effectively, studies show their struggle with evaluating information (Osborne et al., 2022). Considering this, it is a Science Education priority to promote classroom research and practices that foster students' higher order thinking skills development by understanding the social practices of science, evaluating information, and developing a consensus that help justify trust in scientific knowledge.

Therefore, critical thinking (CT) development, which is associated with higher order thinking skills such as the ones described above (Saxton et al., 2012), is essential to promote the development of critical subjects who can identify incorrect scientific information, hopefully reducing its spread. Based on that, this study's objective is to answer the following research question: "*How do high school students mobilize critical thinking skills in the Chemistry class through a writing activity aligned with a global climate simulator analysis?*"

Conceptual Rationale

The development of CT is considered a consensual objective of education by scholars (Jiménez-Aleixandre & Puig, 2022; Kuhn, 1999; Kuhn, 2019). Despite this, its understanding and definition are not homogeneously disseminated, that is, there are different perspectives regarding the characterization of CT, involving its meaning, components and how to develop them (Jiménez-Aleixandre & Puig, 2022).

One of the first and most influential definitions was proposed by Ennis (1985), who defined CT as "reasonable, reflective thinking that is focused on deciding what to believe and do" (p. 45). Later, the American Philosophical Association (APA) held a Delphi panel with various CT experts, including Ennis, aiming to propose a consensual definition of CT. This panel considered that CT include both cognitive and dispositional dimensions, defining it as “purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based” (Facione, 1990, p. 2).

Years later, Davies and Barnett (2015) discussed some shortcomings of the APA definition, such as its rooting in only one kind of CT, related to argumentation and judgment formation. For this reason, they proposed changing this perspective to one in which CT consisted of “criticality,” a perspective including action. Even though both movements considered that the development of skills and dispositions related to CT are important, the criticality movement differed from the traditional CT movement because it defended that the skills and dispositions by themselves weren't enough and they should be guided by ethical values and result in critical action.

Similarly, Jiménez-Aleixandre and Puig (2022) reviewed and proposed a holistic characterization of CT aiming to explicitly integrate both perspectives. This proposal considers that CT is characterized by two sets each one comprised of two components, the first set marked by purposeful judgment, aligned with the APA definition, with the components cognitive and epistemic skills and critical character; and the other set marked by civic participation and social justice, aligned with criticality, with the components independent opinion and critical action.

Regarding CT assessment, it is commonly derived from multiple choice tests, such as Critical Thinking Assessment Test (CAT) (Stein & Haynes, 2011) and Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1964), which arguably do not measure the higher order skills associated with CT, since they are not designed nor have the purpose of formative feedback (Reynders et al., 2020; Saxton et al., 2012). Given this, instead of using tests to assess CT, rubrics could be used, as they provide more detail regarding the scores, what is being measured and what should be included in the data interpretation (Saxton et al., 2012).

Considering this, Saxton et al. (2012) proposed and experimentally studied the Critical Thinking Analytic Rubric (CTAR), heavily influenced by the APA definition, stating that it can be used by raters to score student work samples in a consistent manner. Since they adopted the APA definition as the construct for the instrument, it focuses on the cognitive and epistemic skills and critical character, assessing “five cognitive skills (interpretation, analysis, evaluation, inference, and explanation) and the disposition towards thinking critically” (Saxton et al., 2012, p. 254), rating these categories in scores of 1 to 6.

The categories are briefly defined as follows: *Interpretation* is concerned with understanding and accurately communicating the meaning of information; *Analysis* involves examining and comparing different sources of information; *Evaluation* involves identifying arguments and judging the credibility of sources; *Inference* involves making realistic predictions and forming rational and logical conclusions; *Explanation* is related to the use of evidence and reasoning to support an argument or perspective; *Disposition* towards CT is an affective dimension of CT, highlighting the need for the person to be motivated to mobilize their higher order thinking skills (Saxton et al., 2012).

Methodology And Research Design

This study uses a qualitative approach (Erickson, 2012) and involves the analysis of an activity carried out at a public school in São Paulo (Brazil) in the Chemistry subject by 57 first year high school students (14–16 years old) from different social classes and ethnicities in two classes, divided in 6 groups of up to 5 students per class, totaling 12 groups, anonymized as G1 to G12. The students received tablet computers with internet connection and they had to answer a question while exploring the En-ROADS Climate Solutions Simulator (Climate Interactive & MIT Sloan, 2023), which is a simulation model that allows the exploration of scenarios focused on how changes in taxes, subsidies, carbon prices, and other factors can alter global carbon emissions and the global temperature in the following years. After the activity was completed, we collected the written responses and digitized them following the ethical procedures of informed consent, transparency and confidentiality.

We provided the students with a worksheet that introduced the simulator, explaining it was developed by Climate Interactive and the MIT Sloan Sustainability Initiative. The worksheet also had a summary of a special report of the Intergovernmental Panel on Climate Change (IPCC, 2022) that stated the goal of limiting global warming to around 1.5°C (2.7°F) by 2100. The question was as follows: "*Point out two variables present in the En-ROADS Climate Solutions simulator that you consider most important to be changed to achieve the target proposed by the IPCC. Explain how changing these variables can influence global temperature.*"

In order to analyse the students' responses, we used the CTAR instrument (Saxton et al., 2012) adapted to the question the students had to answer, as presented in Table 1. Both authors scored the responses individually and then compared the scores, discussing and reaching consensus regarding the divergent scores.

Table 1. Overview of the Critical Thinking Analytic Rubric. Adapted from Saxton, Belanger and Becker (2012).

Rubric categories	Scores					
	1	2	3	4	5	6
Interpretation	Demonstrates no ability to work with the key concepts	Demonstrates an extremely limited ability to work with the key concepts	Demonstrates an uneven or shaky ability to work with the key concepts	Demonstrates a clear ability to work with the key concepts	Demonstrates a strong ability to work with the key concepts	Demonstrates confident ability to work with the key concepts and terminology
Analysis	Does not identify patterns in the evidence	Incorrectly identifies patterns in the evidence	Superficially identifies patterns in the evidence	Identifies patterns in the evidence	Accurately identifies patterns in the evidence	Accurately identifies important patterns in the evidence
Evaluation	Ignores all alternative perspectives	Ignores obvious alternative perspectives	Superficially evaluates obvious alternative perspectives	Evaluates all alternative perspectives	Evaluates most perspectives	Thoughtfully evaluates all perspectives

(Table Continues)

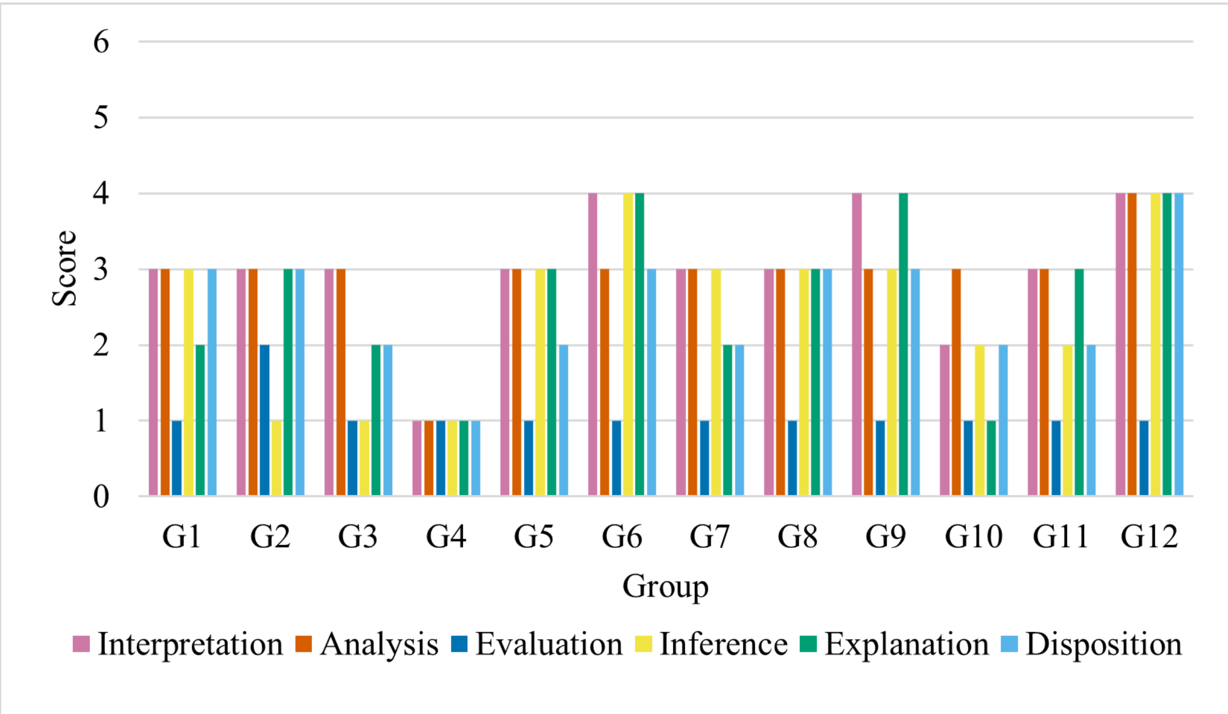
Table 1. Continued.

Rubric categories	Scores					
	1	2	3	4	5	6
Inference	Cannot extend key concepts or make predictions	Inadequately extends key concepts or make predictions	Shakely extends key concepts or predictions	Adequately extends key concepts	Clearly extends key concepts to make predictions	Confidently extends key concepts to make predictions and analyse implications
Explanation	Uses fallacious or irrelevant reasons	Conclusions lack support	Conclusions are weakly supported	Conclusions appear reasonable through use of support	Strong, warranted conclusions	Judicious, persuasive conclusions
Disposition	Close-minded, hostile to reason	Defends only with a single perspective, ignores others	Follows evidence but ignores context	Fair-mindedly follows evidence, addresses context	Fair-mindedly follows evidence, considering context	Objectively considers context and evidence

Findings

The results of the CTAR scores for the response of each group can be seen in Figure 1.

Figure 1. CTAR scores' results for the students' responses.



Although we analysed all the responses, for the purposes of this study, we present, as examples, the analysis of G10's and G12's responses. They presented, for the most part, the lowest and highest scores, respectively, among those with at least *Disposition* score 2, considering that a *Disposition* score 1 means that the response didn't present any justification, which was the situation of G4, whose response only pointed out the variables but did not justify the choice. It is important to note that the student's responses were written in Portuguese and were translated to English for this paper.

G10’s response was as follows: “*The increase in the carbon price and the decrease in methane gas emissions. With the increase in the carbon price, we would use less forms and objects that emit gas, and the decrease in methane gas emissions increases the greenhouse effect on Earth.*”

Regarding *Interpretation*, G10's response showed a limited ability to work with the key concepts, considering that they stated that the decrease in methane gas increases the greenhouse effect when it’s the opposite, hence score 2. As for *Analysis*, both variables listed were related to the decrease of the global temperature, but this relation is presented superficially, therefore score 3. In regard to the *Evaluation* category, the response did not present any alternative perspective, so score 1. Considering *Inference*, as they incorrectly relate the methane to the greenhouse effect and don’t provide much detail about the increase in the carbon price, the category’s score was 1. In the matter of the *Explanation* category, the reasons used to warrant the claim are incorrect, hence score 1. Finally, as regards the *Disposition* category, the response only defended one perspective and didn’t discuss the context that prevents those variables from being changed, so score 2.

G12’s response was as follows: “*The main variables that need to be changed to limit the temperature to 1.5°C by 2100 are: coal, the oil market, and the carbon price, as these are issues that directly impact climate change. Another variable would be the economic development of Third World countries that would reduce the burning of fossil fuels as an energy source and replace it with the use of clean and renewable energy.*”

From this response, we realized that, in relation to *Interpretation*, G12 not only identified both variables, but also perceived the nuance of the relation between the economic situation of a country and its access to sustainable energy, therefore, score 4. Regarding *Analysis*, they identified correctly the patterns in the evidence, thus score 4. As for *Evaluation*, we noticed that they only explained the positive sides of changing each variable, ignoring other points of view and hurdles needed to be overcome in order to change those variables, therefore, score 1. Since they adequately extended the key concepts, especially the relation between economic development and renewable energy sources, *Inference*’s score was 4. Furthermore, as they did not raise any negative side, *Explanation* seems reasonable rather than fully justified, hence score 4. Finally, regarding *Disposition*, we considered that G12 followed the evidence provided by the simulation and addressed the socioeconomic context, therefore, score 4.

Conclusion

In this study, we sought to investigate how high school students mobilize CT skills by structuring their arguments after analysing anthropogenic variables that contribute to the worsening of climate change with the help of an online simulator. Through the data presented, we noticed that most of the categories presented scores equal to 3, stating that in general the students mobilized half of what would be expected to be high cognitive demand skills related to CT.

Interpretation was the skill they managed to mobilize the best, considering most of the groups, except for G4, achieved a score of 3 or higher. This is evidence that the students, in the context of the Chemistry class, are more comfortable with working and applying the key concepts related to the subject.

The poor result in the *Evaluation* category may come from the wording of the question, which did not explicitly ask for alternative points of view. The only group that achieved a score higher than 1 was G2, that pointed out the variable of nuclear energy and briefly talked about its risks, while defending it to be clean and safe due to containment systems.

In future research, we should create the questions considering the opportunity to talk about different perspectives so that the students can better mobilize the *Evaluation* category. The new questions should be created in a way that the students are motivated to mobilize high cognitive demand skills, so that the questions can't be answered with low score mobilizations.

The CTAR, as previously mentioned, is heavily tied to the APA definition of CT (Facione, 1990). So, it doesn't assess critical action and critical virtue defended by the criticality movement (Davies & Barnett, 2015). Considering this, it is important to state that only assessing students' essays and activity responses does not provide the full picture of CT mobilization, as it can be mobilized through actions and dialogue (Jiménez-Aleixandre & Puig, 2022; Kuhn, 2019).

Finally, this study's findings indicate both the urgency of implementing practices that mobilize CT in Science classes, since students don't seem to be used to mobilizing all the CT skills, as well as the need for academic research in Science Education regarding CT development, such as the assessment of students' criticality, as rubrics are limited to assess actions.

Acknowledgement

This study was financed, in part, by the São Paulo Research Foundation (FAPESP), Brazil. Processes Numbers #2023/11360-0 and #2024/21301-3.

References

Climate Interactive & MIT Sloan. (2023). EnROADS Climate Solutions Simulator (Version 23.8.0) [Computer software]. Climate Interactive. https://www.climateinteractive.org/en-roads

Davies, M., & Barnett, R. (2015). Introduction. In M. Davies & R. Barnett (Eds.), *The Palgrave Handbook of Critical Thinking in Higher Education* (pp. 1–25). Palgrave Macmillan US. https://doi.org/10.1057/9781137378057_1

Erickson, F. (2012). Qualitative Research Methods for Science Education. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second International Handbook of Science Education* (pp. 1451–1469). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9041-7_93

Ennis, R. H. (1985). A Logical Basis for Measuring Critical Thinking Skills. *Educational Leadership*, *43*(2), 44–48.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. California State University.

IPCC. (2022). *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157940

Jiménez-Aleixandre, M. P., & Puig, B. (2022). Educating Critical Citizens to Face Post-truth: The Time Is Now. In B. Puig & M. P. Jiménez-Aleixandre (Eds.), *Critical Thinking in Biology and Environmental Education: Facing Challenges in a Post-Truth World* (pp. 3–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-92006-7_1

Kuhn, D. (1999). A Developmental Model of Critical Thinking. *Educational Researcher*, *28*(2), 16–46. https://doi.org/10.3102/0013189X028002016

Kuhn, D. (2019). Critical Thinking as Discourse. *Human Development*, *62*(3), 146–164. https://doi.org/10.1159/000500171

Osborne, J., Pimentel, D., Alberts, B., Allchin, D., Barzilai, S., Bergstrom, C., Coffey, J., Donovan, B., Kivinen, K., Kozyreva, A., & Wineburg, S. (2022). *Science Education in an Age of Misinformation*. Stanford University. https://sciedandmisinfo.stanford.edu/

Reynders, G., Lantz, J., Ruder, S. M., Stanford, C. L., & Cole, R. S. (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. *International Journal of STEM Education*, *7*(1), 9. https://doi.org/10.1186/s40594-020-00208-5

Saxton, E., Belanger, S., & Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing*, *17*(4), 251–270. https://doi.org/10.1016/j.asw.2012.07.002

Stein, B., & Haynes, A. (2011). Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test. *Change: The Magazine of Higher Learning*, *43*(2), 44–49. https://doi.org/10.1080/00091383.2011.550254

Watson, G., & Glaser, E. M. (1964). *Watson-Glaser Critical Thinking Appraisal Manual*. Harcourt, Brace, and World.

