Received: 21 September 2021 | Revised: 10 August 2022 | Accepted: 20 August 2022

DOI: 10.1002/tea.21821

RESEARCH ARTICLE

JRST | WILEY

Using computational essays to foster disciplinary epistemic agency in undergraduate science

Tor Ole B. Odden[1] | **Devin W. Silvia**[2] | **Anders Malthe-Sørenssen**[1]

[1]Center for Computing in Science Education, Department of Physics, University of Oslo, Oslo, Norway

[2]Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, Michigan, USA

Correspondence
Tor Ole B. Odden, Center for Computing in Science Education, Department of Physics, University of Oslo, Sem Sælands Vei 24, 0316 Oslo, Norway.
Email: t.o.b.odden@fys.uio.no

Funding information
Direktoratet for internasjonalisering og kvalitetsutvikling i høgare utdanning; Norges Forskningsråd, Grant/Award Number: 288125

Abstract
This article reports on a study investigating how computational essays can be used to help students in higher education STEM take up disciplinary epistemic agency—cognitive control and responsibility over one's own learning within the scientific disciplines. Computational essays are a genre of scientific writing that combine live, executable computer code with narrative text to present a computational model or analysis. The study took place across two contrasting university contexts: an interdisciplinary data science and modeling course at a large research university in the Midwestern United States, and a third-semester physics course at a large research university in Scandinavia. Over the course of a semester, computational essays were simultaneously and independently used in both courses, and comparable datasets of student artifacts and retrospective interviews were collected from both student populations. These data were analyzed using a framework that operationalized the construct of disciplinary epistemic agency across the dimensions of programming, inquiry, data analysis and modeling, and communication. Based on this analysis, we argue that

computational essays can be a useful tool for fostering disciplinary epistemic agency within higher education science due to their combination of adaptability and disciplinary authenticity. However, we also argue that educational contexts, scaffolding, expectations, and student backgrounds can constrain and influence the ways in which students choose to take up epistemic agency.

KEYWORDS

computational essay, computational modeling, data science, epistemic agency, inquiry

1 | INTRODUCTION

For centuries, science learning was essentially a process of collecting and curating sets of facts (Livingstone, 2010; Rudolph, 2005), and traces of this philosophy still live on in conservative modes of science teaching today (Stroupe, 2014). With the rise of ambitious science teaching, however, this paradigm has shifted toward an educational philosophy that gives students greater control, responsibility, and agency over their own science learning (Miller et al., 2018; Stroupe, 2014; Windschitl et al., 2012). This trend has inspired a wave of science education research on how teachers can empower students to take up this epistemic agency, especially within precollege science education. For example, the field has seen the development of culturally relevant and critical pedagogy (Calabrese Barton & Tan, 2010; Miller et al., 2018), communities of students explaining, debating, and critiquing ideas (Berland, 2011; Berland et al., 2015; Berland & Crucet, 2016; Berland & Reiser, 2011), and, within higher education, a focus on authenticity in scientific laboratory work (Brownell et al., 2012; Cooper, 1994; Holmes, 2020; Holmes & Wieman, 2016, 2018; Kloser et al., 2011; McFadden & Fuselier, 2020). This shift has also been instantiated into curriculum and education standards reform, such as the Next Generation Science Standards (Miller et al., 2018; NGSS Lead States, 2013).

Despite this progress, much of the teaching in higher education STEM remains fixed within a traditional model of instruction that greatly limits the agency students have over their own learning (Apkarian et al., 2021; Henderson & Dancy, 2009; Hora, 2015; Hurtado et al., 2012; Lombardi et al., 2021; Manduca et al., 2017; Norwegian Ministry of Education and Research, 2017; Stains et al., 2018; Teasdale et al., 2017; Viskupic et al., 2019). However, in recent years, a trend has emerged that offers promising new opportunities for helping students take up epistemic agency: the increasing adoption of scientific computation in education (Caballero & Merner, 2018; Young et al., 2019). By *scientific computation*, we mean the use of computational tools (usually computer code) to create and run computational models and data analyses.

We are interested in exploring the ways these computational tools may be leveraged to promote students' epistemic agency. More specifically, we are interested in understanding the opportunities computation may offer for helping students take up epistemic agency within different scientific disciplines; the tools, techniques, and best practices that can help in this endeavor; and the challenges that emerge along the way.

In this article, we report on the development of one such tool, called a computational essay, and its implementation in two contrasting university contexts: an interdisciplinary computational modeling and data science course at major research university in the Midwestern United States, and a third-semester physics course at a major research university in Scandinavia. Over the course of a semester, computational essays were simultaneously and independently used in both of these universities and courses, and comparable datasets were collected from each implementation. Using this data, as well as our own experiences as practicing computational scientists and educational designers, we have operationalized the construct of epistemic agency for computational science education and have used it to analyze the affordances computational essays provide for fostering disciplinary epistemic agency.

2 | THEORY AND MOTIVATION

This study builds off the theoretical construct of epistemic agency, which has garnered increasing interest in the science education research literature over the last three decades. Here, *epistemic agency* refers to the ways in which students, teachers, or social groups take on and distribute cognitive authority—that is, the responsibility for creating and evaluating new knowledge (Calabrese Barton & Tan, 2010; Damşa et al., 2010; Scardamalia, 2002; Scardamalia & Bereiter, 1991; Stroupe, 2014). When students take up epistemic agency, they take control and ownership of their own processes of learning and inquiry, becoming *epistemic agents*: "individuals or groups who take, or are granted, responsibility for shaping the knowledge and practice of a community" (Stroupe, 2014, p. 492). Within science education, granting students epistemic agency involves positioning them as constructors of new scientific knowledge, rather than reproducers of established scientific knowledge, and giving them opportunities to engage in (and shape) epistemic practices that parallel those used by practicing scientists (Hardy et al., 2020; Ko & Krist, 2019; Miller et al., 2018; Stroupe, 2014; Stroupe et al., 2018).

For the purposes of this study, we are specifically interested in how epistemic agency manifests within the context of different scientific disciplines, which we refer to as *disciplinary epistemic agency*. Our conceptualization of disciplinary epistemic agency is built on three foundational pillars of educational philosophy taken from research in science education and the learning sciences.

2.1 | Disciplinary epistemic agency is tied to authentic scientific practices

Drawing on the work of Ko and Krist (2019), Stroupe (2014), and Lehrer and Schauble (2007), our conceptualization of disciplinary epistemic agency is built on a practice perspective of science learning. This perspective posits that learning science involves learning the key theories and principles that scientists use to make sense of the world, the epistemological foundations that they use to evaluate knowledge, social practices for using and communicating ideas, and the tools and resources needed to do scientific work (Stroupe, 2014). Thus, one key element in supporting students' epistemic agency is enculturating them in practices used by actual scientists, such as experimental designs (Brownell et al., 2012; Kloser et al., 2011), aids to measurement (Dounas-frazer et al., 2016; Dounas-Frazer & Lewandowski, 2016), mathematical and statistical models (Watkins et al., 2012), laboratory practices (Cooper, 1994; Dounas-Frazer & Lewandowski, 2018; Holmes, 2020; Sandi-Urena et al., 2012), scientific communication (Blakeslee, 1997; Kloser et al., 2011; Moskovitz & Kellogg, 2005), and scientific computation (Magana et al., 2016; Odden et al., 2019). Ideally, these practices will be acquired within a social community wherein students will be able to share results, questions, difficulties, and critiques with one another, analogous to the professional communities formed by practicing scientists (Irving et al., 2017, 2020; Ko & Krist, 2019; Lehrer & Schauble, 2007; Stroupe, 2014).

2.2 | Disciplinary epistemic agency requires opportunities for scientific inquiry

Our second pillar relates disciplinary epistemic agency to the construct of scientific inquiry. Building on the work of Chinn and Malhotra (2002) and Ko and Krist (2019), we assert that real agency—that is, having some degree of actual control of learning rather than the illusion of control—requires some degree of open-endedness in the problem under investigation (Chinn & Malhotra, 2002; Holmes, 2020; Wieman, 2015). To be epistemic agents, students need the freedom to choose the direction of their project or investigation and change that direction as new information is discovered or challenges arise—making decisions for how to proceed, how to navigate roadblocks, and when to stop or continue (Holmes et al., 2020; Ko & Krist, 2019). They must also have the freedom to solve challenges in different ways depending on preference, knowledge, and skills. Thus, another key element in supporting students' disciplinary epistemic agency is positioning them such that they have opportunities to engage in scientific inquiry, especially on questions that they find interesting and relevant (Chinn & Malhotra, 2002).

2.3 | Disciplinary epistemic agency is tied to shared knowledge creation

Building on the work of Damşa et al. (2010), we see disciplinary epistemic agency as theoretically tied to a knowledge creation perspective, the idea that "learning takes place through collaborative activities whose aim is to create new knowledge through work on shared objects" (Damşa et al., 2010, p. 146). Thus, disciplinary epistemic agency involves students being active participants in a social knowledge-construction process, within which the participants have cognitive responsibility for managing and monitoring their own learning (Scardamalia, 2002; Stroupe, 2014). When taken to its limit, this perspective advocates that students take responsibility not just for their own learning, but also for the decision of what to learn, how they are going to learn it, and how they will judge if they have been successful or not (Stroupe et al., 2018).

In summary, we conceptualize disciplinary epistemic agency in science education has having the following hallmarks: first, that students are making an effort to ask and answer open-ended, personally meaningful questions; second, they are constructing new knowledge using authentic scientific practices; and third, they are evaluating the quality of that knowledge construction in dialogue with others in their local communities (e.g., teachers or students in their classrooms).

