# PRE-SERVICE SCIENCE TEACHERS' COMPUTATIONAL ARGUMENTATION – CASE STUDY IN AN AUTHENTIC RESEARCH CONTEXT

Peitsa Veteli[1] & Maija Nousiainen[2]

[1] Helsinki Institute of Physics

[2] Department of Physics, University of Helsinki

## ABSTRACT

*Constructing a clear and coherent scientific argument is a non-trivial task. There is an argument to be made that in the modern day, using computational essays and similar media that require reproducible programming as a platform for delivering one's ideas could foster development of good scientific communication practices in a disciplinarily authentic manner. We piloted a short field course for pre-service teachers at the Hyytiälä SMEAR-II station, where they participated in lectures and workshops led by practicing researchers, toured the experimental facilities and made their own inquiries into the data collected therein via computational essays. These essays (N = 6) were analyzed for coherent argumentation using an analysis rubric proposed in this study. The results indicate that the medium is well-suited for presenting such argumentation.*

## INTRODUCTION

Many curricula around the world include a broad objective for educational institutions to foster rational, evidence-based thinking and good argumentation skills in their students, yet what that means is often quite vague even in the research literature. Learning to think critically and analytically is certainly an important theme in the wider trend towards emphasizing scientific literacy as an explicit curricular objective (Virtič, 2022), but the practical side of teaching the students to express such thinking in a clear and cogent manner seems too often fall victim to more pressing content knowledge issues (Nousiainen & Vuola, 2023; Vuola & Nousiainen, 2020). This is a rather problematic state of things, especially in natural sciences, where the role of logical arguments, such as making insightful inferences from observations or posing meaningful follow-up questions based on previous experiments, is fundamental to the nature of the disciplines themselves (von Aufschnaiter et al., 2008). The problem is observably wide-ranging from elementary levels to higher education, which may lead to situations where pre-service teachers with less than satisfactory argumentation skills eventually enter the workforce as in-service teachers with the same less than satisfactory skills

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

being passed onto their younger students (Nousiainen, 2013; 2017). Teacher education programs vary widely in addressing these concerns. Regardless, it can be argued that the aim of producing good teachers requires producing people with suitable didactic knowledge and communication skills to become proficient explainers, which in turn requires proficiency in formulating coherent and persuasive arguments that suit the nature of their chosen fields (Lytzerinou & Iordanou, 2020).

In this study, we are interested in the possibilities of authentic inquiry practices using open data and computational essays as tools and scaffolds for fostering such argumentation in pre-service teachers. This was done by organizing a short, voluntary field course at an active multidisciplinary research facility in central Finland with the intent of immersing the participants in the topics studied there and having them report their reflections in an open-ended, evidence-based computational essay about a topic of their own choosing. It is worth bearing in mind that these pre-service teachers are still themselves students first, although some are already developing an identity as a teacher, and thus a voluntary course aimed at them should always offer them something interesting both from an academic as well as a personal angle. With that goal in mind, we used a multifaceted model of authenticity to offer multiple angles for the participants to engage with in a motivating and inspiring experience that they could link to their educational practices.

## AUTHENTICITY AND EPISTEMIC AGENCY IN COURSE DESIGN

Without delving too deep into the nuances of authenticity as it is understood in educational research, we use the term here to mean a complex combination of factors that enhance the conditions of learning by making the experience more 'real' for the learners. In accordance with the model proposed by Schriebl, Müller and Robin in 2022, we can identify three dimensions of authenticity (disciplinary, real-world and personal), each of which encompasses multiple aspects that can be emphasized in course design by educators to increase the intended authenticity of the learning experience. The effectiveness of such choices is always modulated by each individual participant's subjective perception of each aspect's importance, but certain beneficial trends can be leveraged regardless. While measuring lived and experienced authenticity is hard, pre-service teachers are a somewhat special group in the sense that they are required to carry out more reflective assignments during their studies than their fellow students in pure natural sciences. One can thus make a reasonable assumption that, given a suitably open prompt to reflect on why they chose a particular research topic and what implications they might see in it as upcoming teachers, there is a good chance of finding signs of personal meaning in the answers that would also indicate instances of epistemic agency in ways that are typical of scientific thinking. In this study, epistemic agency is taken to mean capabilities that allow an individual to take personal or collective responsibility for their own learning, such as knowledge creation and transformation (Nieminen & Ketonen, 2024).