EUROPEAN SCIENCE EDUCATION RESEARCH ASSOCIATION ESERA 2025 Copenhagen

Deep Dive: Evaluating Students' Use Of Toulmin's Argumentation Pattern In Chemistry Classrooms For Constructing Scientific Arguments About Ocean Acidification

Jennifer Dachauer and Anja Lembens
University of Vienna, Austria

Science education plays a key role in helping students to become responsible citizens, capable of making appropriate decisions based on scientific information, participating in discussions on socio-scientific issues (SSIs), and taking action in the context of climate change. Thus, achieving scientific literacy in general and practising scientific argumentation skills – as part of scientific literacy – in particular, are essential. Therefore, the pilot study presented here focusses on improving upper secondary students' (ISCED 3) scientific argumentation skills by implementing Toulmin's Argumentation Pattern (TAP) in the chemistry classroom. We focus on chemistry education to better understand subject-specific argumentation in chemistry classrooms. This pilot study is part of an interventional study that aims to implement TAP as a construction mechanism to improve students' scientific argumentation skills. Our study is embedded in the SSI of human-induced ocean acidification. In order to gain insight into students' scientific argumentation patterns and develop adequate teaching materials, we asked 29 upper secondary students to write well-founded arguments to confirm a claim about calcification of coral skeletons made in a newspaper headline. Students' arguments were constructed prior to and following the provision of support material based on TAP. These arguments were analysed with regards to their formal quality using qualitative content analysis. Our findings indicate that TAP aligned well with our participants' intuitive arguments. We discuss our findings and propose how using TAP can contribute to teaching and learning chemistry-specific argumentation skills.

Keywords: Scientific Argumentation, Scientific Literacy, Classroom Research

The Need To Improve Scientific Argumentation Skills In Chemistry Education

At present, society is confronted with numerous challenges regarding climate change (IPCC, 2023). To address these challenges, science education plays a key role in helping students to become responsible citizens, capable of taking action in order to protect our environment (Bianchi et al., 2022). In this context, there is a relevant association with scientific literacy as "the ability to engage with science-related issues, and with the ideas of science as a reflective citizen" (OECD, 2019, p. 100). Being scientifically literate is fundamentally tied to using scientific information, i.e., to making appropriate decisions and participating in public discussions about socio-scientific issues (SSIs) (Lederman et al., 2024). Subsequently, on the one hand, students should be able to re-construct already existing scientific arguments and, on the other hand, construct their own ones (Cavagnetto, 2010; Christodoulou & Grace, 2019; Osborne, 2002). At the same time, students' lack of scientific literacy in general (Lederman et al., 2024), and argumentation skills in particular (Osborne, 2002), is problematic. Osborne and colleagues (2016) demonstrate the validity of an overarching instrument for assessing argumentation skills. We adopt one of their major implications here, proposing that "[e]xercises that asked students to distinguish claims, warrants and data would provide a foundation for more demanding exercises which asked students to construct their own arguments" (p. 841). As Toulmin (2003) distinguishes between claims, warrants and data explicitly, we use Toulmin's Argumentation Pattern (TAP) in this study to understand upper secondary students' (ISCED 3) intuitive scientific argumentation in the chemistry classroom.

TAP In Chemistry Education

Following Toulmin (2003), arguments follow a structure based on up to six elements: data, warrants, backing of warrants, modal qualifiers, rebuttals, and the claim itself. TAP is widely used (Henderson et al., 2018) and has proved suitable for analysis of scientific arguments (Erduran et al., 2004; Lieber & Graulich, 2022). Although the idea of implementing TAP (e.g. Erduran et al., 2004; Lazarou & Erduran, 2021) or an adapted version (e.g. McNeill et al., 2006) in the classroom is not new in science education, studies are required that help to delineate a chemistry-specific approach at the secondary level (Christodoulou & Grace, 2019).

Above and beyond subject matters, SSIs are excellent vehicles to practise scientific argumentation (Christodoulou & Grace, 2019). Tackling SSIs via argumentation enables students to become responsible citizens capable of facing global challenges (Sadler, 2004). Argumentation, in terms of formally connecting premises to conclusions, is generisable. Yet, making an argument within SSIs requires domain-specific knowledge (Tricot, 2018). Therefore, our study uses the chemical side of human-induced ocean acidification, which is highly relevant when discussing climate change (IPCC, 2023).

Research Question

This contribution gives insights into a pilot study that poses the following research question, focussing on upper secondary students in chemistry education: How can the implicit implementation of TAP in a chemistry lesson support students in applying this pattern to improve the structural quality of their scientific arguments? In this case, ‘implicit’ refers to the introduction of TAP elements through scaffolds without explicitly naming or defining these elements, allowing students to engage with them intuitively.

Research Method And Design

The present study was designed as a pilot study that aimed to evaluate whether implementing TAP in the chemistry classroom can improve students’ scientific argumentation skills in the context of human-induced ocean acidification. Prior to an actual intervention, we evaluated how students re-constructed preset scientific arguments. 29 students with prior declarative knowledge about human-induced ocean acidification from 11th and 12th grade from a general high school in Austria participated in this study, taking place in a 50-minutes chemistry lesson. At first, a picture of a fictional newspaper headline claiming ‘calcification of coral skeletons is impaired’ was presented to the students.

Phase I: Students were first asked to individually formulate a well-founded written argument supporting the headline, based on their current knowledge about human-induced ocean acidification from the perspective of a climate activist.

Phase II: Afterwards, students were prepared with ‘argumentation cards’. These cards contained part of an argument, each corresponding to an element (e.g. data, warrants) of TAP, and enabled various thematic lines of reasoning (e.g., lime-carbonic acid cycle, pH, solubility of calcium carbonate in solutions of varying pH, solubility of carbon dioxide gas in water). Technically, these cards could be used to form up to seven different arguments to confirm the claim. Cards corresponding to the same element were coloured the same, helping students to recognise patterns between these elements. Apart from this structuring aid, no further information about these elements was given to students. The sole instruction provided to students was that the utilisation of these cards could be in any combination deemed suitable, including the use of more than one card of the same colour, or omitting the entire cards of one colour. In this phase, students worked in pairs discussing which cards were suitable to support their claim but did not formulate a written argument.

Phase III: Subsequently, students individually composed another written argument – this time using the previously selected argumentation cards – to support the headline regarding the impairment of coral skeletons calcification.

Phase IV: In a final phase, students were asked to formulate an additional written argument without any support materials, analogous to Phase I. However, due to time constraints, only a few students reached this phase.

Throughout all phases, students were provided with additional supporting material containing information about basic chemical concepts (e.g. chemical equilibrium, pH) relevant in this context. We chose this strategy to even out individual variance due to prior knowledge, allowing the students to focus on the argumentation process itself.

Analysis And Findings

Students' written arguments from phases I, III and IV were analysed regarding their structural quality through structural qualitative content analysis (Mayring, 2022). Content accuracy was not addressed in this contribution, as it was evaluated in a separate study (Dachauer & Lembens, in press). For the structural analysis, a deductive category system initially based on TAP was used. Due to challenges in assigning sequences of students' arguments to the categories *warrants* or *backing of warrants*, which required intense communicative validation (Steinke, 2004) between the raters - and even then, in some cases more than one categorisation would have been applicable – these two categories were merged into a single one labelled reasoning. This approach aligns with McNeill et al. (2006). As rebuttals were neither incorporated into the construction of the support materials nor observed in students' arguments, they were not included in the category system underlying the structural analysis. Based on the analysis of students' arguments and existing literature (McNeill et al., 2006; Toulmin, 2003), the coding system was further developed and serves as the foundation for a deductive coding approach in the planned follow-up study. This category systems comprises the following categories and their respective definitions:

- *Claim* refers to the statement whose content must be demonstrated and whose validity must be determined.
- *Data* refers to data and/or facts that serve as the basis for the argument and supports the claim.
- *Reasoning* refers to justifications, explanations, or elaborations that provide a basis for the claim in relation to the data. 'Reasoning' functions as a kind of bridge between the claim and the data. It can also reinforce the claim.
- The *Modal qualifier* provides information about the degree of certainty or binding force of the reasoning used to support the confirmation of the claim.
- *Other* refers to a sequence that cannot be assigned to any of the established categories.

During the process of assessing intercoder reliability, distinguishing between *data* and *reasoning* was occasionally challenging. This was due to the fact that some sequences were introduced with linking words such as *due to* or *because*, which could give the impression of aligning with *reasoning*. However, detailed content analysis revealed that these sequences often served as the foundation of an argument and should therefore be categorised as *data*. Two examples are given in Figure 1. Both examples demonstrate that the underlined linking words can introduce either *reasoning* or *data*, depending on the specific content they accompany.

During the communicative validation process between the two raters (Steinke, 2004), aimed at better defining and distinguishing between *data* and *reasoning*, the concept of the *'conceptual* *bridge'* emerged as a visual representation to illustrate the connection between *claim*, *data* and *reasoning*.

Figure 1. Distinction between data and reasoning in students' arguments (students' arguments from phase I, Pos. 5-6 and phase III, Pos. 78, translated from German).

The calcification of coral skeletons is impaired because, due to the intense exchange of CO_2 caused, for example, by the combustion of fossil fuels, carbon dioxide dissolves in seawater and reacts to carbonic acid. From this hydrogencarbonate and oxonium ions are produced in the water. Due to the increased concentration of H_3O^+ ions, the pH of the seawater increases. Although the pH remains in the basic range, it nevertheless harms the calcification of coral skeletons.

Because the pH in the ocean decreases and the CO_2 concentration in the water increases, the calcification of coral skeletons is impaired. This occurs because coral skeletons require solid $CaCO_3$ to form lime, and the decrease in pH increases the solubility of solid $CaCO_3$ in water.

Legend: data – reasoning

This *conceptual bridge* breaks down students' arguments into distinct components, focusing exclusively on *data*, *reasoning* and *claim*, while disregarding *modal qualifiers* and the category *other*. The student argument presented in Figure 2 serves as an example to illustrate the concept of the *conceptual bridge* in Figure 3. As demonstrated in Figure 2, the student's argument includes more than just the elements *data*, *reasoning* and *claim* – in this case, it also contains a modal qualifier. However, this modal qualifier is not relevant to the conceptual bridge.

Figure 2. Example of a student argument (Student argument from Phase III, Pos. 3-4, translated from German).

The pH decreases, as a result, it is certain that the calcification of coral skeletons is hindered. Coral skeletons require solid $CaCO_3$, but the decrease in pH increases the solubility of solid $CaCO_3$ in sea water, which leads to less solid $CaCO_3$.

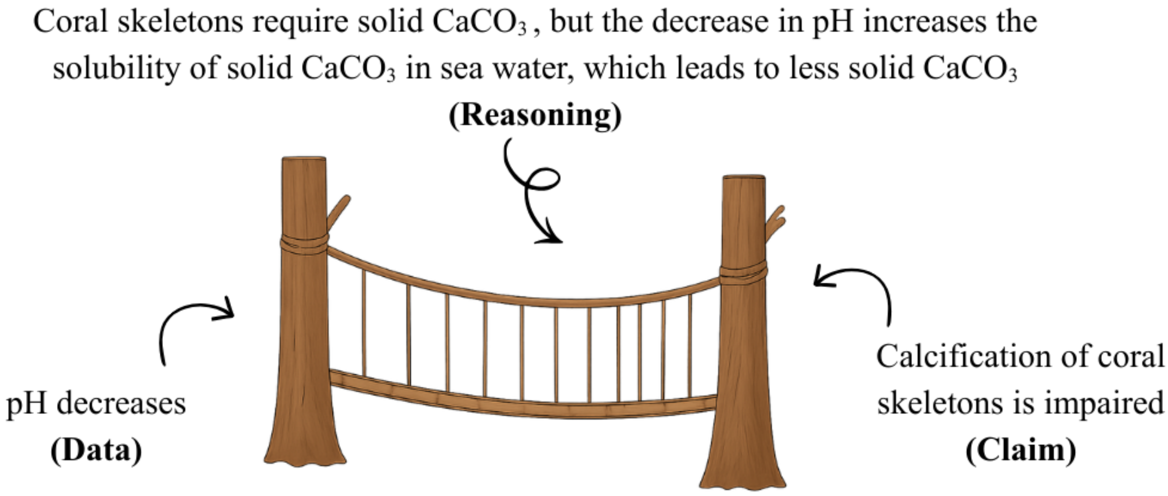
Legend: data – modal qualifier – claim - reasoning

In order to determine whether a sequence of a student's argument serves as either *data*, *reasoning* or *claim*, the *conceptual bridge* can be used as a tool for discussion between raters. Furthermore, it may also serve as an instructional instrument to help students actively distinguish between these elements in their own arguments. As illustrated in Figure 3, *data* and *claim* are visualised as bridge pillars, literally forming the foundation of the bridge – and, figuratively speaking, the foundation of the argument. *Reasoning* functions as the bridge itself, connecting the two pillars and enabling movement from one to the other. This visualisation emphasises the essential link between *claim* and *data* within the structure of scientific argumentation.

A total of three iterative coding cycles took place, each followed by communicative validation between the two raters. Table 1 illustrates the development of intercoder reliability across these three coding cycles.

To assess intercoder reliability, we began by coding students' arguments from phases III and IV – written after the intervention – under the assumption that these arguments would be easier to categorise. We used them to refine the definitions of each category and to add anchor examples to the coding system. Consequently, the category *other* did not exist during the first coding cycle for students' arguments in phases III and IV. This category was introduced during the first communicative validation and is therefore not referenced in the coding of students' arguments from phases III and IV coded in the first cycle.

Figure 3. Conceptual Bridge (Students' arguments from Phase III, Pos. 3-4, translated from German; Bridge generated with ChatGPT).



In addition, intercoder reliability increased significantly across all coding cycles of phase III. Between the first and second cycle, the visualisation of the *conceptual bridge* was created and initially discussed between the two raters. Between the second and third cycle, this visualisation was incorporated into the coding system, along with an anchor example. However, assessing intercoder reliability for students' arguments of phase I remained challenging, even with the visualisation of the *conceptual bridge*. During this process, intense communicative validation took place between the two raters, with each arguing for their own coding strategy, using the *conceptual bridge* visualisation. No agreement on coding modal qualifiers was reached during the first coding cycle of students' arguments of phase I. This can be attributed to the absence of relevant anchor examples needed to define this category. While one rater coded sequences such as *so*, *decisive* or *therefore*, the other only coded sequences only where the degree of certainty was explicitly stated, such as *therefore, it follows with certainty* or *presumably follows*. Consequently, more diverse anchor examples were added to the coding system, resulting in higher reliability in the subsequent coding cycles. Also, intercoder reliability increased significantly between the first and the second coding cycle of phase I, whereas this was not the case during the final intercoder cycle.

Table 1. Development of the intercoder reliability over three iterative cycles.

		Claim	Data	Reasoning	Modal qualifier	Other	κ
Phase I	1st	38.46 %	47.06 %	34.78 %	0 %	22.22 %	**0.22**
	2nd	83.64 %	90.57 %	67.61 %	68.97 %	100 %	**0.74**
	3rd	91.23 %	96.30 %	73.24 %	91.43 %	100 %	**0.84**
Phase III &	1st	89.66 %	45.71 %	23.33 %	66.67 %	-	**0.43**
	2nd	92.96 %	78.87 %	78.26 %	88.89 %	0 %	**0.80**
	3rd	98.59 %	91.43 %	82.86 %	98.18 %	100 %	**0.91**
Total		**95.31 %**	**93.55 %**	**78.01 %**	**95.56 %**	**100 %**	**0.88**

Our analysis of students' approaches to structuring their arguments revealed a broad use of TAP across all phases. However, the significance of these elements was generally low in students' arguments from phase I. While modal qualifiers appeared in all phases, their use became more nuanced after the intervention – for example, shifting from phrases like *so* or *this leads to* to *this* *probably leads to* or *this could lead to*. This enhancement is likely attributable to the provision of the pre-structured argumentation cards. Furthermore, the different elements in students' arguments were not seamlessly linked during phase I. After the intervention, however, the majority of students demonstrated more coherent argumentation. In particular, *data* and *claims* were more effectively connected through deeper reasoning. This progress also likely resulted from the use of pre-structured argumentation cards.

Conclusion And Outlook

We conclude that implementing TAP implicitly through teaching materials such as argumentation cards may lead to an intuitively appropriate use of these tools, thereby laying the foundation for a more explicit implementation of TAP later on. This scaffolded intervention could lead to "[e]xercises that asked students to distinguish [between] claims, warrants and data", thereby to re-construct existing arguments, and thus providing a "foundation for more demanding exercises which asked students to construct their own arguments" (Osborne et al., 2016, p. 841).

Based on our findings, implementing TAP already shows promise in improving students' scientific argumentation skills – even within our short-term intervention, despite limitations such as time constraints, a small number of participants, and a generally high socio-economic status. Therefore, we are going to scale up our follow-up study with a larger, more diverse group of students. To this end, teaching materials will first be developed through design-based research (Bakker, 2018), in close collaboration with both students and teachers to iteratively refine the materials. This will be followed by a two-phase intervention study ($n \approx 150$). From a structural point of view, phase one will implement a simplified version of TAP, focussing on Claim, Data and Reasoning (based on the CER model of McNeill et al., 2006), first implicitly, then explicitly. From a content point of view, phase one focuses on chemistry specific content knowledge. An interactive learning video introduces the topic of human-induced ocean acidification and its effects on coral skeletons. Following this, students are asked to formulate written arguments – later extending to other calcifying organisms as well. Phase two will deepen TAP instruction by extending the CER model to include modal qualifiers and rebuttals. The SSI of human-induced ocean acidification will serve as the central content. Through structured discussions, students strengthen their argumentation skills, reflecting on their actions, and formulating personal courses of action – also considering the broader relevance for humans and ecosystems beyond coastal regions. Given the need for more chemistry-specific research on scientific argumentation in SSI-related subjects at the secondary level (Christodoulou & Grace, 2019), we anticipate that this contribution will offer valuable insights and stimulate further work in this research area.

Acknowledgement

We thank the students and the school for their participation and support.