We acknowledge that this conceptualization of epistemic agency is somewhat restricted, in that we are taking for granted the goal of teaching students established disciplinary norms, practices, and methods. Critical scholars have pointed out that there is a tension between student agency and external structures like disciplinary norms, as well as curricula and class contexts—the so-called structure-agency dialectic (Gutiérrez & Calabrese Barton, 2015). These scholars argue that equitable science education requires educators to position students in such a way that they can bring in their own ideas, cultures, and interests into science classrooms

(Calabrese Barton & Tan, 2010; Kane, 2015) and either explore them using established scientific methods and ideas or push back on, resist, or transform them (Hardy et al., 2020; Stroupe et al., 2018; Varelas et al., 2015). Most of this literature, however, comes from precollegiate science education. Because our study takes place in higher education, we are focusing on disciplinary epistemic agency rather than this more expansive theory of the structure-agency dialectic.

When comparing precollege and collegiate science education, there are two key differences that affect our theorization of epistemic agency. The first is that at the university level students choose to pursue a particular area of study. This means that most students can be assumed to have some general baseline interest in the subject matter, tools, and concepts that they are learning. The second is that university-level science students are training to become scientific and technical professionals: physicists, chemists, biologists, doctors, data scientists, and so forth. Although many students will not seek jobs directly in their fields, students in a scientific degree program are nonetheless required to learn certain skills, concepts, and habits of mind that underly those professions. These disciplinary norms function as the overall structure within which students are situated. Within these structures, however, there can be significant room for student agency: for example, students applying the tools or concepts they are learning to personally relevant problems, communicating their understandings in a variety of ways to different audiences, or making key decisions in laboratory work (Holmes, 2020; Wieman, 2015).

2.4 | Epistemic agency and scientific computation in higher education

Given the relationship between disciplinary epistemic agency and authentic scientific practice, one might expect agency-based teaching to be the default in higher education STEM. However, in practice, all three of the above-mentioned pillars of disciplinary epistemic agency are often suppressed or absent from higher education science, as can be seen from the continued efforts at large-scale curricular reform across STEM disciplines (Cooper & Klymkowsky, 2013; Henderson et al., 2011, 2012; Thompson et al., 2013) and various national reports (American Association For The Advancement of Science, 2009; Association of American Universities, 2017; Olson & Riordan, 2012). In part, this is a function of the fact that traditional teaching methods, which emphasize rote, closed-ended problems, and lecture-based instruction are still widespread at the postsecondary level (Apkarian et al., 2021; Henderson & Dancy, 2009; Hora, 2015; Hurtado et al., 2012; Lombardi et al., 2021; Manduca et al., 2017; Norwegian Ministry of Education and Research, 2017; Stains et al., 2018; Teasdale et al., 2017; Viskupic et al., 2019). There is also the philosophical challenge that as one advances to higher levels within a scientific discipline, the questions one can ask (and phenomena one can investigate) become increasingly constrained by the theory, tools, and previous work done in that discipline. For example, in more advanced physics courses, it quickly becomes apparent that only a small number of problems can be solved using traditional analytical methods.

For all these reasons, it is perhaps unsurprising that most of the science education research literature on epistemic agency comes from precollege science education (Hardy et al., 2020; Ko & Krist, 2019; Miller et al., 2018; Sikorski & Hammer, 2017; Stroupe, 2014; Stroupe et al., 2018). However, this imbalance points to a significant gap in the literature—there is a great need for research on how the theories, frameworks, and principles of epistemic agency can be implemented within higher education science.

In this study, we are specifically interested in the opportunities for epistemic agency inherent to the use of scientific computing in higher education science (Caballero & Merner, 2018;

Chonacky & Winch, 2008; Young et al., 2019). By *scientific computing*, we mean the use of computational tools to solve scientific problems, as exemplified by the disciplines of bioinformatics, computational physics, computational chemistry, and data science (Denning & Tedre, 2019).

Although examples of scientific computing have been present in higher education science since the 1960s (Blum, 1971; Ellis & Lang, 1965; Wilson & Redish, 1989), recent advances in both technology and pedagogy have made it significantly more accessible and useful in higher education. For example, most widely used programming languages are currently available for free and are frequently updated with new tools, methods, and packages. The advent of cloud computing means that computational simulations can be built and run on most computers, or even mobile devices. Several college-level science curricula have been specifically built around the practices of computational modeling (Chabay & Sherwood, 2007; Irving et al., 2017; Silvia et al., 2019). Thus, with the right preparation, students now have unprecedented access to some of the most cutting-edge tools and practices in modern science.

Scientific computing offers several affordances for fostering disciplinary epistemic agency, as conceptualized above. First, scientific modeling and data analysis are key scientific practices (Denning & Tedre, 2019), and the open-source nature of many scientific computing software packages makes the tools and datasets used by professionals available to students at little or no cost. Second, once one has acquired certain key skills in programming and modeling, computational simulations and data analyses are fairly easy to create and modify. This means that students can choose the degree of complexity or sophistication of projects or investigations, which in turn gives them agency over how they use computation to create new knowledge. Third, many different types of analyses can be done with a relatively small set of fundamental techniques such as numerical integration and regression—for example, numerical integration can be used to analyze equations of motion in physics, rate equations in chemistry, and population dynamics in biology. This means that, once they have learned these techniques, students can use them to ask and answer a variety of different kinds of open-ended questions, potentially driven by their own interests.

However, there has so far been very little empirical research on the affordances of computation for fostering epistemic agency in STEM teaching and learning, regardless of level or discipline—a second significant gap in the literature. Thus, we are interested in exploring the ways computation can be leveraged to create more opportunities for disciplinary epistemic agency in higher education STEM. As both educators and practicing scientists ourselves, we see great opportunity in using computation to help our students gain authentic scientific skills, engage in genuine cycles of scientific inquiry, and create a knowledge-building community.

To explore these possibilities, we have implemented and tested a teaching tool called a computational essay (diSessa, 2000; Odden et al., 2019; Wolfram, 2017) within two separate university contexts, collecting comparable data from both. Below, we describe the key features of computational essays which make them useful tools for supporting student agency and the contexts within which the study took place.

3 | EDUCATIONAL DESIGN AND CONTEXT

3.1 | Computational essays: An emerging genre of scientific communication

Computational essays are a new genre of scientific communication that has emerged within the last two decades, whose key innovation is the blending of narrative text with live, executable computer code in order to present an argument, explain an idea, describe an analysis, or tell a story (diSessa, 2000; Odden et al., 2019; Odden & Burk, 2020; Odden & Malthe-Sørenssen, 2021; Wolfram, 2017). A typical computational essay shares many of the features of a traditional essay: a title, introduction, thesis statement, body, discussion, conclusion, references, and possibly graphics, images, or illustrations. However, in a computational essay, a key piece of the essay's argument comes in the form of blocks of code that simulate a phenomenon or analyze a dataset. The outputs of these code blocks are woven together with explanatory text to form the overall narrative of the essay. An excerpt from a student-written computational essay is shown in Figure 1. We have also provided examples of two computational essays in the Supporting Information to this article.

Computational essays are frequently used by practicing scientists, data analysts, and technical professionals to share results and analyses or illustrate computational techniques (Kluyver et al., 2016; Rule et al., 2018; Somers, 2018). They are often written in computational *notebooks*, programming environments that allow users to combine blocks of executable code with text, equations, images, and embedded videos. One of the most popular types of notebooks is the