In accordance with the Schriebl et al. model of authenticity (2022), the course content leaned into real-world authenticity as a motivating dimension, emphasizing such aspects as visiting the research station (learning place and environment), meeting practicing researchers (expertise and enthusiasm of instructors) and seeing the wide variety of research under broader topics like climate change (content related interdisciplinarity). The final assignment brought disciplinary authenticity to the fore, prompting the pre-service teachers to work on open-ended inquiries (complexity) while staying mindful of the methods and limitations inherent to the work they were undertaking (cognitive processes, models and approaches).

## OPEN DATA AS A VECTOR OF AUTHENTICITY

Here we take open data to be information that is openly accessible for the general public, preferentially free of cost. The role of open data for the final assignment was particularly pronounced, as it is a powerful but educationally underused resource with great potential for multiple aspects of authentic learning (Coughlan, 2020). These aspects of authenticity are typically seen to increase the students' interest by steering the educational activities in a direction that resembles 'real' practice of scientific work and posits the students in an active role of knowledge creators, which aligns well with the educational objective of familiarizing pre-service teachers with a pedagogy of 'doing science' rather than 'learning about science' (Miller et al., 2018). As Miller and others point out, while educators want to have their students participate in 'doing science', in practice that often means mimicking what others have decreed fundamental rather than exercising true epistemic agency in authentic, shared knowledge creation.

## COMPUTATIONAL ESSAYS AS AN EPISTEMIC MEDIUM

Computational essays are a modern genre of scientific writing that combines traditional essay prose with digital media, such as diagrams and executable code blocks (diSessa, 2000). They are often used to communicate ideas and arguments between researchers, thus offering an intuitively fruitful tool for fostering similar skills in pre-service teachers. We kept the computational essay assignment intentionally open on this course so that the pre-service teachers could both shape and focus their work in any direction they themselves wished to take it while offering expert advice for the participants to engage with. A concern raised in the Miller et al. 2018 study was that without explicitly resolving tensions between the educators' pre-selected methods and the active epistemic needs of the students, it is likely that educational institutions keep implementing practices and environments that mainly situate the students as passive receivers of facts. Computational essays can be seen as a promising method of clearly putting the students in control of their actions, supporting their ownership of the learning process and generally promoting their epistemic agency (Odden, et al., 2023). Incorporating such tools into teacher education can thus be seen as an attempt to mend this perceived contradiction by granting them access to authentic research methods and encouraging personal epistemic action within suitable structural scaffolds for the intended learning experience. Importantly for future teachers, this would also support their personal experience with meaningful use of digital tools and

prepare them to better participate in the wider discussion relating to the teachers' roles in a digitalizing society (Korhonen, et al., 2021).

As has been argued by diSessa (2018), we are living at a time where a new far-reaching computational literacy is possibly forming alongside established foundational literacies such as textual and algebraic literacies that open completely new avenues of expression, communication and creation for those who possess them. Like these traditional literacies that took centuries to implement as widespread societal defaults, computation can be seen as the next revolutionary step that will have profound impacts on how our societies function and how arbitrary inputs can be turned into presentable, persuasive information. Understanding this literacy can be approached by a threefold model of its material, cognitive and social constituents (diSessa, 2000), of which we will be most interested in the cognitive and social parts. In brief, the cognitive pillar encompasses domain-specific elements of applying computational skills to solve questions relevant to a particular field, like leveraging computation in order to learn more about the physics governing certain phenomena. The social pillar then relates to how the participants communicate about their code, its structure, purpose and results. Like writing a persuasive essay through traditional text requires the writer to understand rhetoric and grammar, so too does writing a persuasive piece of computation-based media require a certain level of mastery over such elements as sourcing, methods and narration of the computer-assisted analysis.