References

Bakker, A. (2018). Design Research in Education. A Practical Guide for Early Career Researchers. Routledge. https://doi.org/10.4324/9780203701010

Bianchi, G., Pisiotis, U., & Cabrera Giraldez, M. (2022). GreenComp. The European sustainability competence framework (Y. Punie & M. Bacigalupo, Eds.). Publications Office of the European Union. https://doi.org/10.2760/13286

Cavagnetto, A. R. (2010). Argument to Foster Scientific Literacy: A Review of Argument Interventions in K–12 Science Contexts. Review of Educational Research, 80(3), 336–371. https://doi.org/10.3102/0034654310376953

Christodoulou, A., & Grace, M. (2019). Argumentation within Societal Contexts in Chemistry Education. In S. Erduran (Ed.), Argumentation in Chemistry Education : Research, Policy and Practice (pp. 197–227). Royal Society of Chemistry.

Dachauer, J., & Lembens, A. (in press). Argumente, die Wellen schlagen: Ozeanversauerung im

Chemieunterricht. In P. Pollmeier (Ed.), Proceedings of the Annual Conference of the Gesellschaft für Didaktik der Chemie und Physik 2025: KI in der naturwissenschaftlichen Bildung (Vol. 46). Frankfurt am Main: Gesellschaft für Didaktik der Chemie und Physik.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. Science Education, 88(6), 915–933. https://doi.org/10.1002/sce.20012

Henderson, J. B., McNeill, K. L., González-Howard, M., Close, K., & Evans, M. (2018). Key challenges and future directions for educational research on scientific argumentation. Journal of Research in Science Teaching, 55(1), 5–18. https://doi.org/10.1002/tea.21412

IPCC. (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 35–115. https://doi.org/10.59327/ipcc/ar6-9789291691647

Lazarou, D., & Erduran, S. (2021). “Evaluate What I Was Taught, Not What You Expected Me to Know”: Evaluating Students’ Arguments Based on Science Teachers’ Adaptations to Toulmin’s Argument Pattern. Journal of Science Teacher Education, 32(3), 306-324. https://doi.org/10.1080/1046560X.2020.1820663

Lederman, J. S., Akerson, V., Bartels, S., & Schwartz, R. S. (2024). Attention science educators, we have a problem: lack of global functional scientific literacy [Comment]. International Journal of Science Education, 1–5. https://doi.org/10.1080/09500693.2024.2361934

Lieber, L., & Graulich, N. (2022). Investigating students’ argumentation when judging the plausibility of alternative reaction pathways in organic chemistry. Chemistry Education Research and Practice, 23(1), 38–54. https://doi.org/10.1039/D1RP00145K

Mayring, P. (2022). Qualitative Content Analysis: A Step-by-Step Guide (A. Owen, Ed.). SAGE Publications.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. Journal of the Learning Sciences, 15(2), 153–191. https://doi.org/10.1207/s15327809jls1502_1

OECD. (2019). PISA 2018 Assessment and Analytical Framework. 97–117. https://doi.org/10.1787/f30da688-en

Osborne, J. (2002). Science Without Literacy: A ship without a sail? Cambridge Journal of Education, 32(2), 203–218. https://doi.org/10.1080/03057640220147559

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S.-Y. (2016). The development and validation of a learning progression for argumentation in science. Journal of Research in Science Teaching, 53(6), 821–846. https://doi.org/10.1002/tea.21316

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

Steinke, I. (2004). Quality Criteria in Qualitative Research. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), A Companion to Qualitative Research (pp. 184–190). Rowohlt.

Toulmin, S. E. (2003). The Uses of Argument (2 ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511840005

Tricot, A. (2018). Scientific Reasoning as Domain-Specific or General Knowledge. A Discussion. In F. Fischer, C. A. Chinn, K. Engelmann, & J. Osborne (Eds.), Scientific Reasoning and Argumentation. The Roles of Domain-Specific and Domain-General Knowledge (1st Edition ed., pp. 271–277). Routledge. https://doi.org/10.4324/9780203731826

EUROPEAN SCIENCE EDUCATION RESEARCH ASSOCIATION ESERA 2025 Copenhagen

Explaining Facts About Electromagnetic Radiation In Podcasts: Students As Science Communicators

Hannah Loidl[1,2], Christoph Kulgemeyer[2] and Thomas Plotz[1]
[1]KPH Vienna/Lower Austria, Vienna, Austria
[2]University of Bremen, Bremen, Germany

Although students often construct explanations in class, they receive limited instruction in communication skills needed to articulate them effectively. In this project secondary school students (N=130-140, age 12-13) were put into the role of science communicators to produce science podcasts on topics of their choosing in electromagnetic radiation (EMR). Two test groups received different introductions prior to the production of podcasts: one on EMR content knowledge (CK) and the other on principles of good explanations. Assessing the depth and type of explanations revealed that students predominantly provide superficial and descriptive explanations. Interestingly, elevated CK did not result in higher-quality explanations. As numerous non-physics explanations were found, we analysed the contexts of the explanations in more detail. Two distinct genres of podcasts emerged: Nature of Science (NOS) podcasts (N=12), which embedded phenomena in a broader scientific context, and science podcasts (N=22), which provided science-based explanations of a phenomenon. We conclude that developing explanatory competence requires explicit instruction in communication skills rather than content knowledge alone. Podcasts offer a promising medium for fostering science communication competence and deepening students' engagement with NOS concepts.

Keywords: Science Communication Competence, Explaining, Podcast

Science Communication And Science Education

Fostering positive attitudes towards science as well as the construction of scientific knowledge are primary goals of both science communication research and science education research. However, the disciplines developed in separate directions. Thus, some scholars advocated for the rapprochement of the two disciplines (e.g., Baram-Tsabari & Osborne, 2015). Kohen and Dori (2019) therefore conducted a literature review to identify similarities and disparities between these "sister disciplines" (p. 526). A key finding concerns the different orientations of the two communities: "the science communication community mostly considers the process of communicating science to the public, whereas the science education community by and large focuses on the product—the knowledge and understanding of science that students gain as a result of learning and engaging with science" (Kohen & Dori, 2019, p. 545). By *process*, the authors refer to the way science content is communicated to the target audience, i.e., through various communication channels. While science communicators prioritize engagement and outreach, science educators (bound to formal settings and regulations) prioritize conceptual understanding and knowledge about science.

Our project seeks to bridge this gap by designing an intervention that targets both objectives: fostering *science engagement* through active participation in authentic communication practices while simultaneously developing students' *scientific knowledge* and *explanatory competence*. Specifically, we implement this approach through a workshop where students are put into the role of science communicators and create podcasts focused on Electromagnetic Radiation (EMR). In the next section, we briefly discuss the relevance of communication competence for science education and present the theoretical framework for the evaluation of the students' explanations.

The Role Of Communication Competence In Science Education

Through communication we make sense of the world, both through passive communication when i.e. obtaining information from written text or listening to explanations from others and through active communication when constructing our own explanations of observed phenomena. Furthermore, the relevance of communication competence becomes clear when examining key science education frameworks. For instance, communication competence is a core element of PISA, as scientifically literate students should be able to "explain phenomena scientifically", and "research, evaluate and use scientific information for decision making and action" (OECD, 2023, p. 9). Similarly, *constructing explanations* and *communicating information* are two out of the seven science practices included in the Next Generation Science Standards (NGSS, 2013). Moreover, explanatory and communication competencies are embedded in several school curricula worldwide, including those of Austria and Germany.

As a result, Kulgemeyer and Schecker (2013b) developed a constructivist model of communication competence. In this model, they distinguish between the subject adequacy and the addressee adequacy of a communication product. Communication between addressee and communicator is considered successful if the communication product is both addressee-adequate and subject-adequate. An addressee-adequate explanation considers the addressee's prior knowledge and interests, whereas subject adequacy refers to the correctness of the communicated information.

To analyse the subject adequacy of the students' explanations in this study, we examine the explanatory type and the explanatory depth. The explanatory type is linked to philosophical accounts of explanation described in detail by Braaten and Windschitl (2011). For this study, we adopted and slightly adapted their three levels of explanation: a) *descriptive explanations* which answer questions about the what of a phenomenon and describe it on the basis of patterns and surface features, b) *process-oriented explanations*, which answer questions about the how of a phenomena by invoking causes and simple if-then relationships, and c) *justifying explanations*, which answer questions about the why of a phenomenon by embedding it within the larger framework and drawing on theoretical concepts and laws. Because these three explanatory types do not capture the factual correctness of an explanation, we additionally examined its explanatory depth.

Research Gap And Research Question

Although the construction of explanations is crucial for the development of students' scientific literacy, the research field is still emerging. There are few existing studies, i.e., Helms (2017) developed a method to improve students' explanations to teach their peers, de Andrade et al. (2019) analysed students' written explanations in the context of chemistry, Alameh et al. (2023) compared explanations of students, teachers and experts and Kulgemeyer and Schecker (2013a) filmed students while explaining physics contents during a roleplay. They analysed the filmed explanations through a qualitative content analysis, focusing on the addressee-adequateness. The more beneficial categories an explanation counted, the better its quality. In this study, we aim to look deeper into the subject-adequacy of students' explanations to answer to what extent the CK affects the subject-adequacy of students' explanations.

In previous studies, students had to explain a certain phenomenon in a fixed setting. It remains unknown how students construct science explanations when put in the role of science communicators and choosing their own topics. We further aim to discover how the students' Content Knowledge (CK) and Communication Competence (CC)[1] affect the subject-adequacy of

[1] This influence of the CC on the explanations is not presented in this paper as the analysis is still in process.

the students' explanations and what contexts students choose for their podcasts as science communicators. The following questions guide our analysis:

RQ 1: Does elevated content knowledge result in increased levels of subject-adequacy of the students' explanations?

RQ 2: What contexts do students choose to communicate their physics knowledge in podcasts and how does the context affect the level of subject-adequacy of the students' explanations?

Research Design

The presented study is embedded in the bigger project *We talk about Science*. In this project, students are put into the role of science communicators and are tasked with the creation of science communication products. The products are presented during science communication events for the school community. To measure influences of the CK and the CC, we used two test groups and one control group. One test group received a communication training (CT group) comprising two school lessons in advance. The training consisted of theoretical knowledge on good explanations and various tasks to practise explaining. The other test group received four physics lessons on EMR prior to the Science Communication Workshops. In this article, we present the results of the produced podcasts with a focus of the students' CK and the podcasts' contexts. The study took place in the summer semesters of 2024 and 2025.

The Science Communication Workshops

The workshops in the second cycle in 2025 consisted of two days each comprising of three school lessons (each 50'). On the first day, students were briefly introduced to the goals of science communication (i.e., inform, entertain) and most important aspects to be considered when producing science communication products (i.e., choosing the right medium, tailoring the product to the target group). Afterwards, we presented the subject EMR to the students. This topic was chosen as we knew that the participating students had not received any training on EMR prior the intervention. Additionally, a teaching concept has been developed recently and evaluated for secondary school students by Zloklikovits (2022) in Vienna. Based on this concept, we introduced the students to some key ideas guided by a prepared power point presentation and interrupted by experiments (i.e., feeling versus seeing IR-radiation, observing the effect of UV-radiation on UV-sensible beads). Afterwards, we asked students how the new acquired knowledge should be explained to their peers. We introduced them to an explanation cycle (see Malotki et al., 2026) that was developed for the communication training. Subsequently, the students performed a paper pencil test (~20') to inquire about their CC and CK on EMR. The first day ended with the students choosing a topic in the field of EMR for their podcasts. We supported the selection process with a collection of links to information texts on the browser-based application *taskcards*. On a pinboard, we arranged the links according to the type of radiation discussed in the texts. The texts ranged from science communication texts from popular German speaking journals (i.e., *Spektrum der Wissenschaft*) to information texts from the *German Federal Office for Radiation Protection*, and physics texts especially for students (i.e., *www.leifiphysik.de*). In groups, the students planned their podcasts with the help of printed storyboards. On the storyboards, they focused on three to five main research questions, chose the genre of their podcast (i.e., interview or discussion format), and planned their explanations. The next day started with the students writing scripts on printed worksheets (~ 50'). In the next 50 minutes, students recorded their podcasts with their mobile phones or tablets using the App C*apCut* (free version). In the last 50 minutes, students edited and handed in their products. The workshop ended with listening and appreciating the students' podcasts. Also, the products were uploaded to our project website. This way, students and teachers could share the podcasts with the school community, family and friends.

In contrast to cycle II as described above, cycle I in 2024 provided maximum freedom for the students. The first day, we briefly introduced the students to the five key ideas of EMR based on the work of Plotz (2017, p. 5). Afterwards, the students posed three questions concerning EMR on their own and then chose one question in groups of three. The students did research on the internet to answer their chosen question. Using a storyboard and worksheets for preparing their scripts (similar to cycle II), they planned their science communication products in their preferred format: podcast, video or Instagram post. On the second day, students created their products. There was no exact timeline for each step of the production process and the teacher as well as the project team only assisted when help was required. The podcasts from cycle I will be examined together as one group (Z1).

Participants

In this study secondary students from the age of 12 to 13 participated (N = 130-140 students from six different classes). We chose middle school classes, as this is the most representative student population in Austria. One of the participating schools is located in Vienna and two in rural areas of Austria (Styria and Lower Austria). In contrast to the two schools in Styria and Lower Austria, the school in Vienna is a private school with a greater percentage of students with migration background. Next to the middle school in Upper Austria, there is a gymnasium–high achieving students are therefore more likely to attend the gymnasium next to the middle school. This information is important, as the location and separation of high-achieving and low-achieving students probably influenced the quality of the learning products.

Research Method

For this study we chose a mixed-methods design to analyse the quality of the students' podcasts and influences of the students' CK and CC. The students' podcasts were analysed by applying a structural thematic content analysis according to Kuckartz and Rädiker (2022). All podcasts (N = 34; duration: approx. one to five minutes) were transcribed and then uploaded to MAXQDA 2024. After transcribing the podcasts, each text was read and structured in explanation sequences. One explanation sequence (usually) equals the proposition of one student. Next, the type of each explanation sequence was determined using the deductive subcodes: descriptive, process-oriented, and justifying explanations. Subsequently, the depth of each student explanation was examined. We distinguished the four subcategories non-physics, insufficient, superficial and concept-oriented explanations. As this distinction resulted in a large number of non-physics explanations, we examined the context of each explanation in more detail and categorised them as dealing with physics, astronomy, history of physics, biology/medicine or Nature of Science (NOS). To have a better impression of the students' associations with science, we analysed NOS aspects that students addressed in their explanations. For this purpose, we used the already existing category system of Sattelkau (2021). She derived the NOS categories from diverse NOS concepts (i.e., Family Resemblance Approach, Whole Science) and reflections on science communication. In our analysis of the explanatory context of each explanation sequence and the number of NOS dimensions, three types of podcasts emerged: Physics Podcast, Biological Podcast and NOS Podcast.

To compare the different groups (CK group, CT group, Control group, and Z1 group) and the different types of podcasts, we calculated scores of the explanatory depth and explanatory type. Each proposition was assigned with one to three points, according to Table 1. The scores were calculated by dividing the sum of the points in depth and type through the number of explanation sequences in the podcast. For example, a podcast containing 4 propositions with three descriptive explanations and one justifying explanation results in six points divided by four, achieving a Type Score of 1.5. Similarly, the depth score of each podcast was calculated. To compare the groups, we used quantitative tests such as ANOVA and Kruskal-Wallis-Tests. Furthermore, we compared the explanatory type and depth of the different podcast types using the Kruskal-Wallis-Test. Finally, the CK test on EMR was evaluated to compare the content knowledge of the three participating groups in the second cycle by running an ANOVA. To measure influences of the students' CK on the explanatory type and depth, we calculated correlations.

Table 1. Distribution of points according to the applied code.

	0 points	1 point	2 points	3 points
Explanatory Depth	Non-physics	Insufficient	Superficial	Concept-oriented
Explanatory Type	-	Descriptive	Process-oriented	Justifying

Results

Analysis Of The Explanations' Subject-Adequacy

We measured the level of the subject-adequacy by examining the type and depth of each explanation in the podcasts. As one can see in Table 2, most assertions were coded as descriptive explanations (n = 128). Those assertions deliver facts or describe observable phenomena, usually answering questions about the "what of a phenomenon". To a lesser extent, assertions were coded as process-oriented (n = 31) or justifying (n = 32). Process-oriented explanations answer questions to the "how of a phenomenon" and include i.e. if-then-clauses, showing simple relations between descriptions ("what") and their trigger ("how"). Justifying explanations answer questions of the "why of a phenomenon" and include causes, theoretical concepts and laws, usually signalled by the word "because". To compare the podcasts of the different groups, a type score of each podcast was calculated as described in the methods section. Running a Kruskal-Wallis Test, no difference was found between the four groups. Neither the communication training nor additional lessons on the concept of EMR seem to influence the explanatory type score.