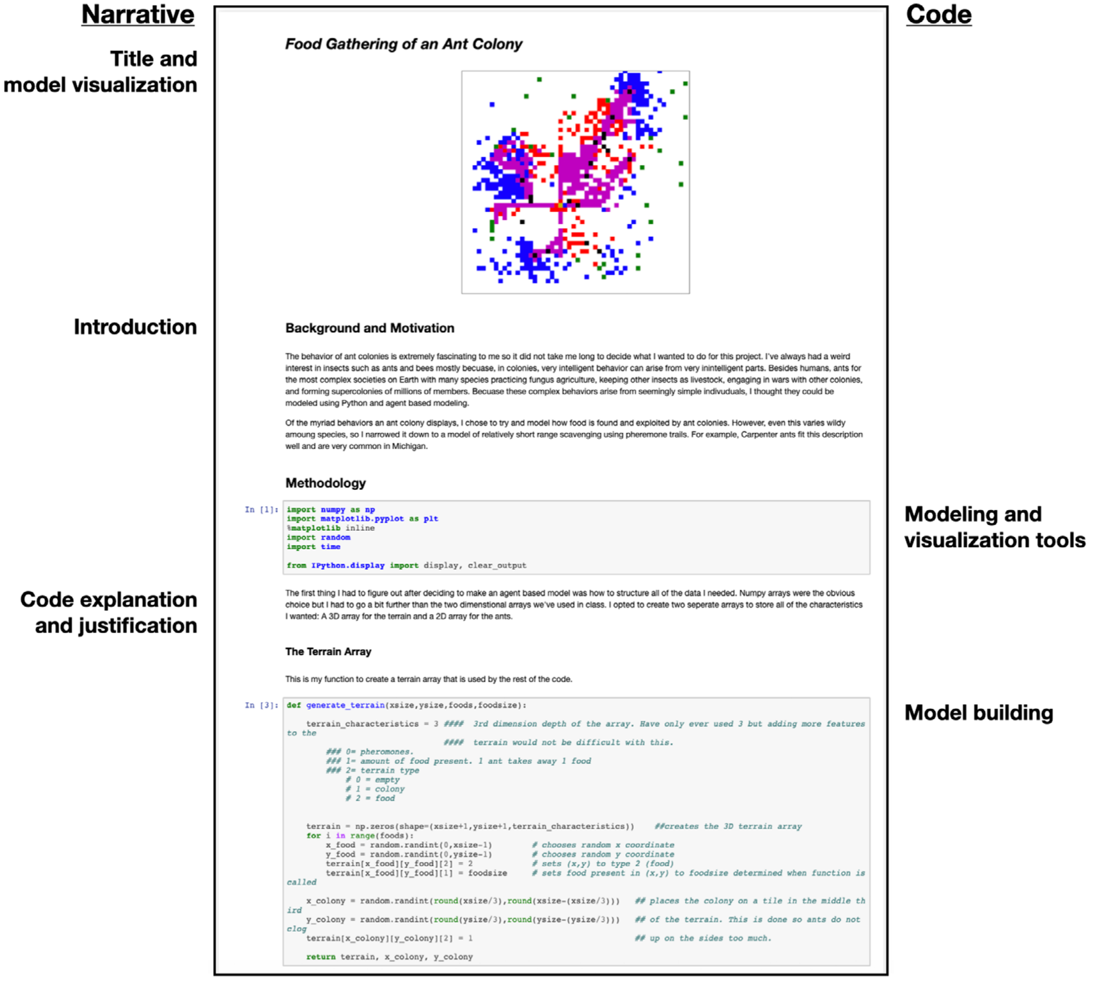


FIGURE 1 Example of a student-written computational essay from an interdisciplinary data science and course

open-source Jupyter notebook (Kluyver et al., 2016), which supports over 40 programming languages and was the programming environment used for all computational essays included in this study. Other tools that support computational essays include Mathematica (Wolfram, 2017) and the MATLAB Live Editor, both of which are available under commercial license.

Computational essays, by their nature, provide several affordances for supporting disciplinary epistemic agency, on top of the general computational affordances described above. First, the process of writing a computational essay is itself an authentic scientific practice. In order to write a computational essay on a scientific topic or problem, one must use code to perform a computational analysis or simulation (a key scientific practice) and explain the steps, results, and relevance of that simulation in writing (another practice). Furthermore, the goal of a computational essay is to convince a reader of the validity of some argument or analysis, which provides opportunities to develop the practice of scientific argumentation (Berland, 2011; Berland & Reiser, 2011).

Second, computational essays can help scaffold open-ended, inquiry-based work (Odden & Malthe-Sørenssen, 2021). Students can write computational essays on a myriad of topics, potentially based on their own interests or completely novel questions. They also can choose how they wish to answer the question, how to address challenges or roadblocks along the way, and how much evidence to provide in support of their argument. In this way, computational essays incorporate certain elements of writing assignments from the humanities, where students have the freedom to construct an argument and the evidence to support it, while also retaining the replicability of computational science simulations.

Third, computational essays are a genre designed for communication and collaboration, providing an ideal medium for shared knowledge creation. Computational essays allow students to share their ideas, analyses, and critiques in much the same way that professional scientists do (Rule et al., 2018). Because students can collaborate on both computational analyses and writing, they thus provide an ideal "shared object" for the building of new knowledge (Damşa et al., 2010). The portability and modifiability of code also open the potential for students to test, reuse, and build off of each other's work, in much the same way that professional scientists share knowledge in the scientific community.

Despite these many potential benefits, there has as yet been little study on the use of computational essays in education, aside from some exploratory work from physics education (Odden et al., 2019; Odden & Burk, 2020; Odden & Malthe-Sørenssen, 2021). There has also been little empirical research on how computational essays might help foster disciplinary epistemic agency for science learners, regardless of educational level. Our current study aims to address this gap in the literature by empirically exploring the affordances of computational essays for helping students take up epistemic agency. Thus, our research questions are as follows:

1. What opportunities can computational essays provide for science learners to take up disciplinary epistemic agency?
2. What kinds of factors affect student uptake of disciplinary epistemic agency when writing computational essays?

3.2 | Computational essay implementation

Based on these research questions, we have studied the use of computational essays in two separate university contexts: an interdisciplinary computational modeling and data science course from a large research university in the Midwestern United States (hereafter MWU), and a 3rd-semester physics course from a large research university in Scandinavia (hereafter ScU). Both of these courses independently used computational essays during the fall semester of 2019. Although the authors of this article, who are also instructors and designers for the respective courses, had been in contact prior to this semester, at the time of data collection the educational designs (which had been developed during previous semesters) were already established. That is to say, no new interventions or activities were introduced into either course for the purposes of this study, and all data collection took place external to the classroom context.

At the Scandinavian University, computational essays were used as the basis for a mid-semester computational modeling project, situated roughly two-thirds of the way through the semester. The goal of the assignment was to give students the opportunity to try their hand at defining and investigating a problem using computational modeling, then present their results to their peers in a scientifically authentic way: computational essays, followed by oral presentations. Prior to this course, most participating students had taken an introductory course in object-oriented scientific programming, as well as one or more physics and math courses that included a computational modeling component. However, this project was the first time most students had had the freedom to define their own investigation questions.

The assignment itself required students to essentially perform a mini computational research project on a topic of their choosing. Students, working singly or in pairs, were first asked to choose a simulated physics system they wished to investigate, which could be built from scratch, adapted from a previous assignment, or built off of several example simulations provided to the students. Then, they were asked to define a question that they felt they could answer using their computational simulation, including considering any necessary assumptions or simplifications they would need to make to answer this question. Next, they were asked to augment (or build from scratch) a computational simulation to answer their chosen question. Finally, they were required to write a computational essay summarizing their question and results, including the computational code used in their simulation, within a Jupyter notebook, and submit that essay for assessment. These essays were then graded pass/fail using a lenient grading rubric (also provided to students) that evaluated the investigation question, code, physics in the simulation, conclusions, and written report. Immediately prior to submission, students also orally presented their assignments to a small group of their peers in mock research-group meetings.

At the Midwestern University, computational essays were used as the basis for an individual, end-of-semester, project-based summative assessment for an introductory, interdisciplinary data science and modeling course. Students in this course were not expected to have any previous programming experience, and over the semester they had learned basic Python programming skills; how to build, run, and interpret computational models; and how to manipulate, analyze, and visualize datasets. Notably, every assignment in the course was written in a Jupyter notebook, and the course included explicit and implicit instructions on code documentation and notebook use.

The actual assignment at MWU required students to begin by defining a question they would try to answer using the modeling and data analysis techniques they had learned in their course. Due to the interdisciplinary nature of the course, students had no restrictions on the types of questions they could ask, and could draw their questions from any discipline or interest. Next, they were asked to find or gather a dataset (or choose a model) that could be used to answer their question. After performing their analysis, students wrote a computational essay summarizing their questions, methods, and results, which was submitted for assessment.

Assignments were graded using a rubric, also provided to the students, that emphasized computational modeling and communication of questions and results. MWU students also orally presented their project results to their peers, using their essays as the basis for their presentation slides; however, in contrast to ScU, these presentations were also factored into project grading.

As can be seen from these descriptions, the two implementations shared certain key features. Both required students to define a question that would try to answer using a computational model or analysis; create a model that could be applied to their dataset or chosen topic; and use computational methods to answer their question. Both required students to communicate their questions and analyses in the computational essay format, and use that computational essay as the basis for an oral presentation to their peers. Both implementations additionally provided students with several forms of scaffolding in the computational essay-writing process, such as an assignment description, a copy of the rubric used to grade the essays, example essays written by staff or previous students, and help-sessions run by instructional staff. And, both implementations placed significant weight on the processes of inquiry, computational modeling, and scientific communication, rather than canonical correctness of models (which was not included in either assessment rubric).

However, there were also certain key differences in implementation. For example, at ScU all projects were physics-focused, since computational essays were situated within a physics course, and students were explicitly required to use physics principles from the course within their projects. Students at ScU had the option to do the project in pairs, whereas all MWU students completed their projects individually. As compared with MWU, at ScU students were provided significantly more example essays, written by faculty and previous students and posted in a publicly-accessible online showroom (Center for Computing in Science Education, 2019). Whereas MWU students were all required to define novel investigation topics, at ScU students were provided with suggestions for questions to pursue in case they had difficulty coming up with a question on their own. These questions were accompanied by "seed" programs that they could build off if they chose, which consisted of stripped-down simulations of physics phenomena that were specifically designed to run without error but provide limited insight into the physical system in order to give students opportunities for exploration.

At MWU, student backgrounds varied considerably, and so essay topics covered subjects from the natural sciences, social sciences, statistics and mathematics, and beyond. MWU students were provided with fewer example essays but were given a template notebook to build off of which specified the structure of the resulting essay (including headers like Background and Motivation, Methodology, Discussion and Conclusions, and References). Students at MWU were also given more oversight throughout their projects: prior to starting their project, MWU students were required to submit a project proposal describing the question they planned to investigate and were given feedback on the feasibility of their chosen topic. A few weeks before the project deadline, students then presented a project update to their peers showcasing what they had accomplished up to that point. Project updates not only served as a motivator to encourage students to begin work on their projects, but also indic to instructors if students were on a path that might not lead to positive project outcomes. Essays from MWU also tended to come in two "flavors" or sub-genres: *modeling essays*, which used a computational or mathematical model to investigate a physical phenomenon, and *data-driven essays* which used statistical and computational techniques to analyze large datasets, whereas essays from ScU were exclusively in the modeling sub-genre.