## ARGUMENTATION AND COMPUTATION IN SCIENCE EDUCATION

The field studying argumentation is wide-ranging: there are different studies which pay attention to, for example, argument structure, conditions for argumentation or possibilities of argumentation (Wohlrapp, 2014) but nevertheless, many researchers agree that learning argumentation and its skills are central goals for science education (Fischer et al., 2014; Rapanta et al., 2013). Argumentation skills include (among other things) abilities to acquire new scientific knowledge and to engage in scientific debates and discourse (von Aufschnaiter et al., 2008). Pre-service teachers need those argumentation skills in order to deliver coherent explanations and logical inferences in their future teaching. Our previous studies have indicated that pre-service teachers need scaffolding and exact teaching on argumentation. If the argumentation task is structurally open, argument structure rarely becomes coherent (Nousiainen & Vuola, 2023). Therefore, in this study, we choose to apply computational essays as the form through which the pre-service teachers would deliver their arguments in order to see whether the medium can assist the argumentative process.

As a platform for scientific communication, computational essays have several virtues (Odden et al., 2023; Odden & Malthe-Sørenssen, 2021) and synergize strongly with the aspects of authenticity in science education promoted by our course design (Schriebl et al., 2022). They are frequently used by active scientists, linking the pre-service teachers immediately to the methods used in the field, in this study's case mainly climate-related physics. Creating a computational essay is authentic scientific work by itself, where domain-specific thinking is explicated

and algorithmically partitioned to a readable and repeatable form that others (including the computer) must be able to understand, avoiding the black box problem where a reader cannot necessarily reliably follow the steps presented in a traditional report. How the writer approaches this is flexible, allowing for epistemic freedom in pursuing and presenting the results in a manner that they find meaningful. In our case, while the assignment itself was left open, the participants were presented with a handful of premade examples that loosely followed a simple 'introduction - question – data – analysis – discussion' structure.

**Identifying argumentation in computational essays: a new scoring rubric**

To analyze the argumentation presented in the pre-service teachers' essays, we followed a four-step model for physics knowledge argumentation structure proposed by Nousiainen and Vuola (2023) to locate coherent arguments. In the previous study, it was proposed that a coherent argument should consist of the following structure: 1. Background for argument, 2. Assertion, 3. Inferences and 4. Conclusions. In this study, we elaborate the argumentation analysis to cover not only episodes of coherent structure but also the quality of the argumentative content. This is done by extending the analysis of each step in a coherent argument by two overarching key features and assessing their importance throughout the whole product. The extension of the argumentation analysis criteria has its roots in the ideas Odden et al. (2023) present in their scoring rubric for instances of epistemic agency in computational essays.

We propose that argumentation in computational essays can be analyzed by the following criteria (see Table 1) via a three-step grading system. Level 0 rating corresponds to no meaningful presence of the criterion. Level 1 rating corresponds to a basic level of competency or engagement necessary for a passing presentation. Level 2 rating corresponds to an advanced effort on the part of the writer beyond the baseline examples given to them.