For determining the depth of explanations, the inclusion of physic concepts, theory and laws as well as the correctness of the assertions was considered. At level zero, there are *non-physics explanations*, which address a physics phenomenon but do not provide a physics-based explanation (i.e., the history of the microwave oven and biological, medical subjects). At the next level are i*nsufficient explanations*, which present a physics perspective but are based on incorrect facts. The next level comprises *superficial explanations*, which include mostly correct facts but do not connect them to underlying physics concepts. In contrast, *concept-oriented* explanations contain physics concepts, draw conclusions, and rely on correct information only. In Table 2, one can find the distribution of codes concerning the explanatory depth. Calculating the depth score and comparing the measures of the groups did not show any differences when computing the Kruskal-Wallis Test, $\chi^2(3) = 3.764$, $p < .288$, $\varepsilon^2 = .114$. However, the CK group shows a slightly higher mean depth score when compared to the other groups. Especially when looking at Dunn's post hoc comparisons, the highest difference was found between the CK and Z1 group ($p = .060$). This might indicate that the additional lessons on EMR before the workshops had some effect on the explanatory depth. Moreover, the Z1 group that received the least information on EMR and the least guidance during the workshop achieved the lowest mean depth score of all groups, however we found no significant differences.

Influences Of The Students' Content Knowledge On The Explanations' Subject-Adequacy

In cycle II, the students took a brief paper-pencil test on EMR to measure influences of their CK on the explanation quality in their podcasts. After verifying the assumptions, we conducted an

ANOVA which revealed significant differences between the groups in their CK ($F(2, 68) = 10.02$, $p < .001$, $\eta^2 = .23$). Post hoc comparisons using Tukey's test showed that the CK group did not differ from the control group, $t(68) = .677$, $p = .777$, $d = .198$. However, we found significant

Table 2. Distribution of subcodes in the categories type and depth of explanations.

Type of Explanations

	Z1	C	CK	CT	Total
Descriptive	45	33	29	21	128
Process-oriented	8	6	8	9	31
Justifying	13	6	6	7	32
SUM	66	45	43	37	191
N = Documents	12	8	7	7	34

Depth of Explanations

	Z1	C	CK	CT	Total
Non-physics	30	17	9	9	65
Insufficient	7	0	7	9	23
Superficial	22	23	24	16	85
Concept-oriented	7	5	3	3	18
SUM	66	45	43	37	191
N = Documents	12	8	7	7	34

differences between the CK and CT group, $t(68) = 4.181$, $p < .001$, $d = 1.207$ and the control and CT group, $t(68) = 3.458$, $p < .003$, $d = 1.009$. Although the mean test score of the CK and control group was significantly higher than the test score of the CT group, comparing the depth scores in the section above, those groups did not present explanations in more depth. This is also evident when examining the correlations between one group's content knowledge and their explanatory type score, as we found no correlation between students' mean test scores and explanatory type or depth.

Context Of The Explanations

During the analysis of the explanations' depth, many non-physics explanations emerged. Nevertheless, the podcasts clearly referred to the field of science. Therefore, we added the main category NOS dimensions to explore which aspects of science the students discussed. As a first step, we decided for each proposition if a phenomenon was explained from a physics, astronomy, history of physics or biological/medical perspective. Explanations that did not fit into any of those categories were coded as NOS. Next, we analysed all explanations regarding NOS dimensions using the coding manual of Sattelkau (2021). An overview of found NOS dimensions in the podcasts and their adapted and summarized descriptions can be found in The Explanations' Subject-Adequacy In The Different Genres Of Podcasts

Given the non-normal distribution of the data, a Kruskal–Wallis test was used to compare the depth scores and type scores of the different genres of podcasts. The test showed that the three genres differed significantly in terms of the explanatory depth $\chi2(2) = 14.333$, $p < .001$, $\varepsilon2 = .434$. In Dunn's post hoc comparisons, Physics Podcasts achieved significantly higher depth scores than NOS Podcasts ($p < .001$, $rrb = .850$) and Biology/Medicine Podcasts ($p = .002$, $rrb = .800$). As expected, there were no significant differences comparing the types of explanations between the three genres, $\chi2(2) = 1.971$, $p = .373$, $\varepsilon2 = .060$. However, this result is not surprising, as NOS Podcasts and Biology/Medicine Podcasts count a higher number of non-physics explanations which are scored with zero points.

Table 3.

By comparing the number of applied codes in each podcast two main genres of podcasts emerged: Science Podcast (n = 22) and NOS Podcast (n = 12). Podcasts coded as Science Podcasts mainly focus on scientific explanations, either from a physics or biological/medical perspective on selected phenomena of EMR. Therefore, we distinguished between two subgenres depending on the referred science discipline: Physics Podcasts (n = 10) or Biology/Medicine Podcasts (n = 12). Physics Podcasts not only included science explanations but also the two dimensions NOS

Products and NOS *Production*. For instance, one group described how the sunlight can be used for energy production through *solar panels* (= subcode *Applications* in the dimension NOS *Products*). In Biology/Medicine Podcasts, we only discovered the NOS dimension *Products* with two subcodes *Applications* and *Risks and Benefits*. Several Biology/Medicine Podcasts presented the application of X-ray imaging, highlighting its benefits for medical examinations, as well as its risks. In contrast, NOS Podcasts displayed a variety of NOS dimensions, for example, they name different nationalities of scientists, describe collaborations and typical tasks of scientists (NOS *Actors*) and name different institutions around the world (NOS *Players*). In the main NOS Podcasts category, we discovered the subgenre History of Physics Podcasts. For instance, one group described the *coincidental* invention of the microwave oven, while another podcast tells the story of the *discovery* of infrared light in 1800 (NOS *Production*). Furthermore, we found evidence in the NOS Podcasts that, alongside the development of *Applications,* one goal of science is the acquisition of *Knowledge* (NOS *Products*).

The Explanations' Subject-Adequacy In The Different Genres Of Podcasts

Given the non-normal distribution of the data, a Kruskal–Wallis test was used to compare the depth scores and type scores of the different genres of podcasts. The test showed that the three genres differed significantly in terms of the explanatory depth $\chi^2(2) = 14.333$, $p < .001$, $\varepsilon^2 = .434$. In Dunn's post hoc comparisons, Physics Podcasts achieved significantly higher depth scores than NOS Podcasts ($p < .001$, $r_{rb} = .850$) and Biology/Medicine Podcasts ($p = .002$, $r_{rb} = .800$). As expected, there were no significant differences comparing the types of explanations between the three genres, $\chi^2(2) = 1.971$, $p = .373$, $\varepsilon^2 = .060$. However, this result is not surprising, as NOS Podcasts and Biology/Medicine Podcasts count a higher number of non-physics explanations which are scored with zero points.

Table 3. Short descriptions of the NOS dimensions, adopted and adapted according to Sattelkau (2021). Specific examples found in the products are given in brackets.

NOS Dimensions	Products (n=71)	Knowledge (n=6)	One intention is the generation of knowledge, differentiation between models, theories and laws (i.e., finding out the origin of the universe)
		Applications (n=65)	Specific applications or technologies, optimization of applications and visions of future applications; Risks and benefits for humans and the environment (i.e., risks of x-ray imaging)
	Players (n=31)	Politics (n=1)	Diverse political actors, Scientific findings inform political decisions, politics can influence the direction of scientific research, i.e. through financing (i.e., Federal Office for Radiation Protection)
		Society (n=10)	Influences of society's needs, development through science, science communication, trust and mistrust (i.e., what radiation types are dangerous, radiation protection)
		Science (n=19)	Institutions, Disciplines, Financing of Science and Lobbying (i.e., scientific facilities around the world, collaborations in large science projects)
		Industry and Economy (n=1)	Influences of industry and economy on scientific projects, input and output of science for industry (i.e., invention of the microwave in the laboratory of an armaments industry)
	Actors (n=40)	Personnel Diversity (n=19)	Scientists have various cultural, religious, gender backgrounds and nationalities, different professions (i.e., physicists)
		Activities (n=15)	Scientists produce knowledge through observing, examining, measuring, … name substances, planets etc.
		Characteristics (n=2)	Scientists have different personalities, typical and atypical characters (i.e., curiosity), they experience success and failure, have vast expertise
		Social Interactions (n=4)	Scientists work together, discuss findings in (inter-)national conferences, use peer reviews, award prizes for outstanding work or name i.e. methods after scientists (i.e., Planck Telescope)
	Production (n=24)	Methods, mechanisms and practices (n=18)	Nature of scientific inquiry, different scientific methods, empirical evidence, formulation and testing of hypotheses (i.e., drawing on empirical evidences, presenting different scientific methods)
		Development and Time (n=6)	Scientific findings take time, there are highs and lows, errors and irregularities (i.e., coincidental invention of microwave oven, year dates)

Discussion

Students Mostly Provided Descriptive And Superficial Explanations

Although children are usually interested in more detailed information when confronted with explanations (*Why is that so?*), they did not show the same inquisitiveness about their own explanations in the analysed podcasts. Students mostly offered *descriptive explanations* and rarely provided *justifying* or *process-oriented explanations*. However, in the second cycle, we explained that in science, a claim is always followed by justifications. This was also one of the points highlighted in the communication training. Nevertheless, the CT group did not provide more justifications than the other groups. We do not know whether students are satisfied with the quality of their explanations or if they simply were overwhelmed by all the steps they had to take in our workshop. It is also possible that students lacked detailed knowledge of the phenomenon to be explained. Note that students did not show signs of time pressure during the research phase.

Furthermore, all groups tended to provide predominantly *superficial explanations*. We anticipated that the CK group might provide explanations in more depth, since they could rely on key ideas and experiences from the lessons on EMR before the workshop. However, the CK group did not provide a higher number of *concept-oriented explanations* than the other groups. This might indicate that students need some kind of guidance to connect prior knowledge to new information gathered from research. Additionally, Schecker and Höttecke (2021) report that in students' explanations claims often lack the connection to physics principles. To guide students in this process, Tschentscher and Berger (2016, as cited in Schecker & Höttecke, 2021, p. 500) developed the methodological tool "Explanatory Chain" (Erklärkette) to support students in formulating subject-adequate explanations.

Interestingly, only about 12 % (n = 23) of the propositions were coded as insufficient, which means that explanations based on wrong facts (i.e., humans cannot see but feel UV radiation, or gamma radiation is radioactive) were not predominant in the student-created podcasts. Nevertheless, we found a large number of explanations that were not physics related. This raises the question of whether students avoid intensive engagement with physics topics, preferring to talk about related sciences (i.e., biology or medicine) rather than explain phenomena from a physics perspective.

Influences Of The Content Knowledge On The Quality Of Explanations

In this study, we did not find statistical evidence that higher levels of CK resulted in higher quality of explanations. This was measured by evaluating the subject-adequacy in terms of the type and depth of explanations. The mean scores of the groups on the CK-test neither influenced the depth score nor the type score. This might indicate that it does not suffice to provide students with information about a certain topic to provide high quality explanations. This is in line with previous studies on CK and its influences on students' communication skills (Kulgemeyer & Schecker, 2013b). Similar results were found for teachers, needing a certain level of pedagogical content knowledge (PCK) to use their CK for delivering high-quality explanations (Kulgemeyer & Riese, 2018). Perhaps students too must develop communication competence skills in science education in order to provide high-quality explanations. Interestingly, students are usually tasked with the formulation of explanations during science lessons or in exams without being prepared with explanatory skills. As part of this study, we therefore developed and evaluated a communication training that focused on the addressee-adequacy of explanations. So far, the CK group has not shown better results in terms of the subject-adequacy of explanations. However, the results regarding the influences of the CC on the subject-adequacy as well as influences of the communication training on the addressee-adequacy are still being processed.

Three Genres Of Students' Podcasts: Physics, Biology/Medicine, Nature Of Science

A deeper analysis of propositions coded with the subcategory non-physics opened a new perspective on the student-produced podcasts. We found two main genres of podcasts, Science Podcasts with the focus of explaining phenomena scientifically (at an age-adequate level) and NOS Podcasts that embed a topic in the field of EMR in a broader picture and address various aspects of science. In this category, we found the subgroup History of Science Podcasts. This is especially interesting, as one propagated way of teaching NOS is through the History of Science approach. Regarding the distribution of genres in the different groups, most NOS Podcasts were found in cycle II (n = 10) in the control group (n = 5). This might be influenced by the fact that we offered them a diverse range of topics concerning EMR on *taskcards*. However, in the end the students chose the topic by themselves. Furthermore, ten out of twelve Physics Podcasts were also created in cycle II. One might anticipate that most of the Physics Podcasts were produced by the CK group, however, in each group of cycle II we found three or four Physics Podcasts. Interestingly, eight out of twelve Biology/Medical Podcasts were created in cycle I. Apparently, students avoid deeper engagement with a chosen topic from a physics perspective, especially if they are not confronted with pre-selected topics/texts.

Limitations

Please consider that there are several limitations in this study:

- The correlations between the CK and the explanation quality in the first cycle is based on a small number of students (N = 71). The analysis of the explanatory type and depth score is based on both cycles with a sum of n = 34 podcasts. However, this is an exploratory study with a focus on the development of teaching materials.
- The findings reported are preliminary, as this is part of the ongoing dissertation project of the first author. Intercoder agreement on the coded data is planned for summer 2026.
- The groups included are not easily comparable, as the students have different socioeconomic backgrounds. Additionally, two of the participating middle schools are located next to a gymnasium. Consequently, high-achieving students are more likely to attend the gymnasium. In contrast, there is no gymnasium near to the third middle school (control group). This may explain why the control group performed better in the CK test. The choice of different schools was made due to the project's focus on science communication and the goal of greater outreach.
- As the podcasts are created in groups, it was not possible to compute correlations between individual CK scores and their skills in providing subject-adequate explanations.
- The used CK test on EMR is based on previous work of Zloklikovits (2024). The adapted and used items were not evaluated in advance. Also, the test only consisted of four items due to time restrictions.

Conclusion

Although we established considerable high standards for the students, they delivered commendable products in a short period of time and demonstrated strong engagement during the podcast production on the second day of the workshops. Incorporating science communication practices, particularly divers communication channels and attention to *communication processes*, appear to be effective approaches to engage students with science. While approximately a third of the created podcasts focused on physics explanations to present a chosen topic in the field of EMR another third of the podcasts focused on facts *about* science. This might indicate that the production of podcasts might be a beneficial way to teach NOS, especially when addressing NOS explicitly. However, aiming at the production of Physics Podcasts, students need more guidance in obtaining information and translating gathered information into valuable science explanations. Formulating subject-adequate explanations is a skill that must be learned. One way of supporting this process is through the explanatory chain developed by Helms (2017). The analysis of the explanations' addressee-adequacy and influences of the developed communication training on the students' explanations are still ongoing. So far, this article presented a useful way of analysing the subject-adequacy in terms of type and depth of explanations. Those categories along with characteristics of the addressee-adequacy will be part of a diagnostics tool for analysing students' explanations. To conclude, we think that next to a certain level of CK, students should be prepared with explanatory skills in science education as are teachers with PCK. By strengthening students' communication skills, we enable them to express and share their knowledge effectively.

Acknowledgement

The authors gratefully acknowledge the use of DeepL and Perplexity AI for language support during manuscript preparation. This work was conducted within the project We talk about Science, a Sparkling Science Project funded by OeAD and the Austrian Federal Ministry of Women, Science and Research.

References

Alameh, S., Abd-El-Khalick, F., & Brown, D. (2023). The Nature of Scientific Explanation: Examining the perceptions of the nature, quality, and "goodness" of explanation among college students, science teachers, and scientists. *Journal of Research in Science Teaching*, *60*(1), 100–135. https://doi.org/10.1002/tea.21792

Baram-Tsabari, A., & Osborne, J. (2015). Bridging science education and science communication research. *Journal of Research in Science Teaching*, *52*(2), 135–144. https://doi.org/10.1002/tea.21202

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, *95*(4), 639–669. https://doi.org/10.1002/sce.20449

De Andrade, V., Freire, S., & Baptista, M. (2019). Constructing Scientific Explanations: A System of Analysis for Students' Explanations. *Research in Science Education*, *49*(3), 787–807. https://doi.org/10.1007/s11165-017-9648-9

Helms, C. (2017). *Entwicklung und Evaluation eines Trainings zur Verbesserung der Erklärqualität von Schülerinnen und Schülern im Gruppenpuzzle*. Logos Verlag.

Kohen, Z., & Dori, Y. J. (2019). Toward narrowing the gap between science communication and science education disciplines. *Review of Education*, *7*(3), 525–566. https://doi.org/10.1002/rev3.3136

Kuckartz, U., & Rädiker, S. (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (5th edn). Beltz Juventa.

Kulgemeyer, C., & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching*, *55*(10), 1393–1418. https://doi.org/10.1002/tea.21457

Kulgemeyer, C., & Schecker, H. (2013a). Schülerinnen und Schüler erklären Physik—Modellierung, Diagnostik und Förderung von Kommunikationskompetenz im Physikunterricht. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. Vollmer (Eds.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (Vol. 3, pp. 225–240). Waxmann.

Kulgemeyer, C., & Schecker, H. (2013b). Students Explaining Science—Assessment of Science Communication Competence. *Research in Science Education*, *43*(6), 2235–2256. https://doi.org/10.1007/s11165-013-9354-1

Malotki, L., Berger, R., & Kulgemeyer, C. (2026). Ein Erklärtraining zur Steigerung der Erklärqualität und der Effektivität kooperativen Lernens. In P. Pollmeier (Ed.), *KI in der naturwissenschaftlichen Bildung.* (46. Tagungsband). GDCP.

OECD (Ed.). (2023). *PISA 2025 Science Framework (draft)*. https://pisa-framework.oecd.org/science-2025/deu_deu/

Plotz, T. (2017). Students' conceptions of radiation and what to do about them. *Physics Education*, *52*(1). https://doi.org/10.1088/1361-6552/52/1/014004

Sattelkau, C. (2021). *Wissenschaftskommunikation mit Texten: Die Kommunikation von Nature of Science mit Textgenres als Brücke zwischen Science Education und Science Communication* [Dissertation].