Additional details on the course contexts and computational essay implementations can be found in Silvia et al. (2019) and Odden and Malthe-Sørenssen (2021). For the purposes of illustration, we have also included copies of example computational essays from both MWU and ScU, as well as copies of their respective grading rubrics, in the Supporting Information of this article.

4 | DATA COLLECTION AND ANALYSIS

4.1 | Data collection

Based on our research questions, this study used a mixed-methods approach (Creswell et al., 2003) that combined collection and analysis of a large number of computational essay artifacts with in-depth student interviews from both institutions. Our goal with this research design was to explore the uses of these tools from both a "top-down" and "bottom-up" perspective—that is, capture a broad picture of how computational essays were used at the two institutions (artifact analysis, top-down), as well as a nuanced picture of student approaches to and reflections on the task (interview analysis, bottom-up). By combining these two forms of data, we aimed to triangulate the large-scale trends in how students took up disciplinary epistemic agency across the two institutional contexts with the students' own perceptions of and reflections on agency in the task.

Accordingly, we gathered the following forms of data.

Artifacts: Completed computational essays from all consenting students or pairs of students. Forty-five of these essays came from MWU, and 58 from ScU. All MWU essays were in English and done by individual students. At ScU, 49 of the 58 essays were in Norwegian, and 20 of the 58 were from students working in pairs. All essays were anonymized by deleting any names or identifying information. The majority of artifacts collected from MWU students were data-driven essays, while ScU essays were exclusively model-based.

Interviews: In addition to collecting completed essays, we interviewed a subset of students (14 from MWU, 12 from ScU) using a similar interview protocol at both institutions to try to capture students' motivations and inquiry processes in their own words. All interviews were retrospective, held around the time students completed and presented their projects. The protocol included questions about the students' academic and programming background; their comfort with programming; their reflections on the ways programming could contribute to science learning; a description of the way they approached the computational essay writing task as well as their motivation for their project topic and challenges they faced; and their explicit reflections on feelings of creative freedom and ownership over the learning process. All interviews also included a component in which students walked the interviewer through the different parts of their essay (displayed on the students' laptop). During recruitment, we made an effort to capture a wide variety of student backgrounds, demographics, and project topics. All MWU interviews, and two of the ScU interviews, were in English; the remaining 10 interviews were in Norwegian. All interviews were transcribed and anonymized (using pseudonyms), although interviewed students were linked to their specific essays in the artifact dataset.

4.2 | Data analysis

Data analysis for this project proceeded in several stages. We began by creating an analytic framework that operationalized the construct of disciplinary epistemic agency for use in analyzing computational assignments like computational essays. We then qualitatively applied this framework to the computational essay dataset, after which a secondary quantitative analysis was performed on the resulting scores. Finally, we used the results of this analysis to select specific interviews, chosen based on assigned levels of disciplinary epistemic agency, for further qualitative analysis.[1] In this way, our data collection and analysis procedure follows a mixed-methods *concurrent nested design* (Creswell et al., 2003) in which multiple forms of data are collected simultaneously, a particular theoretical perspective drives the methodological and analysis choices of the study, and the quantitative analysis is nested within a larger-scale qualitative design in order to capture different grain-sizes of information.

We elaborate on each of these analysis steps below.

4.2.1 | Analytical framework for disciplinary epistemic agency in computational essays

Our first step was creating an analytical framework for disciplinary epistemic agency, in the form of a codebook that could be directly applied to the collected computational essays and thematically applied to interviews. We based this codebook on the epistemic agency literature, our own design philosophies/decisions in the courses, and our experiences as professional researchers in computational science and data science.

Our explicit goal with this codebook was that it would capture most of the ways in which students *could* display disciplinary epistemic agency in their finalized computational essays. That is, we were aiming to "cast a wide net" in order to capture interesting distinctions and nuances in the ways students approached the task, knowing that different students would prioritize different elements (such as programming, investigation question, modeling, and writing or presentation), and also realizing that the two institutions likely supported students in different ways. For this reason, we designed the codebook on a scale of 0–2, with 0 being deficiency in that category, 1 being sufficient to receive a passing grade according to the respective grading rubric, and 2 indicating that the student had gone above-and-beyond the course expectations and taken up agency within this category. Thus, higher scores were meant to capture increased levels of student effort in different aspects of the process of constructing, evaluating, and communicating knowledge using authentic scientific practices.

This codebook went through multiple rounds of revision, as discussed below. Our final codebook included the following categories (elaborated in Table 1):

Programming and data processing (P): This category focused on ways in which students took up agency through the practice of scientific programming. Specifically, some students spent effort on making their code run efficiently and elegantly (subcategory P1), which is an important practice when writing long or complex computational simulations. Other students made a visible effort to make their code organized, readable, and documented (subcategory P2). Still other students took ownership over the coding process by using programming tools and packages in their projects that had not been covered within their courses (subcategory P3).

Investigation (I): This category focused on ways in which students took up agency through the scientific inquiry process. Specifically, some students spent effort on defining novel questions for their investigations (subcategory I1), which is a critical part of scientific inquiry. Other students made an effort to develop their investigations beyond initial, surface-level results, resulting in multi-step investigative narratives (subcategory I2). Still other students made a clear

TABLE 1 Epistemic agency codebook developed for analysis of computational essays

Code category	Code abbreviation	*Description*	Level 0 (no evidence)	Level 1 (limited/basic)	Level 2 (advanced)
Programming and Data Processing	P	*Students taking ownership of coding and data processing by going above and beyond the default expectations. Focus here is on the blocks of code and their organization.*			
Code efficiency and elegance	P1	*Use of code structures and compartmentalization methods such as functions and classes to make code more elegant and efficient*	Little or no code. Code very inefficient or inelegant.	Only basic/default methods used: variables, loops, calls to packages, predefined functions. Lots of copying and pasting.	Effective and efficient use of functions and classes. Code organized efficiently and elegantly to avoid excess copy/pasting or re-definition.
Code organization and self-description	P2	*Use of notebook cells as an organizational tool to make code more readable, and use of variable names and comments/docstrings to make code understandable.*	Code is all lumped together into one chunk and hard to read. Little-to-no use of comments or docstrings.	Code is split somewhat across the notebook, but still heavily lumped together. Limited-to-moderate use of comments and docstrings.	Code is logically split into digestible chunks, sequentially organized. Comments and/or docstrings are used appropriately and effectively. Variable and function names are logical and communicate the idea behind the code.
New packages and coding tools	P3	*Students using new packages and coding tools that they had to learn on their own, appropriate to the disciplinary focus*	No packages or coding tools used. Bare minimum required to run code (or no code present at all)	Default packages and coding tools used (those featured in the course)	Any new packages and coding tools (including modeling or visualization tools) used, and/or students define/create a new computational package for their project
Inquiry	I	*Students taking ownership over the scientific inquiry process by defining novel research questions, developing investigations in novels ways, and making novel interpretations*			
	I1		Research question unclear		

(Continues)

TABLE 1 (Continued)

Code category	Code abbreviation	*Description*	Level 0 (no evidence)	Level 1 (limited/basic)	Level 2 (advanced)
Investigation question		*Students define a novel research question*		Standard/suggested research question and/or research question is overly broad, not well-defined.	Research question is well-defined and clear in scope. If it is a novel, unusual, or personally relevant research question, this makes up for lack of definition/ clarity.
Investigative narrative	I2	*Students developing their models and/or analyses beyond their initial results*	No investigative narrative; all analysis done in a single step, or investigation is inserted into an otherwise-unchanged set of prewritten code.	Investigative narrative is present, but stops at initial results	Multi-step investigation, including refinements on initial results
Discussion and interpretation of results	I3	*Students making novel interpretations and/or describing new ideas they discovered along the way*	No interpretations	Limited interpretations, students most just restate results.	Students provide a summary of the results of the project. Detailed discussion and interpretation of results. Possible reflection on new understandings gained, connections to the real world, or alignment with initial expectations
Modeling and data analysis	MD	*Students taking ownership of the scientific modeling and data analysis process*			
Model development and data exploration	MD1	*Students taking control of the model construction process and/or the exploration of their data*	Students just modify parameters of a pre-existing model, make the very basic plots of the most easily	Students construct a basic new model with significant resemblance to things they have already done or	Students build a sophisticated new model from the ground up and/or extensively modify or adapt

TABLE 1 (Continued)

Code category	Code abbreviation	*Description*	Level 0 (no evidence)	Level 1 (limited/basic)	Level 2 (advanced)
			accessible variables in their data, and/or consider most surface-level variables.	learned about, and/or make minimal modifications to an existing model. Analysis goes beyond surface level variables, but includes limited data exploration. Some use of more advanced data visualizations to extract additional information.	an existing model. Students visualize the data in unique ways that provide critical insight into their understanding of the data (may include types of plots not covered in the course content). Students do a comprehensive exploration of data, including several different kinds of comparisons, and/or examining controlling variables within the given dataset.
Assumptions and limitations	MD2	*Students considering the limitations of their models or their datasets and addressing them*	Limitations not considered or addressed	Assumptions and/or limitations only superficially considered (just noted/stated), little attention to addressing them or unpacking their implications for the results.	Limitations carefully and thoroughly considered and addressed in the model. Students discuss the assumptions inherent to their model or analysis technique, limitations in their modeling technique or dataset, and/or how they might improve the model or analysis by addressing assumptions and limitations.