Table 1. Coding of instances that display argumentative qualities in the essays.

| Thematic category | Code | Level 0 (none) | Level 1 (basic) | Level 2 (advanced) |
|---|---|---|---|---|
| Background | B | How the participants prepare their arguments and convey this through the essay. | | |
| Research question | B1 | Undefined or unclear. | Basic or derivative area of interest. | Novel or complex question incisively presented. |
| Narration | B2 | Progression of inquiry is left uncommented or implied. | Progression of inquiry is adequately presented to the reader. | Progression of inquiry is well documented and serves a purpose in elaborating the thesis of the text. |
| Assertion | A | How the participants present their main arguments and assert their views through the essay. | | |
| Relation to data | A1 | Assertions are not based on the materials. | Assertions relate to the chosen materials directly and shallowly. | Assertions relate to advanced or deduced meanings of the data being presented. |
| Inquiry | A2 | Essay displays a "cook book" style of minimum effort. | Essay displays adequate understanding and ownership of the research process. | Essay presents a process of personal and iteratively deepening path of discovery. |
| Inferences | I | How the participants' arguments relate to existing disciplinary traditions and conventions. | | |
| Justification | I1 | No reasoning is given for decisions made in the research process. | Methods and terms are used purposefully when suited for the task. | Incisive or thoughtful reasoning is presented for decisions made in the research process. |

| Thematic category | Code | Level 0 (none) | Level 1 (basic) | Level 2 (advanced) |
|---|---|---|---|---|
| Disciplinary understanding | I2 | Essay bears no discernible relation to conventions common in natural sciences. | Essay bears similarity to how things are done in natural sciences. | Essay presents an insightful picture of someone who has internalized or is familiar with conventions of the field they are writing about. |
| Conclusions | C | How the participants situate their work in wider contexts. | | |
| Implications | C1 | Essay has no concluding remarks. | Essay discusses basic connections or wider meanings of its topic. | Essay presents insightful broadening of the findings in relation to larger questions. |
| Reflection (teacher context) | C2 | Essay presents no meaningful self-awareness. | Essay recognizes personal or educational connections. | Essay presents personal relevance and meaning in the research process itself beyond mere topical connections. |

## RESEARCH QUESTIONS

We study the quality of argumentation that pre-service science teachers convey in the form of computational essays. In the analysis, we follow the analysis rubric introduced in this study. The research question is:

How the proposed analysis rubric can be used to discern argumentative differences in pre-service science teachers' computational essays?

We aim to describe how the pre-service teachers use the medium of computational essay to study a topic of their choice and how well they make use of it structurally to offer a persuasive argument. If the proposed rubric can effectively discern strong or weak argumentation from these essays, it could be used further to support this facet of teacher education.

## RESEARCH DESIGN AND METHODS

Our sample consists of computational essays (N=6) which were produced as a course task. The course was held over Nov.-Dec. 2023 and marketed towards undergraduate level pre-service science teachers at University of Helsinki. It consisted of a preparatory lecture, a three-day excursion to Hyytiälä SMEAR-II station, a post-excursion meeting and a final computational essay assignment. The essays were written on Jupyter Notebooks in Python using openly available data. The course was intentionally relaxed and open-ended in nature, with some prepared expert lectures and visits at the research site but plenty of room for discussions and debates on topics that the students expressed interest in. Hyytiälä is a multidisciplinary site where research is carried out in atmospheric physics, aerosols, agricultural forestry, biological fluxes and similar topics. Organizers came mostly from a physics background while most of the participants had mainly studied chemistry and had little to no programming background. The concept of a computational essay as presented here was an unfamiliar one to all participants. After the excursion, the group met again two weeks later and each student gave a short talk on the research problem they had chosen for their essay, followed by discussion and comments with experts and their peers. The final essays were turned in two weeks after the follow-up meeting. No grading was given for the coursework. Though each one of the six analyzed essays was turned in by a separate participant, there were more pre-service teachers participating on the course and most worked collectively with their peers during the excursion.

The data was analyzed according to the suggested criteria using a scale of 0–2. Two physics education experts analyzed all reports independently to ensure the credibility of scoring. After the first reading, slight changes were made to the criteria to better fit the material at hand as agreed on by the scorers. The inter-rater agreement between the scorers was 95.5 %, indicating that the scorers had a very high degree of agreement. Given the small number of participants, the criteria are intentionally broad in order to accommodate qualitative notes on the differences between each essay. While the four-step method of analyzing coherent arguments was previously utilized with regards to consecutive sentences, we now applied the criteria over the whole essays, looking at the structure and content of the presentation from the research question to its conclusions.