Christian-Albrechts-Universität zu Kiel.
Schecker, H., & Höttecke, D. (2021). Unterrichtskonzeptionen für die Förderung prozessbezogener Kompetenzen. In T. Wilhelm, H. Schecker, & M. Hopf (Eds), *Unterrichtskonzeptionen für den Physikunterricht: Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis* (pp. 475–519). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-63053-2
Zloklikovits, S. (2022). *Elektromagnetische Strahlung in der Sekundarstufe I unterrichten*. Universität Wien. https://phaidra.univie.ac.at/o:1594025
Zloklikovits, S. (2024). *Elektromagnetische Strahlung in der Sekundarstufe I unterrichten* [Dissertation]. Universität Wien.

EUROPEAN SCIENCE EDUCATION RESEARCH ASSOCIATION ESERA 2025 Copenhagen

Argument Construction System 'Argu-Made 2.0' To Support The Examination Of Evidence And Scientific Judgement Regarding The Bioaccumulation Of Pufferfish Toxins

Yuka Matsuyama[1], Keita Muratsu[2], Shinichi Kamiyama[3], Hideo Funaoi[4], Tomokazu Yamamoto[1]

[1]The Joint Graduate School in Science of School Education Hyogo University of Teacher Education, Japan

[2]Osaka Shin-Ai Gakuin University, Japan

[3]Osaka University of Health and Sport Sciences, Japan

[4]Soka University, Japan

Japanese primary school education is required to develop the ability to make scientific judgements about food safety and prevent food poisoning. For example, the consumption of pufferfish requires a scientific understanding and judgement of the bioaccumulation of pufferfish toxins. An argument consisting of 'claim – evidence – reasoning' is effective in using scientific understanding, and, in particular, the appropriate use of evidence is considered important. Therefore, this study developed an argument construction support system 'Argu-made 2.0', and conducted a class using this system and simulators by Matsuyama et al. (2023) promoting understanding of the bioaccumulation of pufferfish toxin, and the effectiveness of Argu-made 2.0 was evaluated. This class was conducted for sixth-grade food-chain studies. Students learned about the bioaccumulation of pufferfish toxins using simulators and constructed arguments while examining the evidence in Argu-made 2.0. The effectiveness of Argu-made 2.0 was evaluated from two perspectives: 'examination of evidence' and 'achievement of scientific judgement'. In 'examination of evidence', the process of using evidence appropriately in Argu-made 2.0 was analysed, while in 'achievement of scientific judgement,' the content of the students' arguments was analysed. The results showed that 48 [87%] students selected appropriate evidence, and 42 [76%] constructed arguments by correctly constructing all elements. These results indicate that Argu-made 2.0 was effective in supporting sixth-grade students in considering appropriate evidence and in making scientific judgements about the presence or absence of pufferfish toxins.

Keywords: Argument, Primary school, Argu-made 2.0

Introduction

Japanese primary school education is necessary to make effective scientific judgements about food behaviours, which is crucial for food safety especially preventing food poisoning. However, not only in Japan, recent studies have indicated that many consumers make incorrect judgements about food safety due to limited knowledge and awareness (e.g., Sharif & Al-Malki, 2010). There is a growing need for teaching methods that enable primary school students to learn to handle food safely and make informed judgements using this knowledge.

One example of scientific judgement regarding food safety is food poisoning by pufferfish toxins. In this case, a scientific understanding of the bioaccumulation of pufferfish toxins and a scientific judgement regarding the safety of consuming pufferfish by applying this understanding are necessary (e.g., Malloggi et al., 2023). Students must understand how the concentration process works to gain a scientific understanding of pufferfish toxin bioaccumulation. The simulator developed by Matsuyama et al. (2023) to understand the bioaccumulation of pufferfish toxins was effective in promoting this process. Developing the ability to form scientifically valid ideas (claims) is key to making scientific judgements. Arguments structured as 'claim (answer to a question) - evidence (experimental results) - reasoning (scientific principle)' have proven effective (e.g., Kamiyama et al, 2015; Iordanou et al, 2019). In particular, since the use of appropriate evidence provides higher scientific credibility and validity to one's claims, it is important to consider the appropriateness of evidence in terms of it being obtained for proper purposes and methods and it being linked to one's claims while being based on theoretical support (scientific conventions) (Duncan et al., 2018).

Matsuyama et al. (2024a) developed an argument construction support system, 'Argu-made', to help users share, compare, and examine evidence online to promote the appropriate use of evidence. Matsuyama et al. (2024b) developed and evaluated a teaching method to encourage scientific judgement on the subject of bioaccumulation of pufferfish toxins, using a simulator developed by Matsuyama et al. (2023) in combination with arguments consisting of 'claim-evidence-reasoning'. The results showed that some students did not appropriately use their knowledge of the mechanism of bioaccumulation in their scientific judgements, indicating the need for improved instruction on the structure of evidence and reasoning.

Therefore, this study focused on examining this evidence, which was one of the challenges identified by Matsuyama et al. (2024b). To improve evidence examination, support is needed to allow opinions on the appropriateness of evidence to be shared and compared online and reconsidered from multiple perspectives. Therefore, we improved Argu-made by Matsuyama et al. (2024a) and developed an argument construction support system called 'Argu-made 2.0' that has the functions to realise these improvements. Unlike traditional paper media, this system enables asynchronous commenting based on individual progress, improving learner participation. Another feature is that past opinions can be confirmed and revised based on the comment history. This study aimed to develop and implement an instructional method using the simulator by Matsuyama et al. (2023) and Argu-made 2.0, and to analyse the effectiveness of this method by focusing on the process of examination of evidence by students and the achievement level of scientific judgement.

Research Question

This study verified the effectiveness of Argu-made 2.0. The two research questions are as follows:

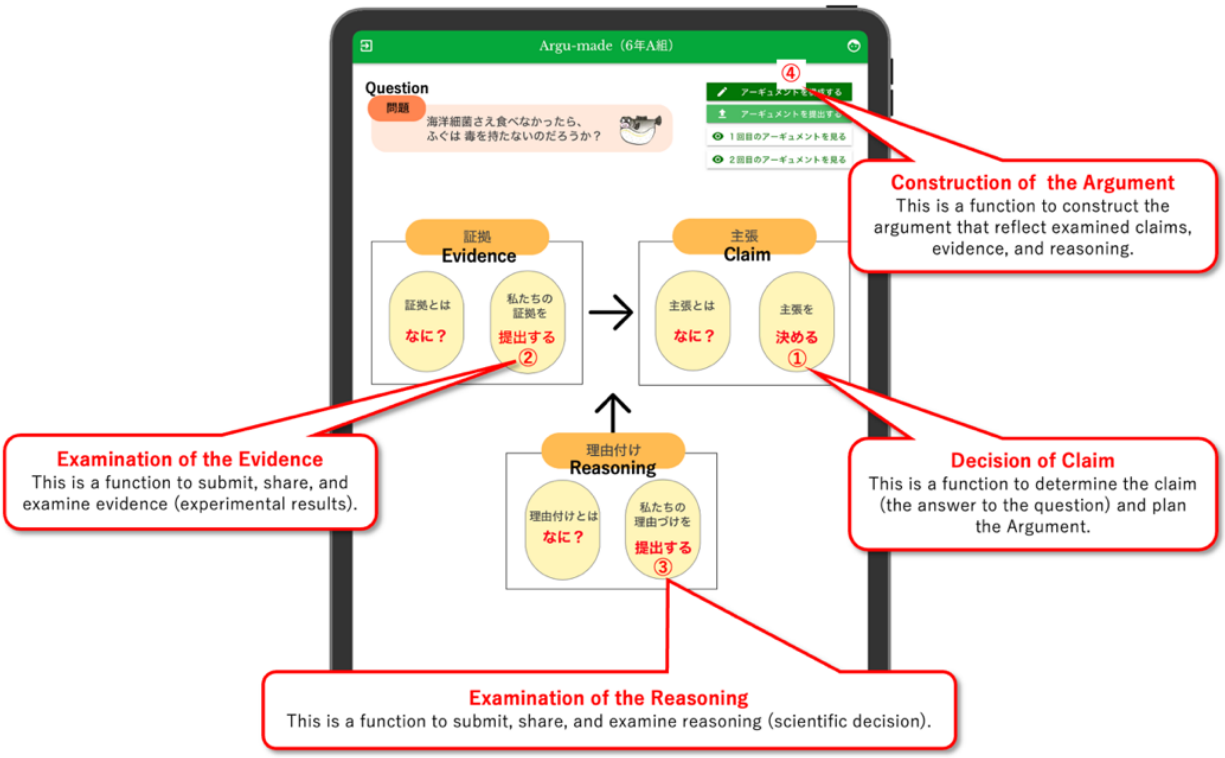
Did Argu-made 2.0 support the evidence for bioaccumulation of pufferfish toxin?

Was Argu-made 2.0 effective in achieving scientific judgement?

Argument Construction Support System 'Argu-made 2.0'

Argu-made 2.0—a system developed using the no-code application development tool 'Adalo'—supports the construction of an argument consisting of 'claim (answer to a question) - evidence (experimental results) - reasoning (scientific principle)'. Figure 1 shows the home screen of Argu-made 2.0. Using the 'Decision of Claim (①)' function, users decide on the claim (the answer to the question) and make a plan for the construction of the argument, such as considering the characteristics of evidence to enhance its scientific reliability and validity (hereinafter referred to as 'appropriate evidence'). Using the 'Examination of the Evidence (②)' function, they can compare evidence and discuss the appropriateness of each evidence while considering the evidence to be used in the argument. The 'Examination of the Reasoning (③)' function allows them to construct reasoning that strongly supports the scientific credibility of their claims (hereafter referred to as 'appropriate reasoning'), just as when considering the selection of evidence.

Figure 1. Home screen.



Selecting 'Examination of the Evidence (②)' takes the users to the screen shown in Figure 2. Several pieces of evidence are submitted and shared here. They can extract appropriate evidence with a good mark (②-2) using the characteristics of the appropriate evidence noted in ②-1 as a hint. Furthermore, by selecting the comment mark (②-3) on each piece of evidence, they can move to the screen shown in Figure 3 to examine the appropriateness of the specified evidence. They can share their opinions, such as reasons, with others using the comment-sending function (②-4), and check the contents in the list (②-5).

Figure.2 Screen for submitting and sharing evidence

Figure.3 Screen for discussing the appropriateness of evidence

Overview Of The Class

A four hour long class was conducted by the second author in the science course ‘Connections among living creatures’ for 55 sixth-grade students (11–12 years old). Students had already studied the food-chain on land and in rice paddies. At the beginning of this class, they also knew that pufferfish toxin is not produced endogenously but by marine bacteria. They were asked the following learning question: ‘If only marine bacteria were not eaten, would pufferfish be poisonous?’ They made scientific judgements about the presence or absence of pufferfish toxins in the following two scenarios.

In Scenario 1, using the simulator and the knowledge of the content of the previous class, the students were asked to organise the food-chain with the five creatures (pufferfish, crabs, starfish, clams, and marine bacteria) and understand the basic mechanism of bioaccumulation: All the toxin in the prey is absorbed into the predator’s body. In Scenario 2, through simulations, they confirmed the mechanism of bioaccumulation: The amount of toxin accumulated depends on what and how much a predator preys on, and more toxin is accumulated as it gets closer to the top of the food-chain. Subsequently, they constructed arguments using Argu-made 2.0. First, the students were given the prospect of constructing the argument by deciding their claim with the function of ‘Decision of Claim (① in Figure 1)’. Next, they submitted and shared their evidence using the function of ‘Examination of the Evidence (② in Figure 1)’, and extracted the evidence they considered appropriate with a good mark (②-2 in Figure 2). Only one appropriate piece of evidence was selected when discussing the appropriateness of each piece of evidence with comments (②-4, 5 in Figure 3). Finally, the function of ‘Examination of the Reasoning (③ in Figure 1)’ was used to construct the appropriate reasoning just as when considering the selection of evidence.

Analysis

The class was evaluated from two perspectives: ‘examination of evidence’ and ‘achievement of scientific judgement’. In the ‘examination of evidence’, the extraction of evidence with good marks (②-2 in Figure 2) from students’ activity history in Argu-made 2.0 in scenario 2 was analysed by classifying the students into four types (Table 1).

For the ‘achievement of scientific judgement’, the content of the students’ arguments in scenario 2 were analysed by scoring them, and their judgements about the presence or absence of pufferfish toxins were evaluated. We scored the claims, evidence, and reasoning independently on a 1-point scale, with the sum of the scores on a 3-point scale, based on a rubric. Concerning claim, one point was awarded if the student was able to write ‘poisonous’ in response to the question: If only marine bacteria were not eaten, would pufferfish be poisonous? Concerning evidence, one point was awarded if the student was able to select the experimental result: the pufferfish fed on zero marine bacteria and at least one other creature of one or more species. For reasoning, one point was awarded for correctly describing the basic explanation of how bioaccumulation works: All the toxins in the prey are absorbed into the body of the predator.

Results

From the ‘examination of evidence’ section, Table 1 shows the status of evidence extracted with good-mark. Of the students who selected appropriate evidence, Type A (42 [76%]) extracted only appropriate evidence with good marks. Type B (6 [11%]) selected appropriate evidence at the end, although they extracted appropriate and inappropriate evidence with good marks.

Among the students who ultimately failed to select appropriate evidence, Type C (2 [4%]) extracted appropriate and inappropriate evidence with good marks. Type D (4 [7%]) extracted all inappropriate evidence with good marks. Type E (1 [2%]) did not select anything at the extraction or selection stage.

In the ‘achieving scientific judgement’ section, 54 out of 55 [98%] students received a score of 1 for the claim, 48 [87%] for the evidence, and 47 [85%] for the reasoning in each component. In addition, 42 students [76 %] correctly constructed all the elements (students who scored a perfect score of 3). Most students used the food-chain and bioaccumulation mechanisms to make scientific judgements regarding the presence or absence of pufferfish toxins while providing evidence of appropriate simulation results.

Table 1. Extraction of evidence with the good-mark to its adoption into the argument

Type	Number of students [The ratio]
A: All evidence given a good mark is appropriate, and the appropriate evidence is ultimately selected.	42 [76 %]
B: The evidence given a good mark is a mixture of appropriate and inappropriate evidence, but the appropriate evidence is ultimately selected.	6 [11 %]
C: The evidence given a good mark is a mixture of appropriate and inappropriate evidence, and the inappropriate evidence is ultimately selected.	2 [4 %]
D: All evidence given a good mark is inappropriate, and the inappropriate evidence is ultimately selected.	4 [7 %]
E: None of the evidence was given a good mark, or no evidence was ultimately selected.	1 [2 %]

Discussion

Argu-made 2.0 developed in this study was able to support the examination of evidence regarding the bioaccumulation of pufferfish toxins. Based on the status of the extraction of evidence with good marks, Argu-made 2.0 effectively helped sixth-grade students examine appropriate evidence. This effectiveness was attributed to Argu-made 2.0 encouraging the step-by-step selection of evidence using the good-mark function and supporting discussion its appropriateness during the selection process.

Furthermore, Argu-made 2.0 was effective in achieving scientific judgements. Most students used food-chain and bioaccumulation mechanisms to make scientific judgements about the presence or absence of pufferfish toxins while providing appropriate evidence. This success was likely because the selected evidence could be uploaded immediately to the argument, reducing the workload for the students and allowing them to focus on constructing arguments without interruptions. Additionally, since Argu-made 2.0 allows repeated modifications after uploading, this feature likely encouraged the students to re-examine the appropriateness of their evidence.

Future studies should investigate these factors in greater detail through questionnaires and follow-up interviews.

Acknowledgement

This study was supported by JSPS KAKENHI Grant Number JP23K25719.

References

Duncan, G. R., Chinn, A. C., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the next generation science standards’ conceptualization of evidence. *Journal of Research in Science Teaching, 55*(7), 907-937.

Iordanou, K., Kuhn, D., Matos, F., Shi, Y., & Hemberger, L. (2019). Learning by arguing. *Learning and Instruction, 63*, 101207.

Kamiyama, S., Yamamoto, T., Yamaguchi, E., Sakamoto, S., Muratsu, K., & Inagaki, S. (2015). Instructional strategies for teaching primary students to construct arguments with rebuttals. In J. Lavonen, K. Juuti, J. Lampiselkä, A. Uitto & K. Hahl (Eds.), *Electronic Proceedings of the ESERA*

2015 Conference. Science education research: Engaging learners for a sustainable future, Part7 (co-ed. Maria Andrée & Maria Pilar Jimenez-Aleixandre), 997-1003.

Malloggi, C., Rizzo, B., Giusti, A., Guardone, L., Gasperetti, L., Dall'Ara, S., & Armani, A. (2023). First toxicological analysis of the pufferfish sphoeroides pachygaster collected in Italian waters (strait of sicily): Role of citizens science in monitoring toxic marine species. *Animals, 13,* 1873.

Matsuyama, Y., Kamiyama, S., Funaoi, H. & Yamamoto, T. (2023). Evaluation of simulators to promote the understanding of bioaccumulation among elementary school students. In Shih, JL. et al. (Eds.), *31th International Conference on Computers in Education Conference Proceedings, 2*, 863-865.