(Continues)

TABLE 1 (Continued)

Code category	Code abbreviation	*Description*	Level 0 (no evidence)	Level 1 (limited/basic)	Level 2 (advanced)
External sources	MD3	*Students consulting external sources for inspiration and evaluation of model*	No external sources consulted or cited	Small number of sources consulted (1–3), and/or mostly general-purpose sources (i.e., Wikipedia)	Large number of sources consulted (>3), and/or sources taken from specific domains or literatures (i.e., reports, papers, or articles)
Communication	C	*Students taking ownership over the scientific communication and presentation process, and fully using the communicative capabilities of the notebook for this purpose*			
Code explanation and justification	C1	*Students going above-and-beyond in the narrative text explanations of their code*	Little or no writing, just code	Some explanation of code throughout the notebook, but limited to simple descriptions of what the code does and/or basic comments	Detailed prose, fleshed out and persuasively written throughout the notebook. Students not only explain what code does, but also how it works and why it is necessary (motivates the code).
Writing genre and polish	C2	*Students explicitly trying to write their essays in a particular genre, such as a scientific report or personal essay, and attending to the quality of writing in their essay*	Report is very basic, poorly written, and/or lacking. Little or no writing, and/or prose is limited to very basic explanations of what the students did	Report is adequately written. Writing is mostly limited to explanations of what they did, but students have introduced some structure and organization into their report	Report is well-written and polished. Clear markers that the students have tried to write their essays for in a particular genre/style, such as use of informative section headers for organization
Graphics, images, and illustrations	C3	*Use of graphics, images, and animations to illustrate key pieces of the essay and/or set the tone for the work*	Limited-to-no use of graphics, illustrations, animations, or graphs and plots beyond the most basic graphs needed to communicate	Moderate use of illustrations; visualizations, graphics, and plots; enough to communicate both the project design and the project results, with	Students have put significant effort into the visual design of the computational essays. Frequent use of illustrations, graphics, animations, and plots to

TABLE 1 (Continued)

Code category	Code abbreviation	*Description*	Level 0 (no evidence)	Level 1 (limited/basic)	Level 2 (advanced)
			results. Students have not produced anything new.	visualizations sprinkled throughout the report. Limited consideration of aesthetics or understandability.	communicate ideas, key points, and/or improve aesthetics of essay. Clear thought put into the colors, organization, or other elements to make it more aesthetically appealing and/or understandable.

effort to provide detailed discussion and interpretation of the results of their analyses, including reflections on understandings gained or connections with real-world systems (subcategory I3).

Modeling and data analysis (MD): This category focused on ways in which students took up agency through the processes of scientific modeling and data analysis. Specifically, some students paid explicit attention to the sophistication and realism of their model or analysis (subcategory MD1). These students often produced unique visualizations and/or detailed analyses of their chosen phenomena. Other students paid explicit attention to addressing limitations or assumptions inherent to their modeling or data analysis techniques (subcategory MD2), a key part of any scientific endeavor. Still other students spent effort consulting multiple external sources and using these to improve their models or analyses (subcategory MD3).

Communication (C): This category focused on ways in which students took up agency through the scientific communication process by fully using the communicative capabilities of computational notebooks. Specifically, some students spent extra effort in explaining and justifying the meaning and structure of their code in the text of their reports (subcategory C1). Other students spent visible effort in polishing their written reports, sometimes deliberately writing in a particular genre (like scientific paper or personal essay; subcategory C2). Still other students made a deliberate effort to use graphics, plots, and illustrations in their written reports to either communicate key findings or simply generate visual interest (subcategory C3).

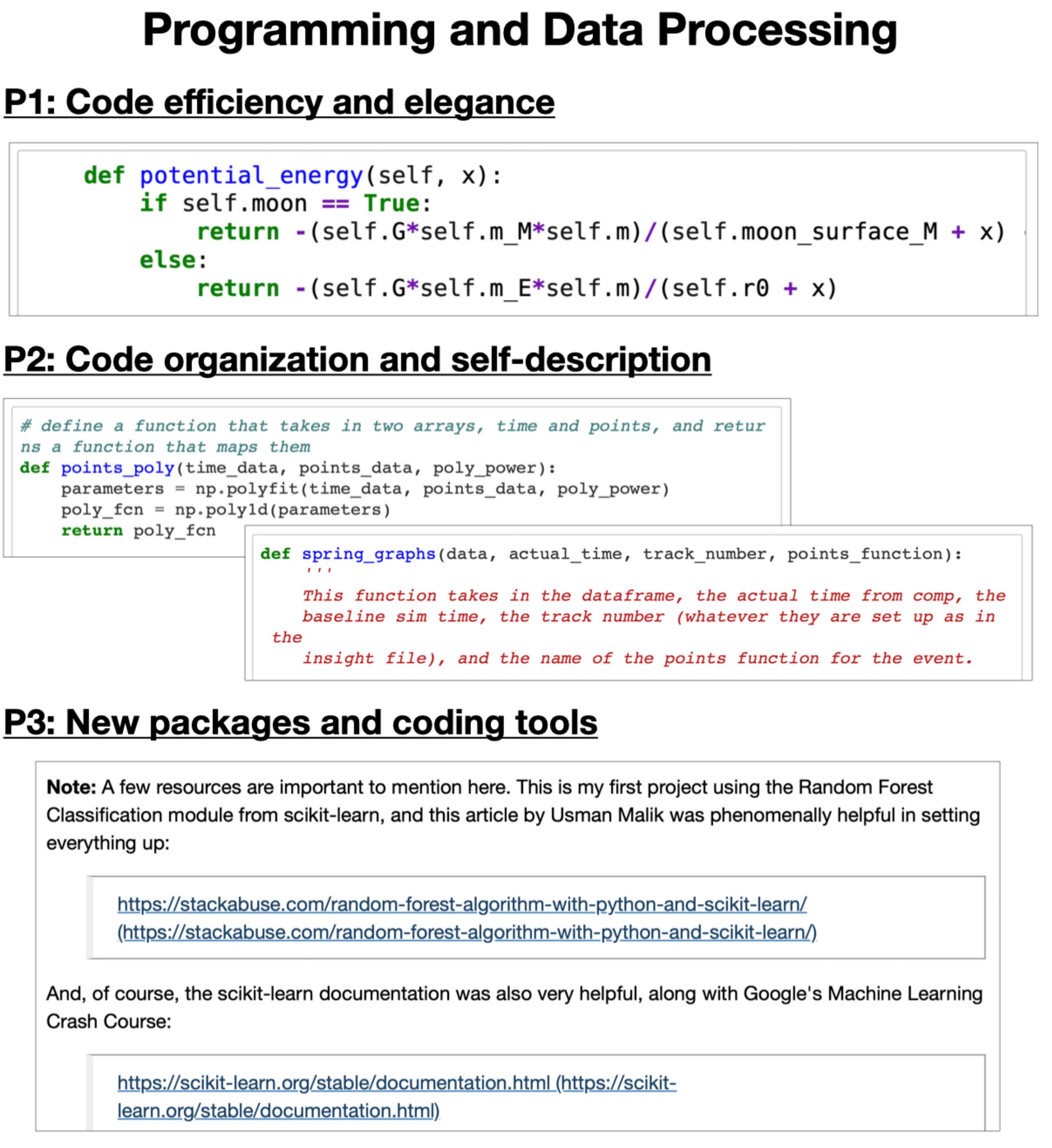
Taken together, these categories and subcategories operationalize the three pillars of disciplinary epistemic agency discussed above within the context of scientific computation. In reference to the practice pillar, all four categories represent authentic scientific practices that are used whenever professional scientists perform and present computational analyses. In reference to the inquiry pillar, inquiry was deliberately included as a category within the rubric, and we paid explicit attention to the ways in which students defined and pursued their investigation questions. In reference to the final pillar of shared knowledge creation, the definition, investigation, and presentation of a novel question to a reader represent a clear example of this kind of social knowledge construction. Finally, across all these categories, the primary focus was on visible markers that students had made a deliberate effort to take ownership over this category, rather than judgments on correctness of contents or practices.

4.2.2 | Application of codebook to student artifacts

After creating an initial draft of this codebook, our analysis proceeded as follows: First, we engaged in iterative coding on a subset of the data, composed of roughly 20% of the essays (19 in total: the 9 English essays from ScU plus 10 randomly chosen essays from MWU) in order to refine our codebook and establish initial estimates of inter-rater reliability. These initial rounds of coding resulted in several small-scale changes and clarifications to the codebook (resulting in the final codes shown in Table 1), and also revealed the necessity of sharing additional information about the course contexts between coders to clarify some of the evaluations made.

For the purposes of illustration, Figures 2–5 show excerpts from computational essays in the dataset that would lead to high scores in each of the 12 categories.