## RESULTS

The essay scores are shown in table 2 and elaborated further in this section. Based on the scores received, the essays fell into three emergent groups of structural argumentative merit: weak (zeroes in multiple categories, low overall points), adequate (no more than one zero) and commendable (high overall points, focused text). While the number of essays was low, it was clear that the weaker essays were lacking (scoring zeros) in categories that were more prominent in the better essays and the distinction can be generalized. Overall, the participants were good at setting up non-trivial research questions and building their arguments on top of the data they were analyzing but did not present a particularly advanced appreciation of clarifying their own decisions made during the inquiry process.

Table 2: Results of the analysis for each of the six participants (P).

| Category | Code | P1 | P2 | P3 | P4 | P5 | P6 | Mean |
|---|---|---|---|---|---|---|---|---|
| **Background** | | | | | | | | |
| Research question | B1 | 1 | 2 | 2 | 2 | 2 | 2 | 1.83 |
| Narration | B2 | 0 | 2 | 2 | 1 | 1 | 0 | 1 |
| **Assertion** | | | | | | | | |
| Relation to data | A1 | 1 | 2 | 2 | 2 | 2 | 1 | 1.67 |
| Inquiry | A2 | 0 | 2 | 2 | 1 | 1 | 0 | 1 |
| **Inferences** | | | | | | | | |
| Justification | I1 | 1 | 2 | 2 | 1 | 1 | 1 | 1.33 |
| Disciplinary understanding | I2 | 1 | 2 | 2 | 1 | 1 | 1 | 1.33 |
| **Conclusions** | | | | | | | | |
| Implications | C1 | 1 | 2 | 2 | 1 | 1 | 1 | 1.33 |
| Reflection | C2 | 1 | 2 | 2 | 1 | 1 | 0 | 1.17 |
| **Total (X/16)** | | 6 | 16 | 16 | 10 | 10 | 6 | |

**Essay overviews and open data sources**

All six works tapped into the publicly accessible SMEAR database that is also used by active researchers in their associated fields. Most participants used one or two measuring stations or instruments as a basis for their arguments and findings. Chosen variables often included atmospheric temperatures, carbon dioxide fractions, ecosystem fluxes and aerosol particles.

P1 wrote a sparse report with no frills or ambition. The topic was climate science.

P2 wrote a well-executed report that delve into a relationship between two variables. The topic was biology. Report also used data from the national weather service archives.

P3 wrote a well-executed report and self-aware inquiry from a clearly grounded teacher perspective. The topic was climate science. Report also used data from the national weather service archives.

P4 and P5 wrote basic essays attempting to ascertain news or factoids. The topic was climate science in both.

P6 wrote a sparse report, that was very roundabout and lacking reflection. The topic was climate science.

**Background - B**

B1 describes the framing of the essay topics. It was the strongest category that emerged. Participants 2-6 set themselves questions that required either comparisons between multiple datasets or set out to personally ascertain something interesting that they had heard in passing during the course. In the pre-service teacher context, the latter could be argued to be of particular interest for epistemic reasons and supporting critical thinking in their students. For example, P4 got curious about an off-the-cuff remark that you could spot volcanic eruptions from Hyytiälä's data and set out to hunt signs of a 2010 eruption in Iceland from there. P1, faring weakly throughout, settled for a trivial comparison between two stations' radiative measurements.

B2 describes the way that the essays present the writers' thought process to the reader. This is of particular interest with the medium of the computational essay, as it does present the reader a replicable document rather than a purely pre-processed product. Together with the A2 category, this area differentiated the three groups most cleanly from each other. While a rich, considered text uses the essay's structure to comment on each part's purpose and why a particular step is taken to enhance the validity of their resulting assertions in persuading the reader, a weaker text leaves most or all of that out in a black box way of jumping over the fundamentals. This difference in communication is something that has been previously discussed as well (Odden, Lockwood & Caballero, 2019), with a gulf between those who skip over the technical parts to get to what they see as the main content and those who take their time in explaining what it takes to get there.