Matsuyama, Y., Kamiyama, S., Funaoi, H. & Yamamoto, T. (2024a). Development and evaluation of classes encouraging judgment of masticatory behavior based on scientific evidence: A fourth grade elementary school class "structure and movement of human body" with "Argu-made." *Journal of Science Education, 48*(4), 349-364.

Matsuyama, Y., Kamiyama, S., Funaoi, H. & Yamamoto, T. (2024b, November 11-15). *Evaluation of a Japanese elementary school class using simulators and arguments to promote scientific decisions on food poisoning prevention regarding pufferfish-toxin* [Oral session]. XXI International Organization for Science and Technology Education 2024, Kuching, Sarawak, Malaysia.

Sharif, L., & Al-Malki, T. (2010). Knowledge, attitude and practice of Taif University students on food poisoning. *Food Control, 21*(1), 55-60.

Gender Differences In Science Groupwork Participation: Exploring Students' Voices Using Automatic Speech Recognition And Social Network Analysis

Jingoo Kang and Stephen Adofo
Univeristy of Eastern Finland, Finland

This study explores gendered participation patterns in secondary school science group work by integrating Automatic Speech Recognition (ASR) and Social Network Analysis (SNA) in authentic classroom settings. Audio data were collected during science activities involving 16 students in a Finnish secondary school, using individual microphones with separate-channel recording to capture student speech. ASR was applied to generate time-stamped transcriptions and speaker labels, and interaction networks were subsequently constructed using SNA to identify participation patterns and student roles. Student speech was also manually coded into functional categories to examine discourse roles over time. The results demonstrate that ASR, when combined with individual microphone systems, can accurately differentiate speakers and capture student interactions in natural classroom environments. SNA effectively identified central and peripheral participation roles, while discourse analyses revealed how gender, science identity, and extroversion intersect to shape students' participation patterns. Although exploratory, this study highlights the methodological potential of combining ASR and SNA for fine-grained and scalable analyses of equitable participation in science classrooms.

Keywords: gender difference, automatic speech recognition, social network analysis

Introduction

Gendered Participation In School Science Group Work

Over the last decade, in many countries including Finland, female students have outperformed male students in school science subjects (OECD, 2016; 2023). In addition, previous studies indicate that female students tend to report stronger preferences for collaboration (Li et al., 2021) and higher levels of extroversion than their male peers (Weisberg et al., 2011). Taken together, these characteristics—higher academic performance, greater cooperativeness, and stronger extroversion—would suggest that female students are likely to participate actively in science group work.

Contrary to this expectation, participation in science group work often follows gendered patterns. Research consistently shows that boys tend to dominate group interactions and assume leadership roles, whereas girls are more frequently positioned at the periphery of group activities (Wieselmann et al., 2020). One potential explanation for this pattern lies in differences in science identity. Boys, on average, report higher levels of science identity, encompassing recognition, confidence, interest, and perceived competence, compared to girls (Williams & George-Jackson, 2014).

However, science identity alone does not fully explain observed participation dynamics. Empirical evidence suggests that some girls with high levels of science identity still participate passively in group work, while some boys with relatively low science identity engage actively (Kang, 2023). Moreover, it remains underexplored why girls' reported extroversion and collaborative preferences are not consistently manifested in their observable participation during science group activities. These inconsistencies point to a need for more nuanced analyses that consider the intersection of gender, science identity, and personal characteristics in shaping participation patterns.

To date, however, most studies investigating gendered participation in science group work have relied on small samples and labor-intensive manual observation or coding. While such approaches have provided important insights into classroom interaction, they offer limited temporal resolution and constrain the ability to examine participation patterns systematically across groups and contexts. There is therefore a clear need for methodological approaches that enable scalable, fine-grained analyses of student participation in authentic classroom settings, making it possible to link individual dispositions with observable interactional roles and participation trajectories.

Emerging Technologies For Capturing Participation Patterns

Traditionally, participation patterns during group work have been examined through direct classroom observation or video-based coding schemes, both of which are time-intensive and limited in scalability (e.g., Wieselmann et al., 2020). While these approaches allow for rich qualitative interpretation, they pose challenges for capturing fine-grained interactional dynamics across larger samples and multiple classroom contexts.

Recent advances in educational data science have introduced computational approaches that offer new possibilities for analysing classroom interactions in more systematic and scalable ways. One such approach is Automatic Speech Recognition (ASR), which enables the automated transcription of spoken interactions and the quantification of verbal participation (Jensen et al., 2021; Pugh et al., 2021). In classroom research, ASR has primarily been applied to teacher talk, as it is logistically easier to equip a single teacher with a high-quality microphone than to provide individual recording devices for students (Jensen et al., 2021). Only a small number of studies have applied ASR to student speech. For instance, Pugh et al. (2021) examined collaborative dialogue among students aged 12–15 using headset microphones; however, their data were collected outside natural classroom environments, with students interacting in pairs over Zoom. Consequently, the potential of ASR to capture student participation in authentic, face-to-face science group work remains underexplored.

While ASR enables the identification of who speaks and how much, it provides limited insight into the relational structure of interactions within a group. To address this limitation, Social Network Analysis (SNA) has increasingly been adopted in educational research as a powerful analytical framework for examining interactional patterns among learners (Saqr et al., 2020). SNA conceptualizes classroom interactions as networks, where students are represented as nodes and communicative exchanges as ties. This approach makes it possible to identify participation structures, such as central or peripheral roles, leadership, and dominance, that may not be readily visible through frequency-based measures alone.

In the context of group work, SNA has been used to examine collaboration quality, knowledge construction, and social positioning within learning groups (Saqr et al., 2018; 2020). Importantly, SNA offers a theoretically grounded way to operationalize participation as a relational phenomenon rather than an individual attribute, making it particularly well suited for investigating gendered participation patterns.

Integrating ASR with SNA offers a promising methodological advancement. ASR can provide automated, fine-grained data on verbal interactions, while SNA can model these interactions as dynamic networks that reveal students' participation patterns and roles within group work. Despite this potential, empirical studies that combine ASR and SNA to analyse student participation in natural classroom settings remain scarce. This study seeks to address this gap by leveraging the complementary strengths of ASR and SNA to examine gendered participation patterns in science group work.

Research Aim And Questions

Addressing these gaps, the present study aims to examine students' participation patterns during science group activities conducted in natural classroom settings. Specifically, it investigates how gender, science identity, and personality characteristics—particularly extroversion—are associated with students' participation patterns and roles within group work. Methodologically, this study explores the integration of ASR and SNA as complementary tools for identifying participation structures and interactional roles in an automated and scalable manner.

The study is guided by the following research questions:

RQ1. Can ASR capture and differentiate students' speech during science group activities in natural classroom settings?

RQ2. Can SNA effectively identify participation patterns and student roles within science group work?

RQ3. How are gender, science identity, and personality traits (extroversion) related to students' participation patterns and roles in science group activities?

Methods

Sample, Data Collection, And Measures

The study was conducted with high school students participating in a mandatory two-day science project on sustainability in Finland. A total of 28 students took part in the project, of whom 16 volunteered to have their voices recorded during the first two-hour brainstorming session. These 16 students were organized into four groups, each consisting of four members. Three groups comprised two female and two male students, while one group consisted of one female and three male students.

To capture individual student speech during group work, each participating student was provided with a wireless microphone. Audio signals were recorded on separate channels using an LR20 mixer to ensure high-quality speech input suitable for ASR processing. After recording, the audio channels were grouped according to team membership and merged into four final audio files—one per group—while preserving channel-level information for speaker identification.

In addition to audio data, students completed questionnaire measures assessing science identity and personality characteristics, including extroversion. These self-report measures were used to contextualize participation patterns identified through ASR and SNA and to explore potential associations between individual characteristics and participation roles.

Data Analysis

ASR processing generated time-stamped transcriptions and speaker labels, which served as the basis for constructing interaction networks for SNA. To address RQ1, the quality of ASR outputs was evaluated using two complementary metrics: Word Error Rate (WER) and Diarization Error Rate (DER). WER quantifies the proportion of incorrectly recognized words relative to a human reference transcript, with lower values indicating greater transcription accuracy. In speech recognition research, lower WER values (e.g., below approximately 20–30% in noisy multi-speaker environments) have been considered acceptable for interaction-level analysis rather than for perfect lexical accuracy alone (Attia et al., 2023).

On the other hand, DER measures the proportion of total recording time that is incorrectly attributed to speakers, combining false alarms, missed speech segments, and speaker confusions into a single error metric. DER is the most widely used metric for evaluating speaker diarization systems, especially in multi-speaker contexts where correctly identifying "who spoke when" is critical for interaction analysis. In natural, spontaneous speech environments such as classroom group discussions, DER values up to mid-30% range have been reported in recent diarization research as indicative of a feasible diarization system for exploratory interaction analysis (Wang et al., 2024).

The outputs of ASR (time-stamped transcriptions and speaker attributions) were used to construct participation networks for SNA. In these networks, students were represented as nodes, and ties were defined based on temporal co-occurrence of speech segments and turn-taking interactions. SNA indicators (e.g., degree and eigenvector centrality) were then used to distinguish between more central and more peripheral participants within each group, addressing RQ2.

Survey data on science identity and personality traits were integrated with ASR- and SNA-derived indicators to explore associations between individual characteristics (gender, science identity, extroversion) and participation patterns, addressing RQ3. Given the pilot nature of the study and the small sample size, these analyses were exploratory and aimed at identifying preliminary trends rather than testing causal relationships.

To complement the automated analyses, a manual content analysis of student speech was conducted. Each utterance was coded into one of six functional categories: directing, elaborating, questioning, sharing, disagreeing, or agreeing. This coding provided (a) richer interpretive insight into participation patterns beyond structural measures, and (b) a labelled dataset for future development of machine learning models for automated speech function classification. In the present pilot study, the manually coded speech functions were incorporated into the participation analysis to capture both quantitative interaction structures and qualitative discourse roles.

Results

RQ1. Can ASR Capture And Differentiate Students' Speech During Science Group Activities In Natural Classroom Settings?

Audio data were collected from four student groups in a secondary school science classroom. To address RQ1, an initial performance evaluation was conducted using the audio data from the first group to compare several commercially available ASR systems, including Whisper, Google Cloud Speech-to-Text, AWS Transcribe, and AssemblyAI. The comparison focused primarily on the systems' ability to accurately differentiate between individual speakers in a natural classroom environment. Based on this evaluation, AssemblyAI was selected for subsequent analyses, as it demonstrated the highest speaker differentiation accuracy.

For the first group, ASR performance was quantitatively evaluated using WER and DER. The word-level comparison between ASR output and human transcripts yielded a WER of approximately 2.1% (63 word errors out of 3001 reference words), indicating very high lexical accuracy. At the utterance level, comparison of human and ASR speaker labels resulted in a confusion-based DER of approximately 19.5%, corresponding to a speaker attribution accuracy of about 80.5%. These values substantially exceed typical feasibility thresholds reported for classroom-based, multi-speaker ASR applications, where considerably higher error rates are often considered acceptable for participation-level analysis. Following this initial evaluation, AssemblyAI was applied to the remaining three groups.

Despite the overall high accuracy, minor errors were observed in both word recognition and speaker identification. For example, as shown in Table 1, in Occurrence 1, the utterance "I don't have any ideas" was incorrectly transcribed as "I do have any ideas," representing a word-level recognition error. In Occurrence 298, an utterance was incorrectly attributed to a group member when it originated from a brief interruption by a neighboring group, resulting in a speaker attribution error. Such cases were rare and did not meaningfully affect the identification of speaker turns or overall participation patterns.

As illustrated in Table 1, the ASR system detected approximately 300 speech occurrences among the four members of the first group, providing detailed information on speech duration and speaker identity. Each entry includes an occurrence number, timestamps marking the start and end of speech, an assigned speaker label (A, B, C, or D), and the corresponding transcription.

Overall, the results demonstrate that ASR can capture and differentiate students' speech with high accuracy during science group activities in natural classroom settings. The high level of performance observed in the initial evaluation, together with consistent outputs across all groups, supports the feasibility of using ASR for analysing student participation patterns, addressing RQ1.

Table 1. Sample conversation excerpt from ASR output.

Occurrence	Duration	Speaker: Statement
1	00:00:01,439 --> 00:00:03,020	Speaker A: I do have any ideas.
2	00:00:06,879 --> 00:00:10,259	Speaker B: What kind of inequality do you want to focus on?
3	00:00:12,480 --> 00:00:15,660	Speaker A: Unfortunately, I haven't experienced any inequality.
4	00:00:20,239 --> 00:00:24,300	Speaker C: Last year they did the accessibility.
5	00:00:24,719 --> 00:00:50,466	Speaker D: The first thing that pops into my mind is like, you know, language inequalities, you know, like if you don't speak Finnish, like, how much that sets you back. Like in general, living in XXX or in Finland, of course it like depends on the area. Like in XXX, like, there might be more English translations and stuff, but like, think about how hard it is to participate in everything.
….	….	….
298	00:46:16,420 --> 00:46:57,019	Speaker D: How are you guys doing? What is it? What is it? That is interesting. That is such a good idea. I know. Great.
299	00:46:59,159 --> 00:47:13,619	Speaker A: Yeah. We're doing about nutrition in relation of malnutrition to poverty and reinforcing this free school food with nutrients which are, like, less common in the poorer.

RQ2. Can SNA Effectively Identify Participation Patterns And Student Roles Within Science Group Work?

Following ASR-based transcription and speaker identification, Social Network Analysis (SNA) was conducted to examine interaction patterns within student groups. Due to space limitations, the results presented here focus on a single 45-minute recording from Group 1, consisting of two male and two female students. The speakers' characteristics were as follows: Speaker A (Male 1, high science identity, extroverted), Speaker B (Female 1, low science identity, introverted), Speaker C (Male 2, average science identity, extroverted), and Speaker D (Female 2, average science identity, extroverted).

Frequency Of Speech Transitions And Speaking Duration

A total of 299 speech transitions were identified and ranked by frequency (Table 2). The most frequent transitions occurred between Speaker A and Speaker C (A→C: 70; C→A: 66), indicating that a substantial portion of the discussion was concentrated between the two male students. The next most frequent transitions were between Speaker C and Speaker D (C→D: 50; D→C: 45), suggesting that Speaker C actively engaged with multiple group members. In contrast, transitions involving Speaker B were rare, with only three transitions observed between Speaker C and Speaker B and between Speaker B and Speaker A, indicating limited participation and a peripheral role in the discussion.

Analysis of speaking duration further supported these patterns. Speaker C had the longest total speaking time (751.52 seconds), followed by Speaker D (571.53 seconds) and Speaker A (564.83 seconds), whereas Speaker B spoke substantially less (119.76 seconds). Together, the high frequency of transitions and extended speaking duration position Speaker C as the most central participant in the group, while Speaker B appears comparatively marginalized.

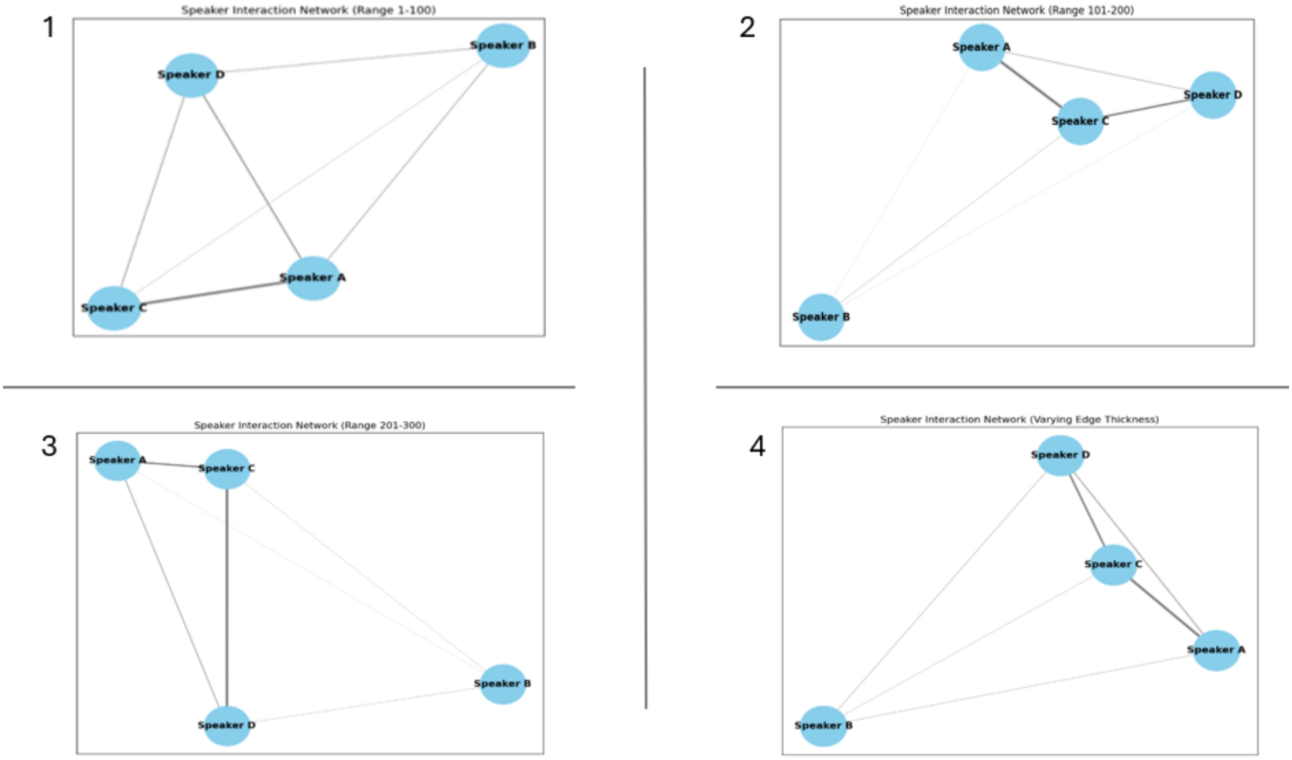
Table 2. Speech transition occurrences between group members.