To evaluate inter-rater reliability, we initially used the standard Cohen's Kappa statistic. However, after inspection of the results from the initial rounds of coding, we realized that despite substantial percent agreement (>70%) across hundreds of ratings in different categories, the IRR statistic was low (<0.6). Upon further investigation, it turned out that the value was


Programming and Data Processing

P1: Code efficiency and elegance

```
def potential_energy(self, x):
    if self.moon == True:
        return -(self.G*self.m_M*self.m)/(self.moon_surface_M + x)
    else:
        return -(self.G*self.m_E*self.m)/(self.r0 + x)
```

P2: Code organization and self-description

```
# define a function that takes in two arrays, time and points, and retur
ns a function that maps them
def points_poly(time_data, points_data, poly_power):
    parameters = np.polyfit(time_data, points_data, poly_power)
    poly_fcn = np.poly1d(parameters)
    return poly_fcn
```

```
def spring_graphs(data, actual_time, track_number, points_function):
    '''
    This function takes in the dataframe, the actual time from comp, the
    baseline sim time, the track number (whatever they are set up as in
the
    insight file), and the name of the points function for the event.
```

P3: New packages and coding tools

Note: A few resources are important to mention here. This is my first project using the Random Forest Classification module from scikit-learn, and this article by Usman Malik was phenomenally helpful in setting everything up:

> https://stackabuse.com/random-forest-algorithm-with-python-and-scikit-learn/
> (https://stackabuse.com/random-forest-algorithm-with-python-and-scikit-learn/)

And, of course, the scikit-learn documentation was also very helpful, along with Google's Machine Learning Crash Course:

> https://scikit-learn.org/stable/documentation.html (https://scikit-learn.org/stable/documentation.html)

FIGURE 2 Excerpts from computational essays indicating high levels of disciplinary epistemic agency in programming and data processing

being suppressed by the "paradox" inherent to Cohen's Kappa, in which categories with high agreement will produce low Kappa values (Feinstein & Cicchetti, 1990). For example, categories that produced nearly 100% agreement caused the Kappa statistic to go to 0. For this reason, we instead chose to use a "paradox-resistant" IRR statistic, Gwet's AC2 (Gwet, 2014), benchmarking the degree of inter-rater reliability using the procedure described by Gwet (2014).

Using AC2 as our metric, the first two authors, TOBO and DWS, then repeatedly coded and modified the codebook until we reached "very good" agreement based on this benchmarking

Inquiry

I1: Investigation question

1.0.1 Introduction

This computational essay aims to explore the plausibility of on-site radiation shielding from a magnetic multipolar system. More specifically, the investigation focuses on a system consisting of one or more pairs of Helmholtz coils in different geometric configurations with the purpose of deflecting charged particles on the outside of a given perimeter, and at the same time keeping the magnetic field inside the perimeter close to zero.

I2: Investigative narrative

1.2 METHODOLOGY

We will look at two models that explain how the movement of energy in the body, more specifically in glucose and fat storage with compartmental modelling. One model is depicting the general traditional and more common idea of energy movement in body. The second model will simulate the right movement of energy that is more prominent in the medical field. Lastly, we

1.3 Simplest Model: All Calories Are The Same

In our first step, we will model our problem with these assumptions in our mind: 1. We only consider carbohydrate and dietary fat intake since protein has very little amount of Cal per gram,

1.4 Realistic Model: Carbs Burns First, Fat Burns Later

This model is inspired by Dr. Fung's study of metabolism of human body where the energy movement in body actually is affected by the **presence of insulin** which manipulates the preference of

I3: Discussion and interpretation of results

3 Discussion

In our model, we use an element of randomness to simulate the path the lightning would take. In the real world, we do not have this random factor. We assume homogenous air mass with no wind, pollution or other things that could affect the path of a lightning. Even though a lightning strike looks random for us, it will always choose the path of least resistance.

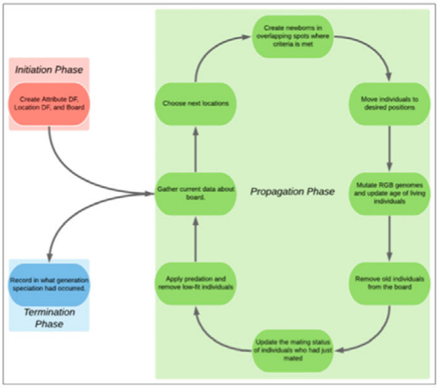
In our simple model, if we would disregard the random element, the highest point would always win, since it covers the shortest distance to the V=1-boundary - the only contributor to voltage. The Poissons'/Laplace equation works like a blur effect, and the shorter distance gives the biggest rate of voltage change, and therefore the biggest voltage different from the lightning. The result: a boring straight line from the rod to cloud.

FIGURE 3 Excerpts from computational essays indicating high levels of disciplinary epistemic agency in inquiry

scale (AC2 > 0.85). Thereafter, the dataset was split based on institution and every remaining essay was independently coded by two researchers, with discrepancies being resolved through discussion. All remaining MWU essays were independently coded in batches by DWS and TOBO. Prior to discussion, percent agreement ranged from 72% to 86% (increasing with each batch) and AC2 scores ranged from 0.87 to 0.9, consistently meeting the "very good" threshold of agreement. All remaining ScU essays were independently coded in batches by TOBO and a

Modeling and Data Analysis

MD1: Model development and data exploration



MD2: Assumptions and limitations

First we will however briefly investigate to what extent the gravity will have an affect on the a particle in a cyclotron. Since the acceleration period of a particle is very short, we do not expect it to have a significant effect. We will for simplicity assume that the gravity is uniform inside the accelerator with a value of $g = -9.81$ m/s^2 in the z-direction, wich form our ordinary experience seems to be a good assumption. The forc

One major problem with our model is our disregard of frictional forces. In a real cyclotron, altough there is a vaccum there will inevitably be frictional forces due to the production of heat in the materials and other interactions. We have also disregarded the fact that in reality the magnetic fields produced by the electromagnets and the electric field over the gap does not necessarily have only one directional componet or beeing completly uniform. With today´s technology this is however not necessarily the greatest concern.

MD3: External Sources

References

Amadeo, K. (2019, November 16). Universal Health Care in Different Countries, Pros and Cons of Each. In The balance. Retrieved from https://www.thebalance.com/universal-health-care-4156211 (https://www.thebalance.com/universal-health-care-4156211)
Berwick, D. M., & Hackbarth, A. D. (2012). Eliminating waste in US health care. Jama, 307(14), 1513-1516.
CDC. (2003, July). National Health and Nutrition Examination Survey. In CDC. Retrieved from https://www.cdc.gov/nchs/data/nhanes/databriefs/adultweight.pdf (https://www.cdc.gov/nchs/data/nhanes/databriefs/adultweight.pdf)

FIGURE 4 Excerpts from computational essays indicating high levels of disciplinary epistemic agency in modeling and data analysis

graduate researcher who had previously taught other computationally-focused courses in the ScU physics department. Prior to discussion, percent agreement ranged from 67% to 75%, with AC2 scores ranging from 0.74 to 0.88, reaching the "good" and "very good" thresholds of agreement.

After all essays had been analyzed, we performed several secondary, quantitative analyses on the resulting codes. These included examinations of overall levels of agency across multiple dimensions: aggregate, split by institution, and split by type of essay (modeling vs. data analysis). This allowed us to evaluate the prevalence of different categories of agency in the data, in

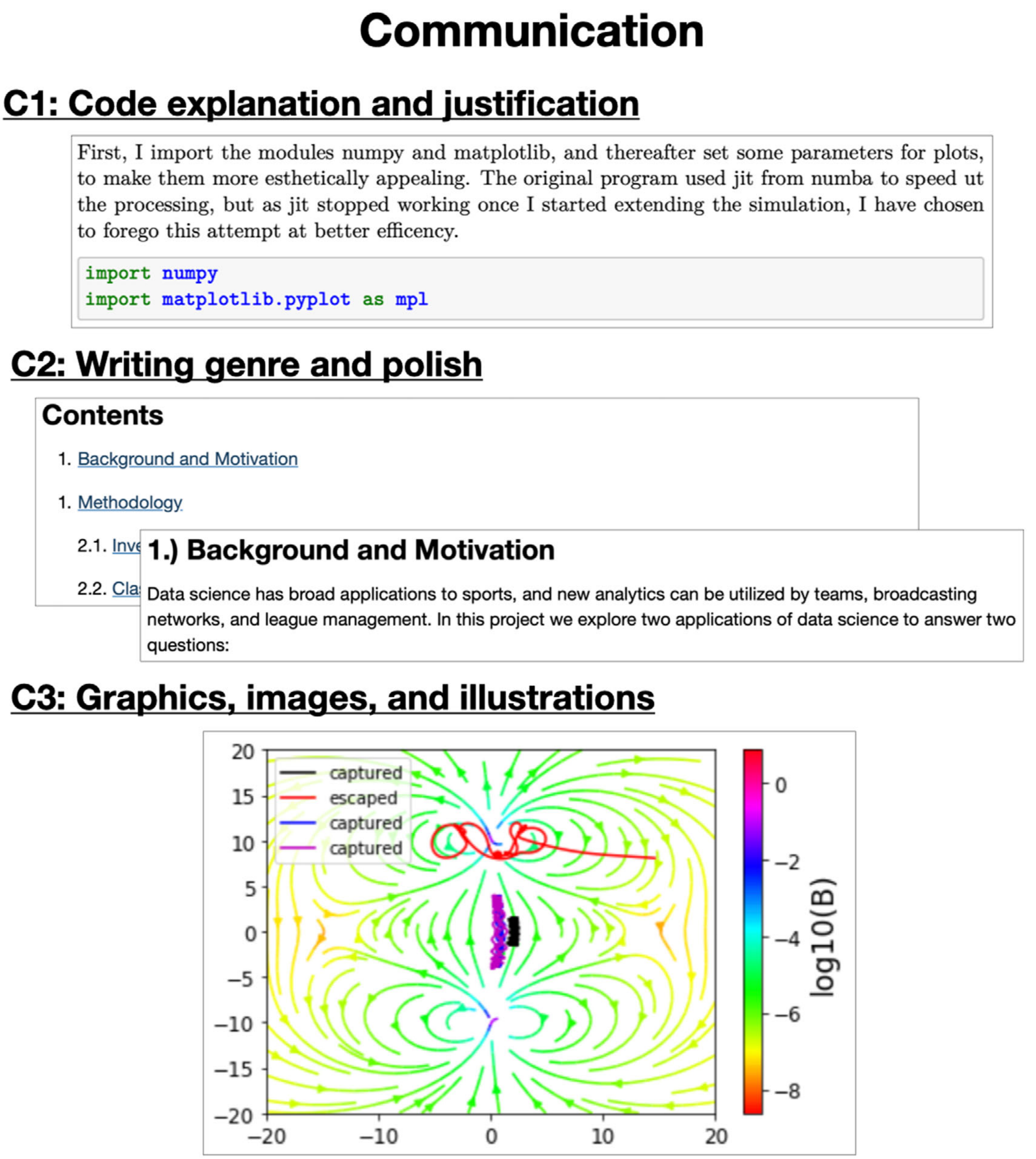


FIGURE 5 Excerpts from computational essays indicating high levels of disciplinary epistemic agency in communication