**Assertion - A**

A1 describes how the presented arguments relate to the data behind them. Second strongest category. A scientific argument is nothing without the data to back it up, but likewise merely stating what you can see in the shape of a plot is at best a first step towards understanding the phenomena being observed. In the weaker group, P1 and P6, we can see this exact problem emerge as any assertions made are very surface level descriptions of obvious things or vague musings on the original topic based on previous knowledge. P2-5 show better usage of the data by creating plotted graphs and referring to them right after, giving additional contextual explanations for broader thinking on the topic as they move on.

A2 describes the epistemic agency of the writers and their ownership of the inquiry process. Together with B2, this category most starkly differentiates the three groups along the same lines. The weaker group, P1 and P6, presents very little work beyond initial surface level analysis and thinking beyond that minimum achievement. In the adequate group, P4 and P5, the inquiry is cut similarly short after the initial plots and answers are produced, but the text goes somewhat further into what experimental considerations have led to those results or what they represent. In the highest achieving group, P2 and P3, there is clear iterative progress from one step to another as initial results guide further analysis steps towards a more suitable and persuasive discussion of the original research question. In addition, unsurprisingly coinciding with high scores in narration, this

group demonstrates personal understanding of the steps involved and their limitations while expressing their own growing appreciation for the research process as a whole.

**Inferences - I**

I1 describes any reasoning presented for choices made, such as analysis methods, mathematical functions or selected variables. While all participants used methods that fit their topics, like averaging longer time-series into forms that better represented their intentions, only P2 and P3 took the time to properly explain these explicitly. For the purposes of argumentation, such explication both strengthens the validity of the assertions given and works to form a more coherent narrative throughout the text, helping the reader to process it all.

I2 describes the presentations' relation to the wider scientific endeavor, in form and content. As above, all essays followed a discernible structure familiar to anyone who has read research papers in natural sciences but only P2 and P3 displayed deeper mastery of their chosen fields' terminology or content knowledge in a way that noticeably guided the essay work from intuition to hypotheses and onwards. Having an example essay at hand undoubtedly scaffolded the work along, providing a firm but malleable armature even for the weaker essays.

**Conclusions - C**

C1 describes the writers' ability to tie their thoughts and results to a broader discussion beyond their immediate findings. Overall, this tends to be shallow. The commendable group manages to situate their texts better in a rich discussion without losing focus on their original topic, while the rest are going for two distinct filler techniques: pondering what other variables might have been interesting to look at in order to better understand their chosen topic or listing slightly connected topics influenced by theirs. Concluding an essay like this with something that isn't necessarily even all that profound, but clearly demonstrative of understanding the implications of the phenomena just analyzed, is evidently a non-trivial task.

C2 describes the writers' reflective capabilities in motivating the research work, the topic and the lessons learned from the whole process. Though we did suggest the pre-service teachers to use this exercise as an excuse to think about their own educational practices, only P3 made a solid effort on that. P2 made a commendable display of elaborating on their personal interest but disregarded the educational side completely. P6 included no reflection at all, treating the essay as a regular student would a weekly assignment: purely topic-oriented, a question and an answer. It is interesting to see the original hypothesis of more forthcoming reflection from the pre-service teachers fail, possibly due to a lack of continuously expressed demand for it or by the low-stakes nature of the course causing the participants to treat it as a less important part of the exercise.

## DISCUSSION AND CONCLUSIONS

### Findings in the context of computational literacy and epistemic agency

With regards to diSessa's three-pillared model of computational literacy (2000), practitioners often see the social pillar as a neglected facet that gets left on the wayside. In this study, we have proposed a way to analyze computational essays since they are suitable tools to communicate about the research process and their results. Such an approach is rarely in the focus when compared to developing coding practices or content knowledge. While there are undoubtedly reasons for this, such as curricular focus on regular physics or data science courses, it could be argued to be a significantly different case in the educational field.