Rank	Speaking direction	Frequency	Rank	Speaking direction	Frequency
1	'Speaker A' → 'Speaker C'	70	7	'Speaker A' → 'Speaker B'	8
2	'Speaker C' → 'Speaker A'	66	8	'Speaker B' → 'Speaker D'	8
3	'Speaker C' → 'Speaker D'	50	9	'Speaker D' → 'Speaker B'	5
4	'Speaker D' → 'Speaker C'	45	10	'Speaker B' → 'Speaker C'	5
5	'Speaker D' → 'Speaker A'	22	11	'Speaker B' → 'Speaker A'	3
6	'Speaker A' → 'Speaker D'	14	12	'Speaker C' → 'Speaker B'	3

Network Structure And Temporal Dynamics

Figure 1 illustrates interaction networks derived from SNA, where node proximity reflects interaction strength and edge thickness represents transition frequency. To examine temporal changes in participation patterns, the interaction data were segmented into three consecutive phases of 100 speech occurrences, followed by an aggregated network representing all 300 transitions.

Figure 1. Temporal changes in interaction patterns among four participants based on SNA results.



During the first phase, interactions were relatively balanced, although Speakers A and C were already more strongly connected than other pairs. In the second phase, Speaker C became increasingly central in the network, while Speaker B engaged less frequently and moved farther from the core interaction structure. This pattern intensified in the third phase, with Speakers C and A forming even stronger ties. The aggregated network highlights Speaker C's dominant role in the discussion and Speaker B's consistently peripheral position.

Overall, these results demonstrate that SNA effectively captures participation patterns and reveals distinct student roles within science group work. By integrating transition frequency, speaking duration, and network structure over time, SNA enabled the identification of central and peripheral participants, providing clear evidence in support of RQ2.

RQ3. How Are Gender, Science Identity, And Personality Traits (Extroversion) Related To Students' Participation Patterns And Roles In Science Group Activities?

To further explore how individual characteristics relate to participation patterns and roles, each speech occurrence was manually coded into one of six functional categories. This analysis complemented the SNA results by examining not only how much students participated, but also how they contributed to the ongoing discussion.

Among the six categories, cumulative elaboration statements were first examined to trace how participants contributed to advancing the discussion over time. As shown in Figure 2, the two male students exhibited a crossing pattern in cumulative elaboration, indicating that they alternated in taking the lead during the conversation. In contrast, Speaker B (Female 1, low science identity, introverted) contributed to elaboration primarily during the early phase of the session but gradually withdrew from active participation around the midpoint. Speaker D (Female 2, average science identity, extroverted) showed a gradual increase in elaboration contributions over time; however, her participation followed a different trajectory from that of the male students and did not converge with the dominant interaction pattern.

Figure 2. Cumulative elaboration statements over time.

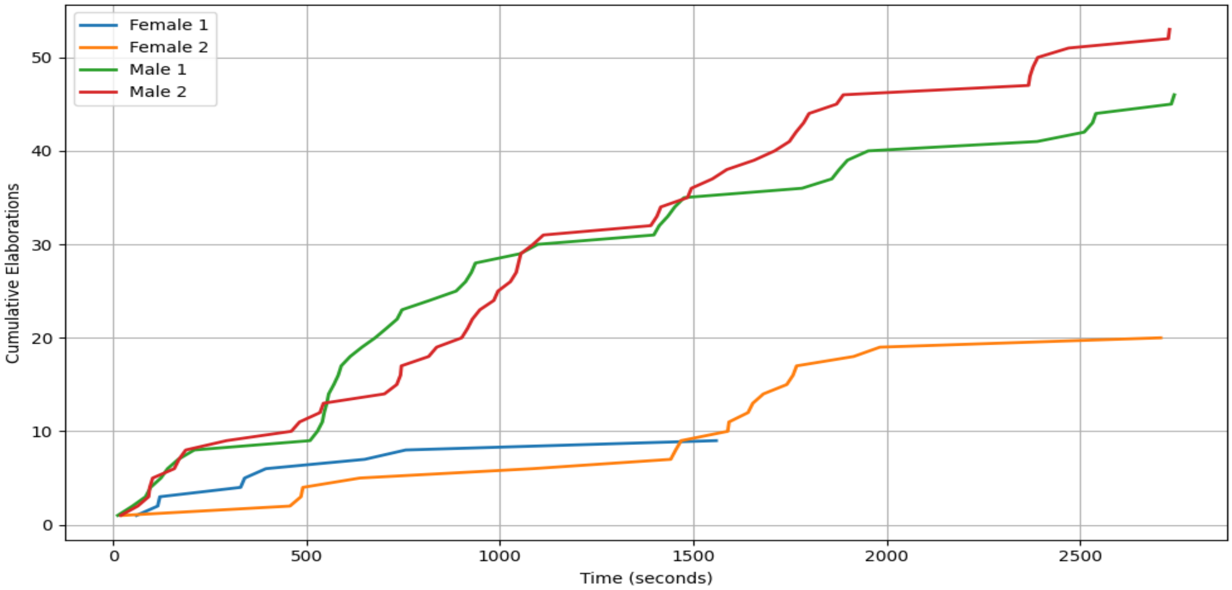
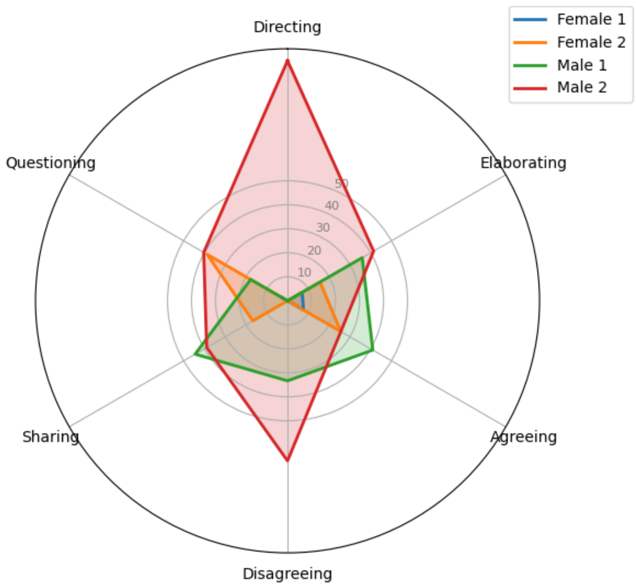


Figure 3 presents the proportion of speech functions attributed to each speaker. Speaker C (Male 2, average science identity, extroverted) dominated across multiple conversational functions, including directing, elaborating, and disagreeing, indicating a central and influential role in shaping the discussion. Speaker A (Male 1, high science identity, extroverted) contributed primarily through elaboration, suggesting a supportive but still central role. In contrast, Speaker B (Female 1, low science identity, introverted) accounted for a substantially smaller proportion of all speech categories, reflecting a marginalized participation role. Speaker D (Female 2, average science identity, extroverted) contributed across several categories but was less involved in directing or disagreeing, functioning primarily as an active follower within the group interaction.

Taken together, these findings suggest that participation patterns and roles in science group work are shaped by an interplay of gender, science identity, and personality traits. In this group, male students—regardless of differences in science identity—occupied more central and directive roles, while female students displayed more varied participation trajectories that were sensitive to both science identity and extroversion. Although exploratory in nature, these results provide initial evidence that individual characteristics are meaningfully related to how students participate and position themselves within science group activities, addressing RQ3.

Figure 3. Proportion of each category attributed to speakers (%).



Discussion

Building on the results addressing RQ1, the ASR system demonstrated performance that exceeded commonly reported feasibility thresholds for classroom-based, multi-speaker environments (Wang et al., 2024; Attia et al., 2024). This high level of performance can be attributed not only to recent advances in ASR technology but also to the use of individual microphones with separate-channel recording, which substantially improved speaker separation and audio quality. Together, these results provide empirical evidence that ASR, when supported by individual microphone systems, can be successfully applied to investigate student interactions in authentic, co-located classroom activities involving more than 16 students.

In addition, whereas previous studies have primarily focused on online learning contexts (Saqr et al., 2018; 2020), the integration of SNA in the present study proved effective for identifying participation patterns and distinguishing central and peripheral roles within authentic, face-to-face science group work (RQ2). By modelling interaction structures over time, SNA made visible participation dynamics that are difficult to capture through traditional observational approaches alone. When combined with functional analyses of student talk, the results further revealed how participation roles are shaped by a complex interplay of gender, science identity, and personality traits such as extroversion (RQ3). These findings underscore the importance of considering individual characteristics when examining collaborative learning processes in science education.

While the present study is exploratory in nature and based on a limited number of groups, it demonstrates the methodological promise of combining ASR and SNA for scalable and fine-grained analyses of classroom interaction. Accordingly, future work should further validate the use of ASR and SNA in educational research and deepen understanding of equitable participation and inclusion in science classrooms with larger student populations.

Acknowledgement

This study is funded by the Research Council of Finland [Grant No. 364000]

References

Kang, J. (2023). Gendered science practice at secondary school and its effects on science motivations. International Journal of Science and Mathematics Education, 21(6), 1725-1747.

Attia, A. A., Demszky, D., Ogunremi, T., Liu, J., & Espy-Wilson, C. (2024). Continued pretraining for domain adaptation of wav2vec2. 0 in automatic speech recognition for elementary math classroom settings. arXiv preprint arXiv:2405.13018.

Kang, J. (2023). Gendered science practice at secondary school and its effects on science motivations. International Journal of Science and Mathematics Education, 21(6), 1725-1747.

Jensen, E., L. Pugh, S., & K. D'Mello, S. (2021, April). A deep transfer learning approach to modeling teacher discourse in the classroom. In LAK21: 11th International Learning Analytics and Knowledge Conference (pp. 302-312).

Li, J., Zhang, M., Li, Y., Huang, F., & Shao, W. (2021). Predicting students' attitudes toward collaboration: Evidence from structural equation model trees and forests. Frontiers in Psychology, 12, 604291.

OECD. (2016). PISA 2015 results (Volume I). Excellence and equity in education. OECD Publishing.

OECD. (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. OECD Publishing

Pugh, S., Subburaj, S. K., Rao, A. R., Stewart, A. E., Andrews-Todd, J., & D'Mello, S. K. (2021, November). Say what? Automatic modeling of collaborative problem solving skills from student speech in the wild. In Proceedings of the 14th international conference on educational data mining. Educational Data Mining.

Saqr, M., Fors, U., Tedre, M., & Nouri, J. (2018). How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. PloS one, 13(3), e0194777.

Saqr, M., Viberg, O., & Vartiainen, H. (2020). Capturing the participation and social dimensions of computer-supported collaborative learning through social network analysis: which method and measures matter?. International Journal of Computer-Supported Collaborative Learning, 15(2), 227-248.

Wang, J., Dudy, S., He, X., Wang, Z., Southwell, R., & Whitehill, J. (2024). Speaker Diarization in the Classroom: How Much Does Each Student Speak in Group Discussions?. In Proceedings of the 17th International Conference on Educational Data Mining (pp. 360-367).

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. Frontiers in psychology, 2, 178.

Wieselmann, J. R., Dare, E. A., Ring-Whalen, E. A., & Roehrig, G. H. (2020). "I just do what the boys tell me": Exploring small group student interactions in an integrated STEM unit. Journal of Research in Science Teaching, 57(1), 112-144.

Williams, M. M., & George-Jackson, C. (2014). Using and doing science: Gender, self-efficacy, and science identity of undergraduate students in STEM. Journal of Women and Minorities in Science and Engineering, 20(2).

Unpacking Science Teachers' Discursive Practices In Instructional Modelling.

Alba Cortés Morales and Ainoa Marzabal
Pontificia Universidad Católica de Chile, Chile

This study explores the discursive practices of science teachers in modelling-oriented teaching contexts, particularly emphasising dialogic classroom decision-making and its impact on students' learning opportunities. Teachers' speech turns are analysed when implementing a task with modelling potential to visualise these discursive processes. This is achieved through a discourse analysis of the productive conversation, including the breakdown of the dialogue into modelling phases (familiarisation, discussion, and consensus) and the classification of interaction types such as validation, information, clarification, challenge, and consolidation. The unpacking of these modelling-oriented talk moves highlights how teachers perceive and respond to student responses and how their discourse can facilitate or hinder the modelling process. The results suggest that the talk moves teachers use to enable the development of a productive conversation in which students' cognitive engagement in modelling processes is promoted.

Keywords: Instructional Modelling, Talk Moves, Teachers' Discursive Practices

Theoretical Framework

In accordance with the needs of the learning process in science education is critical for teachers to promote interactive talk in the classroom, as questioning students increases students' motivation to learn, supports productive classroom discussion, and fosters knowledge construction (Tiffany et al., 2023; Tytler & Aranda, 2015). Questioning is a process of collaborative learning that requires active, reciprocal participation that promotes a tension between students' need to express their ideas and teachers' need to mediate their reasoning in terms of scientific knowledge (Hamnell-Pamment, 2023). In modelling-based instruction, this tension is palpable in the mediation process teachers enact to facilitate students' mental model-building. As a teaching approach, modelling instruction requires teachers to promote an instructional task sequence that allows students to engage in modelling practices (Garrido & Couso, 2024). This performance is associated with a student-centred discourse, where teachers encourage students to express their ideas and promote divergence and convergence to help them gradually revise and refine their initial models.

Promoting the performance of modelling practices through dialogic teaching involves spaces to listen to other ideas, reasoning with others, and negotiating new ideas (Vergara et al., 2025b). So far, the study of the dialogic opportunities given in science instruction has been focused in the count of instances in which discursive strategies are used (Christodoulou & Osborne, 2014), the comparison of the distribution of linguistic movements in the dialogue (Balyan et al., 2022), the presentation of dialogue in a particular episode (Cui & Teo, 2023), the time spent on the stages of a task (Newman, 2017), flow diagrams (Larusso et al., 2023), and the use of discursive categories over time (Tytler & Aranda, 2015). In science modelling processes, visualisations have focused on the flow of the chain of discursive strategies (Vergara-Sandoval et al., 2025a) and graphs of discourse variation in different episodes (Garrido Espeja, 2016). However, to understand how teachers promote modelling practices, a deeper analysis of their dialogical decisions is required.

Exploring productive conversations between teachers and students through conversation analysis can shed light on teachers' discursive strategies to facilitate model construction (Hamnell-

Pamment, 2023). While embedded within the intricate, in-the-moment decision-making processes of teaching, these strategies can be identified and visualised. To achieve this, it is essential to analyse teachers' talk moves in terms of their specific purposes. This approach also highlights the importance of examining the balance between allowing students to articulate their reasoning and the teacher's role in guiding instructional modelling (Hamnell-Pamment, 2023).

Analysing teacher discourse in dialogical science classrooms can reveal patterns and strategies that contribute to refining students' ideas as they develop a school scientific model (SSM). However, the learning opportunities available depend significantly on how tasks are implemented (Krist et al., 2023). Given that task implementation can often diminish a task's modelling potential, it becomes crucial to integrate an analysis of teachers' discursive practices that foster students' intellectual engagement (Tekkumru-Kisa et al., 2023). Although efforts have been made to enhance students' epistemic agency (Krist et al., 2023), the role of teachers' discursive decisions in shaping discussions within the specific context of task implementation has not been extensively explored.

Therefore, it is important to identify how talk moves within a modelling process mediates students' learning (Tytler & Aranda, 2015), particularly when considering how these moves are shaped by teachers' noticing, interpretation, and responses during task implementation and student interactions. This research aims to examine visualisations of classroom interactions between teachers and students by analysing modelling-oriented talk moves and their patterns, with a focus on decisions that support students' modelling processes. By characterising these talk moves, the study seeks to illustrate the learning opportunities created by science teachers through their modelling-oriented discourse.

Research Question

In characterising classroom interactions between teachers and their students in which dialogic decision-making affects learning opportunities in a task with modelling potential, we seek to respond to the following research question:

> RQ: What discursive decisions do science teachers make to facilitate modelling processes in their students?

Methods & Context

To achieve the study's objective, a qualitative research approach with discourse analysis was employed. The study involved 3 Chilean chemistry teachers, Chris, Marina, and Caroline, who constitute the participants of this research. The data collection included the audiovisual recording of one lesson for each secondary education teacher.

Context Of The Study

The teachers in this study are part of a learning community in model-based teaching. The learning community lasted a year and a half, during which they revised their ideas about models and modelling, discussed their experiences, and received feedback on the tasks with modelling potential that they enacted in their classrooms. In this context, in the last months of the learning community, the teachers co-designed a modelling-oriented teaching and learning sequence (TLS) that follows the criteria for a modelling-oriented task proposed by Cortés-Morales & Marzabal (2025a) and the design principles for a modelling-oriented TLS proposed by Tena & Couso (2023).

To recognise teachers' discursive practices and establish possible talk move patterns in the modelling conversations, the implementation of task 5 was selected for its characterisation: "Represent and describe the final state of the reaction". This task is part of an experiment that compares the reaction between copper sulphate and iron shavings in a TLS about ponderal laws.