order to address our first research question by seeing whether certain categories were over- or under-represented. In some cases, the trends revealed by these analyses seemed connected to the specific design decisions made in the computational essay implementations described above. Unpacking these trends helped us address our second research question. The statistical significance of differences between the two institutions was evaluated using Fisher's exact test, which is appropriate for categorical data where certain categories may have low counts.

4.2.3 | Selection and analysis of interview case studies

After analyzing the essays, we triangulated these results with narrative case studies of specific interviewed students in order to both validate our codebook with interview data and provide a more detailed illustration of the different ways in which students demonstrated agency through this project. For this part of the analysis, we isolated the codes for all interviewed students and chose three groups of students who displayed differing patterns of agency uptake (while also aiming to capture a mix of narratives across the institution, gender, and group composition). We then reviewed their interviews and essays to try to reconstruct the story of their essay-writing process and capture their reflections on creativity and agency within the project. Student narratives were compared with the assigned codes for their essay in order to look for points of overlap or conflict between the narratives and assigned categories. The student reflections from these cases allowed us to both gauge the degree of intentionality behind the trends we had noticed in our artifact analysis (further helping address our first research question) and hear students' own reflections on the factors or decisions that led to these trends (addressing our second research question).

5 | RESULTS

5.1 | Overall trends in the computational essay dataset

In this section, we unpack the large-scale trends revealed by this analysis and describe potential connections to essay characteristics and educational design decisions across the two featured institutions. Figure 6 shows a heat map of all computational essay scores, disaggregated by institution and color-coded according to level of demonstrated agency (darker = higher score). Note that in this figure, we have highlighted the three groups whose narratives are featured in Section 5.2.

5.1.1 | Aggregate trends

When we combined all essays from both ScU and MWU into one group, our analysis revealed several interesting large-scale trends and patterns in the way students demonstrated disciplinary epistemic agency. These trends are visible in Figure 7, which shows the number of students in the entire dataset who received scores of 0, 1, and 2 in each category.

As can be seen in Figure 7, in most categories/subcategories the majority of students demonstrated a level of performance that would be considered sufficient for a passing grade (1 or 2). This result can also be seen in Figure 6, which shows that only a few essays demonstrated low levels of disciplinary epistemic agency across multiple categories. This result, however, is not necessarily surprising as both institutions provided students with substantial scaffolding and support, including rubrics that specified the criteria necessary for a passing grade.[2] The only category with a significant number of 0 scores was “External Sources” (MD3); however, this was likely due to the fact that students at the Scandinavian University were encouraged but not explicitly required to cite external sources. This effect can also be seen in Figure 8.

Figure 7 also shows that for certain subcategories (e.g., all of the *Inquiry* subcategories) high levels of student agency were often the default rather than the exception. A qualitative review of the collected computational essays supports this conclusion. For instance, many of the essay

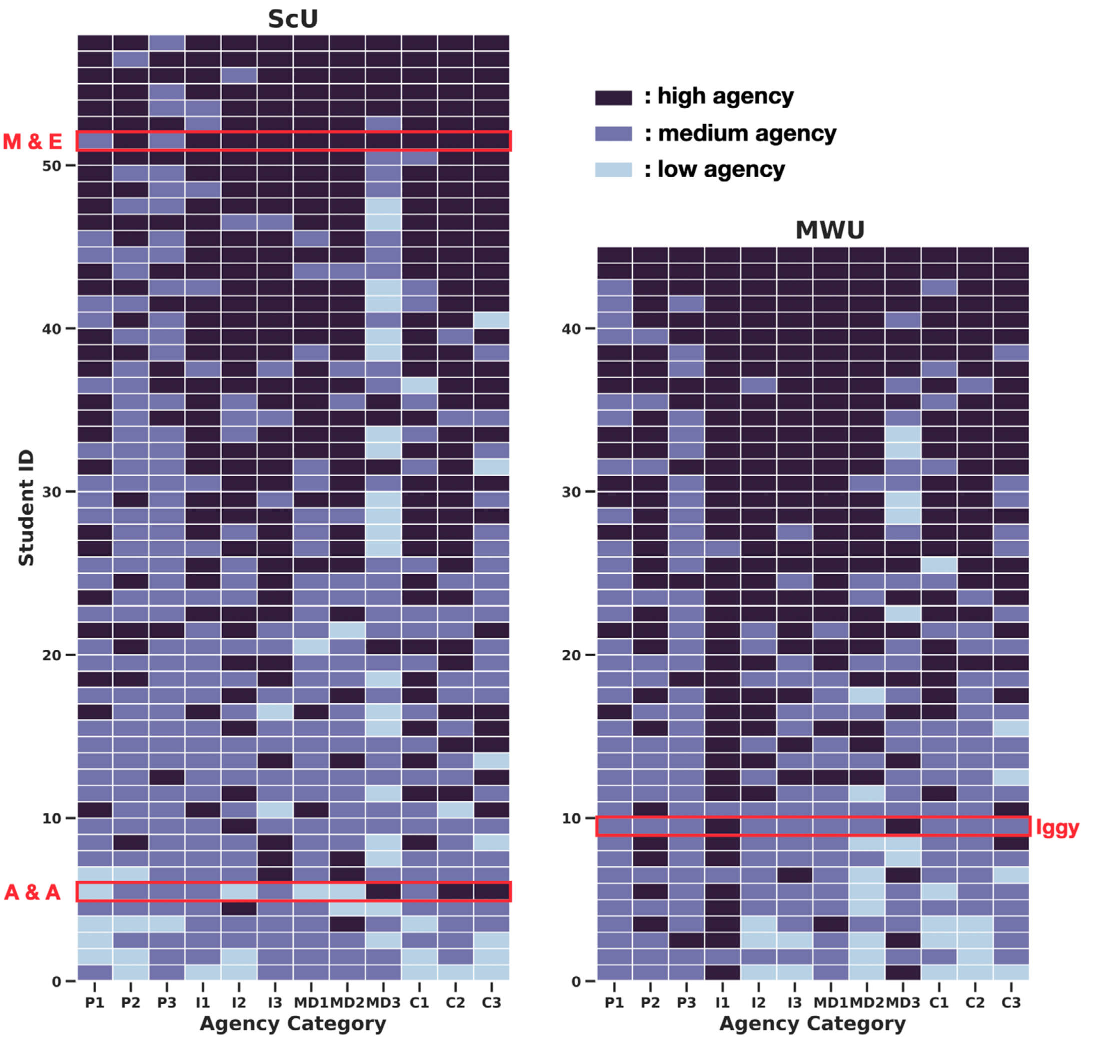


FIGURE 6 Heat map of disciplinary epistemic agency codes, disaggregated by institution and highlighting students featured in narratives. Darker colors indicate higher levels of demonstrated agency

topics were personally relevant to the students. This fact was often made clear by explicit references in the introduction (e.g., an essay on baseball statistics which began "I have always enjoyed playing and watching baseball. It is a data-driven sport that…") or the topic of the essay (e.g., "Could we use the concept of a rail gun to make a Space Elevator?"). Some essays also displayed clear markers of student agency throughout the narratives: for example, many essays from the Norwegian dataset included an analysis of the approximate cost of electricity for simulated tasks, like the aforementioned railgun-driven space elevator. Because Norway's economy is heavily energy-driven, this emphasis likely represents an example of students' home culture being integrated into the computational essay task.