Epistemic literature places a great emphasis on the community where knowledge is created (Stroupe, 2014). In this literacy framework, each community eventually develops their own 'literature', a set of ways to use their tools. This process is inherently a social one and guided by the nature of any given community. One could make a naive hypothesis on upcoming teachers possibly emphasizing narrative parts over technical details, but the stark differentiation between the three groups found in this study highlights the results of many previous studies (Nousiainen & Vuola, 2023; Odden et al., 2019; Virtič, 2022): students need explicit, clearly defined guidance for developing particular practices in any context. Developing and nurturing a literature of educationally beneficial methods and carrying those over into the teacher education programs is a necessary step towards a sustainable production of computationally, and thus nowadays scientifically, literate future teachers. A physicist will use computational materials differently from a mathematician, who will use them slightly differently from a person teaching those subjects, but the teacher in particular needs to understand and be able to explicate why the others do what they do to communicate the meaning of that work for a prospective student.

As argued by Nieminen & Ketonen (2024), our current higher education systems often use assessment methods that commodify learning into a performative product, something that positions students as customers rather than creators of shared knowledge. We argue that this is a precarious state of being, especially for pre-service teachers, who ought to be able to understand and promote student-centric engagement in the scientific endeavour. In the age of open data and accessible programming through communicative media such as computational essays, we see underutilized benefits for nurturing precisely this type of epistemic agency amongst learners in ways that allow them to better identify themselves as legitimate members of the epistemic process.

### Findings in the context of advancing teacher education

The criteria presented in table 1 were found to be suitable for analyzing essays like this by both scorers in the light of the four-step model previously established (Nousiainen & Vuola, 2023). After the analysis, no glaring omissions stood out either. The selected categories were broad enough to encompass some variation in the approaches taken by the participants but expressive enough to also differentiate between them effectively, as they were meant to do in order to assess the

many facets of cohesive argumentation. Given the high level of agreeability between both scorers' results, the analysis rubric works and can be repeatedly applied to larger pools of participants in the future to assess and hone the argumentative quality of similar works. The rubric can be expanded via further sub-categories as needed.

Previous works have noted that pre-service teachers struggle to create coherent arguments in open assignments (Nousiainen & Vuola, 2023). Most strikingly, such incoherent arguments often lack rational ordering of the steps taken to arrive at the conclusions as well as being overtly reliant on background knowledge over the remaining steps of a persuasive argument. Learning to explicate and elaborate on these steps is not focused on even in higher education, often expected to implicitly develop alongside other studies. Computational essays could offer a useful tool for improving their skills in that regard, as the medium intuitively lends itself very well to presenting arguments in a manner structurally similar to scientific reports, allowing them to focus on the actual content of their assignments. Like Odden et al. (2019), we saw the participants adhering reasonably close to a presented example style, but most still took it as their own and expanded on it, engaging with it in the same way that they took ownership of their research process. We argue that it is important to openly help pre-service teachers develop their skills and knowledge to work in ways that are authentic to their chosen domains as well as conducive to quality argumentation required for effective teaching, which nowadays almost always requires more than a passing familiarity with computational methods as well.

While the course was short and participation numbers small, as is often the case with these case studies, the pilot was otherwise successful and set to continue in 2024. As the intention of formulating a helpful tool for discerning solid argumentation is to assist in improving the communicative power of the participants, it would be interesting to see another similar course apply the rubric presented here midway through the process to further hone in on aiding the pre-service teachers in formulating their thoughts out as cogent arguments.