By presenting a scientific phenomenological context associated with low cognitive demand in an open-ended question, it can be established that the tasks have low modelling potential (Cortés-Morales & Marzabal, 2025a). This low potential is related to a selected context that is distant from the students' experiences and knowledge, so an extensive familiarisation process may be required to enable students to relate what they know about the phenomenon.

Strategies For Data Collection And Analysis

The video recordings were made following the analytical strategy proposed by Lehesvuori et al. (2018) to analyse teaching processes. This meant using a camera positioned at the back of the classroom to record the teacher's interactions with the whole group and a mobile camera to record the teacher's interactions with the small groups. This meant that a camera and microphone followed the teacher to record the conversations.

A coding scheme was developed to examine the talk moves used by science teachers to support their students' modelling; the analysis required arranging the codes to visualize the discourse. Previous work on modelling conversations made by Tytler & Aranda (2015) and Vergara-Sandoval et al. (2025a) led the researchers to make adaptations in their proposals in order to consider how the nature of the talk move could change the purpose of the discourse action taken by the teachers. Consequently, this tool for characterising conversations considers the purpose of the talk move and its nature depending on how the move is framed. In this sense, a distinction was made between the cognitive demand imposed by the teacher and whether that demand entails a process of requesting or providing ideas for constructing an SSM (Table 1).

The dialogue between teacher and students during a task with modelling potential was coded according to the following aspects of discourse:

a) Communicative act. The identification of turns of speech, differentiating between student and teacher turns, and communicative intentions to differentiate one turn among several together (Hennessy et al., 2016)

b) Typology: The distinction between productive conversation and general classroom conversation (Soysal, 2019)

c) Organisation: The organisational difference between students in conversation, i.e., interaction with the teacher in a small or large group (Alexander, 2018).

d) Identification of the phase of the modelling cycle that teachers promote with their discourse: familiarisation, discussion, and consensus (Garrido & Couso, 2024).

e) Characterisation of the type of interactions in Table 1 (next page).

This qualitative research proposes analysing data to visualise the discourse led by teachers through their turns of speech (Category A), which are framed by the organisation of the students with whom they interact (Category C) in a phase of the modelling cycle (Category D). These turns are differentiated by the type of conversation (Category B) and the type of interaction the teacher promotes (Category E).

For this, in the first instance, how teachers interact with their students during the same task was characterised and presented in the visualisations by their location on the axis. In a second instance, the purpose of the interactions was characterised by the sub-code indicated in the turn-of-speech box (P: Provide; R: Request). The discourse display is a chronological map of the teacher's talk moves as they implement a common task.

Table 1. Talk moves for modelling conversation (Cortés-Morales et al., 2026).

Talk move	Teacher request…	Teacher provides…
Acknowledge		A word or phrase to signify the value of the student's contribution, to validate or acknowledge a point of view, or time for students to think about what they have expressed. *i.e., ok, very good, right*
Inform	Information to get a first appreciation of the students' ideas *i.e., How are we doing? What happened? What were you able to observe?*	
Clarify	Students to expand on their ideas with a more robust response or to explain what they meant. *i.e., You say it's an oxidation, what was oxidized? And that, to what equate? Ok, we would see particles, particles of what?*	A thought-provoking idea by providing information or an analogy for the revision of an idea about the phenomenon. *i.e., But if it were to dye it blue and we don't see it blue, then maybe the change doesn't go hand in hand with dying it*
Challenge	Students to check the ideas that have been put forward or ask for an interpretation that helps to contrast ideas to elicit an idea when something doesn't fit. *i.e., ok, it melts [repeating student's answer] how so?*	A questioning of the students' ideas and rationales about the phenomenon. *i.e., Do we have the same results if we have different amounts of copper sulphate?*
Consolidate	Students to integrate the ideas discussed into a single answer. *i.e. [To end the task] Is it the same if we have two volumes of copper sulphate in the reaction?*	A synthesis of the students' ideas in a single answer. *i.e., That is also relevant because when I have a chemical change, the changes are important; besides the change in color, we also see a change in temperature.*

Findings

The analysis of teachers' discourse employs visualisations that allow us to observe how discourse and talk moves are structured in a productive conversation for a task with low modelling potential. This first finding yields a methodological tool that enables us to observe the rhythm and consistency of the dialogue in a modelling-oriented classroom. These discourse visualisations (Figure 1) show, on the left, the modelling-oriented talk moves, and, on the right, their uses by a teacher in chronological order. On top, we can examine if the conversation is between the teacher and the whole group (WC), or with a particular small group (G. number). Below the class organisation, we will find the modelling phase that the teachers promote in their discourse, colour-coded: cyan for familiarisation, blue for discussion, and purple for consensus. The descriptions of the expected conversations for each modelling phase follow the studies by Garrido

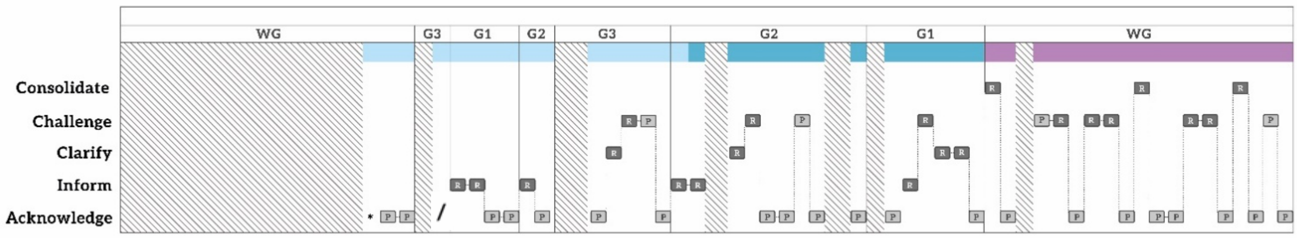
& Couso (2024) and Cortés & Marzabal (2025b). In a familiarisation phase, the teachers approach a contextualised problem situation and expect students to recognise that they need a model, express it, and, in their conversations with others, understand that there can be different versions of the model in the classroom. In the discussion phase, teachers ask questions to help students use their ideas iteratively, testing and contrasting them to revise their initial model. In the consensus phase, teachers facilitate a dialogue in which students agree on a final model that emerges from collective understanding.

In the case of the turn-of-speech boxes, we can identify whether the teacher is requesting (R) or providing (P) an idea in the modelling-oriented talk move. They have been coloured as well to facilitate the reading of the visualisation, light grey is providing, and dark grey is requesting.

Finally, every conversation that is not dialogic and interactive as a communicative act (Mortimer & Scott, 2003) is levelled with general classroom conversations and appears in the visualisation as diagonal shading. Every speech turn in a teacher's general classroom conversation has the same width as the turn-of-speech boxes with the talk moves.

In addition to the codes typified in the visualisation, two icons have been used to show how the tasks are presented to students. The asterisk (*) marks the teacher reading the task, in some tasks teachers take a time before reading the task to talk with students about something that they need to be able to complete the task, in such cases as in figure 1, we can see that the task does not start with the reading of the task, but with an explanation about how to use a balance. Another icon is a slash (/) that indicates turns of speech in which the teacher repeats the task with some alteration, providing a reformulation or simplification of the task instructions to the students.

Figure 1. Visualisation of Marina's talk moves in the task.



Science Teachers' Discursive Practices In Instructional Modelling

The analysis of the visualisation for the three chemistry teachers allows us to examine how the productive conversations unfold in instructional modelling. The findings show patterns of the dialogue that promote students' revision of their SSM.

Firstly, most of the speech turns relate to productive conversations. Although most of the modelling conversations are productive conversations, some of them are general classroom conversations that provide students with the learning opportunities they need to refine their initial model. This means that, in implementing tasks with modelling potential, an alternation between productive and general classroom conversations is required. This shift between communicative acts reflects teachers noticing and acting in accordance with the ideas expressed by their students, a key aspect of modelling conversation, as well as the responsiveness of teachers in the dialogue they mediate.

Similarly, teachers' dialogue involves extensive conversations that promote students' familiarity with the phenomena, accounting for 100% of Caroline's dialogue, 46% of Marina's turns, and 18% of Chris's dialogue. Chris's turns of speech in the familiarisation phase, unlike Marina's and Caroline's, are not grouped together, as each conversation with the small groups starts with familiarisation, shortening the amount of talk moves used in this phase. However, there is a consistency in the phase for this low modelling potential task, which relates to the request for information each time the teachers start a familiarisation dialogue with the small groups.

Secondly, the use of talk moves oriented towards modelling shows that most of the moves have a nature of requiring (R) students' ideas in different patterns. On the one hand, teachers ask for clarification, followed by a challenge to their ideas when they are ending a familiarisation phase or starting a discussion phase. On the other hand, teachers ask them to challenge their ideas, then require clarification during the discussion phase. In the latter, teachers expect students to be clear about their ideas and how they differ from their initial model.

Lastly, there is constant use of validation moves, which are the most used talk moves. Teachers use validation moves in every phase, with every small group, and with the whole group, and they alternate them with other modelling-oriented talk moves. These moves account for 45%, 42% and 20% of the modelling-oriented talk moves used by Marina, Chris, and Caroline, respectively. The repeated use of these moves relates to teachers giving students the opportunity to see their ideas as valid, as a starting point, or as a way to explain the phenomena. They are a transitional move that teachers put in place to motivate students during the process and keep the dialogue going. Even though their use does not promote modelling practice, it allows it to unfold.

Discussion & Implications

The findings in this study lead us to believe that, when implementing the task, the teachers can maintain a productive conversation oriented toward an instructional modelling process that goes beyond the modelling potential. In this sense, Chris, Marina, and Caroline decide to facilitate the modelling process for their students by introducing a recurrent familiarisation phase into whole- and small-group conversations.

As the potential depends on the practice to be developed, in which the description (demand of the task) could lead the teacher to close the conversation in search of a particular description that teachers expect from the students, we can see in Marina's and Chris's cases that the modelling potential increases as the teacher is inviting students to reason out the observations that lead them to a representation or description of the phenomena (Park et al., 2021) instead of searching for a particular answer.

Consequently, regarding the management of modelling processes, the teachers can involve students in productive discussions. The actions of the observed science teacher are fruitful in engaging students in modelling processes. Nevertheless, the same could not be said of Caroline's productive conversations, as she closed the dialogue, ending the task before the conversation reached the discussion phase of the modelling process.

The shifts between productive and general classroom conversations are not interruptions, but rather a key discursive strategy that teachers use to respond to students' emergent needs during modelling. In the same way, each teacher exhibits common patterns but a particular style in managing classroom dialogue, demonstrating a responsive and adaptive mediation that supports transitions between phases of the modelling process, depending on the group they are talking with and the task context. The discourse decisions teachers make to manage and promote the continuation of a conversation relate to understanding when a general classroom conversation will help students in the development of the modelling phase, choosing when validation will secure students expressing their initial representations and when a particular pattern of talk moves will be more appropriate to help them review their model and negotiate the personal ideas in a group setting.

The analysis of the results suggests that mapping these interactions could serve as an analytical tool for discussing and improving teachers' discursive practices, thereby optimising students' intellectual engagement in modelling processes.

Disclosure. It should be noted that part of this proceeding was published in an article in Spanish that can be found at https://orcid.org/0000-0002-3804-7942

Acknowledgement

This research was funded by the grant programme of the Chilean National Research and Development Agency (Agencia Nacional de Investigación y Desarrollo: ANID) under Grant Doctorado Nacional 2021 number 21210524, as well as the Ministerio de Ciencia e Innovación of Spain (PID2022-138166NB-C22b) and carried out within the SGR ACELEC research group, ref. 2021 SGR 00647w.

References

Alexander, R. (2018). *Towards dialogic teaching: Rethinking classroom talk.*

Balyan, R., Arner, T., Taylor, K., Shin, J., Banawan, M., Leite, W. L. & McNamara, D. S. (2022). Modeling One-on-one Online Tutoring Discourse using an Accountable Talk Framework. Proceedings of the 15th International Conference on Educational Data Mining, EDM 2022. https://doi.org/10.5281/zenodo.6852936

Christodoulou, A. & Osborne, J. (2014). The science classroom as a site of epistemic talk: A case study of a teacher's attempts to teach science based on argument. *Journal of Research in Science* Teaching, 51(10), 1275–1300. https://doi.org/10.1002/tea.21166

Cortés-Morales, A. & Marzabal, A. (2025a). The design of science teaching and learning tasks with modelling potential. *Studies in Science Education*, 1–29. https://doi.org/10.1080/03057267.2025.2493990.

Cortés-Morales, A., Marzabal, A., & Couso, D. (2025b). Dialogical modelling processes: conversations for the social construction of scientific models in the science classroom. *Chemistry Education Research and Practice*. https://doi.org/10.1039/d5rp00017c

Cortés-Morales, A., Marzabal, A. y Couso, D. (2026). Movimientos lingüísticos del profesorado en el aula de ciencias orientada a la modelización. Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, 23(1), 1103. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2026.v23.i1.1103

Cui, R. & Teo, P. (2023). Thinking through talk: Using dialogue to develop students' critical thinking. *Teaching and Teacher Education*, 125. https://doi.org/10.1016/j.tate.2023.104068

Garrido, A. & Couso, D. (2024). The IPM cycle: An instructional tool for promoting students' engagement in modeling practices and construction of models. *Journal of Research in Science Teaching*, *September 2024*. https://doi.org/10.1002/tea.21979

Garrido Espeja, A. (2016). Modelització i models en la formació inicial de mestres de primària des de la perspectiva de la pràctica científica. In TDX (Tesis Doctorals en Xarxa). http://www.tdx.cat/handle/10803/399837

Hamnell-Pamment, Y. (2023). Making sense of chemical equilibrium: productive teacher-student dialogues as a balancing act between sensemaking and managing tension. *Chemistry Education Research and Practice*, *25*(1), 171–192. https://doi.org/10.1039/d3rp00249g

Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., García-Carrión, R., Torreblanca, O. & Barrera, M. J. (2016). Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction*, *9*, 16–44. https://doi.org/10.1016/j.lcsi.2015.12.001

Krist, C., Machaka, N., Voss, D., Mathayas, N., Kelly, S. & Shim, S. (2023). Teacher Noticing for Supporting Students ' Epistemic Agency in Science Sensemaking Discussions. *Journal of Science Teacher Education*, 1–21. https://doi.org/10.1080/1046560X.2022.2155355

Larusso, M. D., Al-adeimi, S., Kim, H. Y., Harbaugh, A. G. & Connor, C. O. (2023). Capturing Classroom Talk With the Low- Inference Discourse Observation Tool : A Validation Study. *Journal of Early Adolescence*, 0(0), 1–31. https://doi.org/10.1177/02724316231182291

Lehesvuori, S., Ramnarain, U. & Viiri, J. (2018). Challenging Transmission Modes of Teaching in Science Classrooms: Enhancing Learner-Centredness through Dialogicity. *Research in Science Education*, *48*(5), 1049–1069. https://doi.org/10.1007/s11165-016-9598-7

Newman, R. M. C. (2017). Engaging talk: one teacher's scaffolding of collaborative talk. *Language and Education*, 31(2), 130–151. https://doi.org/10.1080/09500782.2016.1261891

Mortimer E. F. and Scott P., (2003). Meaning making in secondary science classrooms, Open University Press, DOI: 10.1002/sce.20103.

Park, B.-Y., Campbell, T., Kelly, M., Gray, R., Arnold, C., Chadwick, C., Cisneros, L. M., Dickson, D., Moss, D. M., Rodriguez, L., Volin, J. C. & Willig, M. R. (2021). Improving NGSS focused model-based learning curriculum through the examination of students' experiences and iterated models. *Research in Science and Technological Education*, 1–25. https://doi.org/10.1080/02635143.2021.1978962

Soysal, Y. (2019). Indicators of productive classroom talk and supporting discourse moves: a systematic review for effective science teaching. *Academy Journal of Educational Sciences*, *3*(2), 114–137. https://doi.org/10.31805/acjes.642246

Tekkumru-Kisa, M., Akcil-Okan, O., Kisa, Z. & Southerland, S. (2023). Exploring science teaching in interaction at the instructional core. *Journal of Research in Science Teaching*, *60*(1), 26–62. https://doi.org/10.1002/tea.21790

Tena, È. & Couso, D. (2023). ¿Cómo sé que mi secuencia didáctica es de calidad? Propuesta de un marco de evaluación desde la perspectiva de Investigación Basada en Diseño. *Revista Eureka*, *20*(2). https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2023.v20.i2.2801

Tiffany, G., Grieger, K., Johnson, K., & Nyachwaya, J. (2023). Characterizing students' peer–peer questions: frequency, nature, responses and learning. *Chemistry Education Research and Practice*, *24*(3), 852–867. https://doi.org/10.1039/d2rp00146b

Tytler, R. & Aranda, G. (2015). Expert Teachers' Discursive Moves in Science Classroom Interactive Talk. *International Journal of Science and Mathematics Education*, *13*(2), 425–446. https://doi.org/10.1007/s10763-015-9617-6

Vergara, C., López-Simó, V. y Couso, D. (2025a). El ciclo de modelización a través de la interacción oral. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 22(2), 2301. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2025.v22.i2.2301

Vergara, C., López-Simó, V. y Couso-Lagarón, D. (2025b). Dialogar en un aula para construir ideas científicas. *Ápice. Revista de Educación Científica*, 9(1), 21–38. https://doi.org/10.17979/arec.2025.9.1.11311