**REFERENCES**

von Aufschnaiter, C., Erduran, S., Osborne, J. & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, 45(1), 101–131. https://doi.org/10.1002/tea.20213

diSessa, A. (2000). Changing Minds: Computers, Learning, and Literacy. The MIT Press.

diSessa, A. (2018). Computational literacy and "The Big Picture" concerning computers in mathematics education. *Mathematical Thinking and Learning 20*(3), 3–31. https://doi.org/10.1080/10986065.2018.1403544

Coughlan, T. (2020). The use of open data as a material for learning. *Educational Technology Research and Development* 68, 383–411. https://doi.org/10.1007/s11423-019-09706-y

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R. & Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, *2*(3), 28–45. https://doi.org/10.14786/flr.v2i2.96

Korhonen, T., Juurola, L., Salo, L. & Airaksinen, J. (2021). Digitisation or Digitalisation: Diverse Practices of the Distance Education Period in Finland. *CEPS Journal*, *11*, 165–193. https://doi.org/10.26529/cepsj.1125

Lytzerinou, E. & Iordanou, K. (2020). Teachers' ability to construct arguments, but not their perceived self-efficacy of teaching, predicts their ability to evaluate arguments. *International Journal of Science Education*, *42*(4), 617–634. https://doi.org/10.1080/09500693.2020.1722864

Miller, E., Manz, E., Russ, R., Stroupe, D. & Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. *Journal of Research in Science Teaching*, *55*(5), 1053–1075. https://doi.org/10.1002/tea.21459

Nieminen, J. H. & Ketonen, L. (2024). Epistemic agency: a link between assessment, knowledge and society. *Higher Education*, *88*, 777–794. https://doi.org/10.1007/s10734-023-01142-5

Nousiainen, M. (2013). Coherence of Pre-service Physics Teachers' Views of the Relatedness of Physics Concepts. *Science & Education, 22*, 505–525.

Nousiainen, M. (2017). Organization of Physics Knowledge for Teaching Purposes: From Knowledge Justification Schemes to Didactical Schemes. *European Journal of Science and Mathematics Education*, *5*(2), 210–221.

Nousiainen, M. & Vuola, K. (2023). Analysing argumentation episodes: A case study from physics teacher education. *FMSERA journal*.

Odden, T., Lockwood, E. & Caballero, M. (2019). Physics computational literacy: An exploratory case study using computational essays. *Physical Review Physics Education Research,* 15, 020152. https://doi.org/10.1103/PhysRevPhysEducRes.15.020152

Odden, T. & Malthe-Sørenssen, A. (2021). Using computational essays to scaffold professional physics practice. *European Journal of Phys*ics, *42*, 015701. https://doi.org/10.1088/1361-6404/abb8b7

Odden, T., Silvia, D., & Malthe-Sørenssen, A. (2023). Using computational essays to foster disciplinary epistemic agency in undergraduate science. *Journal of Research in Science Teaching*, 60(5), 937–977. https://doi.org/10.1002/tea.21821

Rapanta, C., Garcia-Mila, M. & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research*, *83*(4), 483–520. https://doi.org/10.3102/0034654313487606

Schriebl, D., Müller, A. & Robin, N. (2022). Modelling Authenticity in Science Education. *Science & Education*, *32*(2), 1021–1048. http://dx.doi.org/10.1007/s11191-022-00355-x

Stroupe, D. (2014). Examining Classroom Science Practice Communities: How Teachers and Students Negotiate Epistemic Agency and Learn Science-as-Practice. *Science Education*, *98*, 487–516. https://doi.org/10.1002/sce.21112

Virtič, M. P. (2022). Teaching science & technology: components of scientific literacy and insight into the steps of research. *International Journal of Science Education*, *44*, 1916–1931. https://doi.org/10.1080/09500693.2022.2105414

Vuola, K. & Nousiainen, M. (2020). Physics knowledge justification: an analysis framework to examine physics content knowledge. *Nordic Studies in Science Education, 16*(2), 149–166.

Wohlrapp, H. (2014). *The Concept of Argument*. Springer.