We see a tentative connection between several of these trends and the scaffolding and expectations provided for the students. For example, students at MWU were explicitly required to define their own investigation question (including finding their own dataset if they were doing a data-driven project) and could not complete their project without doing so. For these students,

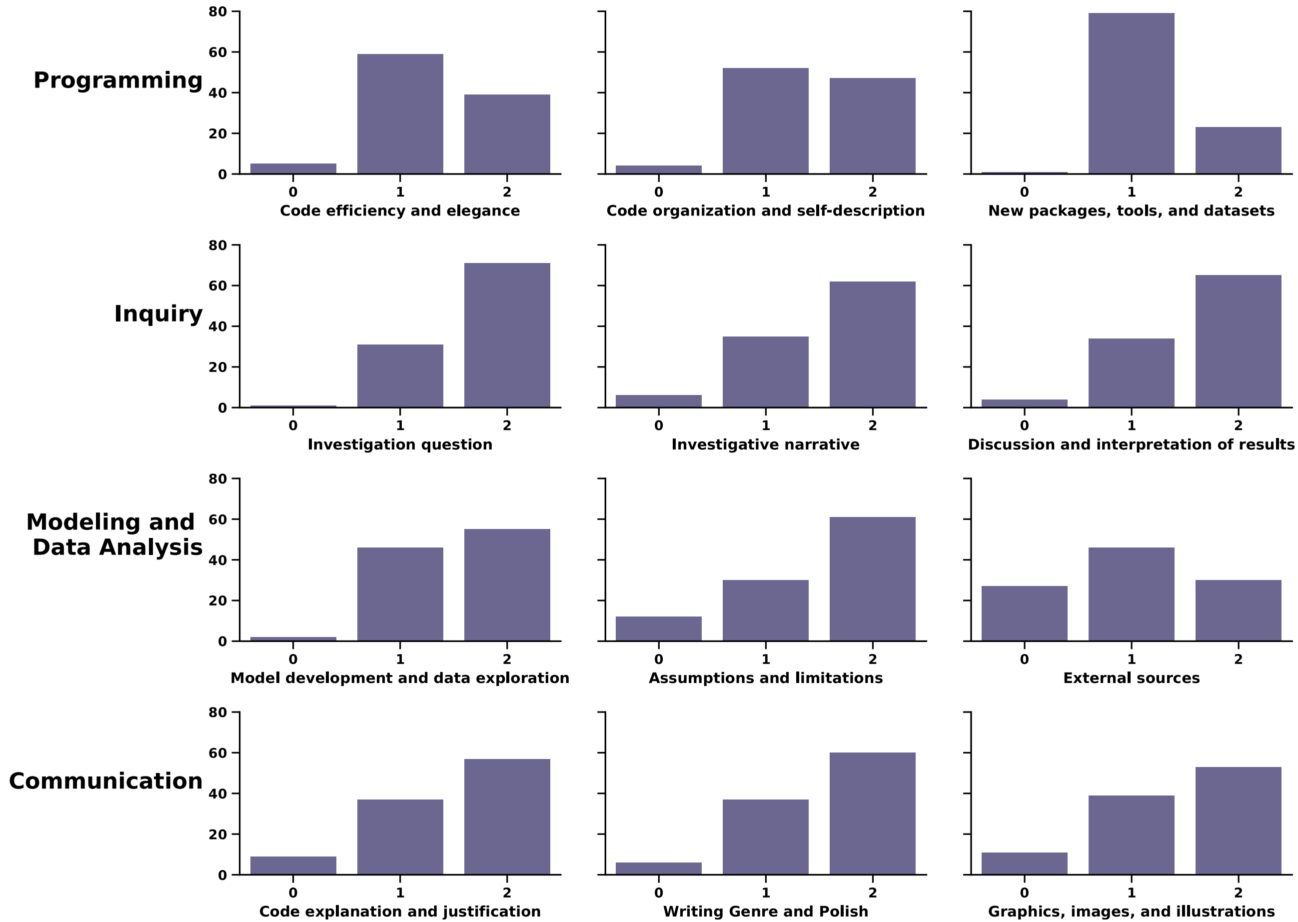


FIGURE 7 Number of students receiving specific agency scores across each of the 12 subcategories, aggregating data from the Scandinavian University and Midwestern University

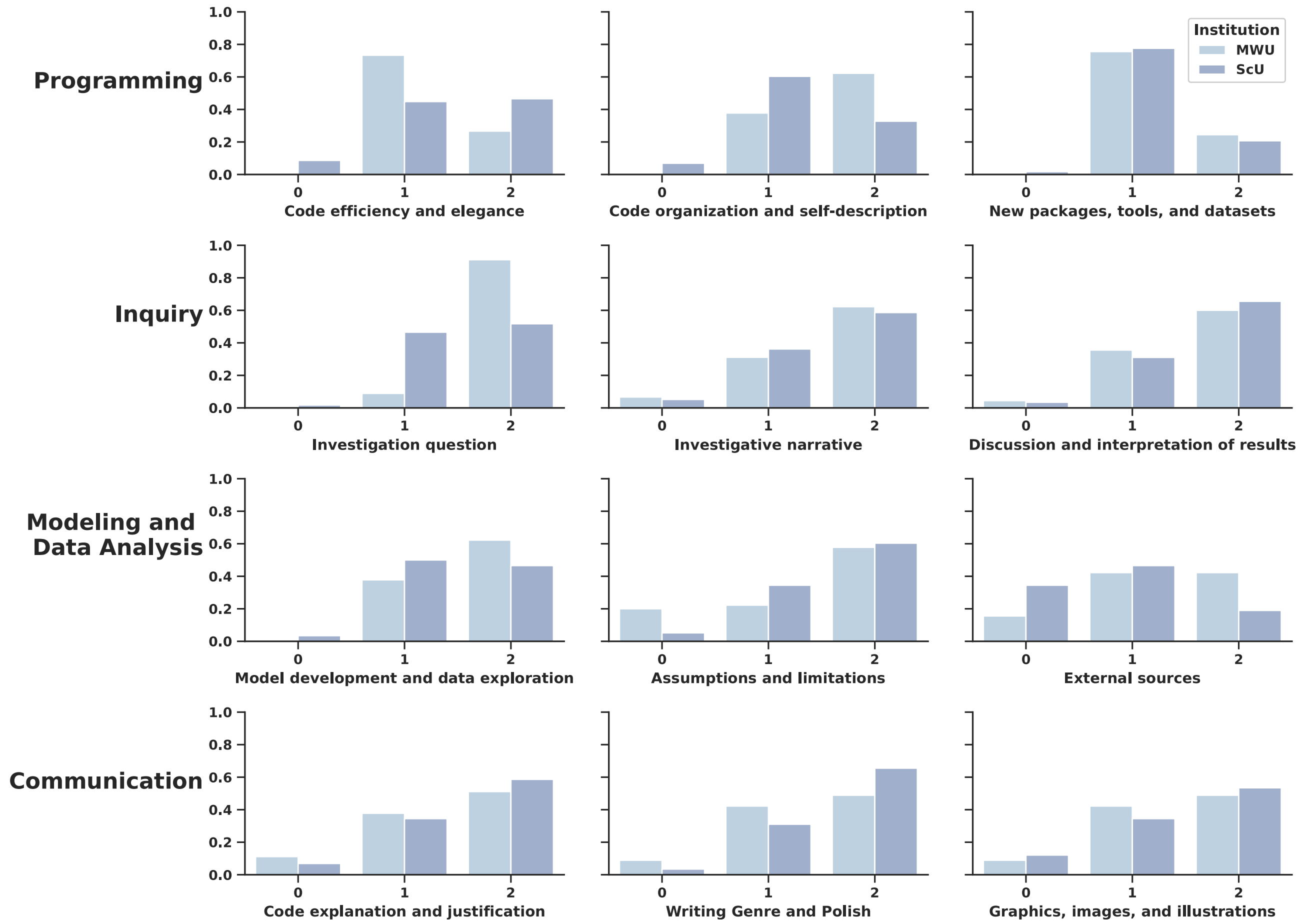


FIGURE 8 Percentages of students receiving specific agency scores across each of the 12 subcategories, separated by institutions (MWU vs. ScU)

the only reason that they would receive a "1" in the corresponding subcategory (I1) would be if they either failed to include their question in their essay or had a particularly ill-defined question. We explore these connections in greater detail in the next section.

Returning to our first research question, these results show that when writing computational essays, the ma students in this study population successfully took up agency within the areas of coding, inquiry, modeling, and communication. In other words, every subcategory showed high levels of agency in at least 20% of the dataset, as can be seen from Figures 2 and 3, indicating that every one of our theoretically defined categories of disciplinary epistemic agency was represented in the dataset. Furthermore, as can be seen from Figure 6, nearly every student in the dataset had at least one area in which they demonstrated high levels of agency, and many demonstrated high agency across multiple categories. However, there were clearly differences in uptake of disciplinary epistemic agency between these two institutions, differences that could be due to course context, expectations, scaffolding, or other factors. To unpack these differences, we must disaggregate the dataset by institution.

5.1.2 | Differences between the two institutions

By splitting the dataset by institution, we were able to make more direct comparisons between the degrees of agency demonstrated by the students across these two educational contexts. A normalized histogram of the codes, split by institution, is shown in Figure 8.

Looking at this comparison, we can make several observations. Several subcategories appear to be fairly isomorphic across the two institutions (e.g., P3, I2, I3, C1, and C3). We suggest that these similarities indicate points of alignment between the two educational designs: for example, both encouraged students to spend time presenting and justifying their analysis using both a mix of text and code (C1) and graphics, images, and illustrations (C3). Similarly, both designs encouraged students to present their results in a narrative form (I2), including discussion of relevant project results (I3). However, neither design explicitly required students to learn new programming tools for the project (P3) although some students did, in fact, do so.

Other categories show clear differences across the two institutions: these include P1, P2, I1, MD2, and MD3. All five of these differences were evaluated to be statistically significant according to Fisher's exact test, as shown in Table 2. We interpret these differences as reflecting differences between the two educational designs. For example, ScU students demonstrated higher overall levels of agency on *Code Efficiency and Elegance* (P1) and *Assumptions and Limitations* (MD2). We suspect that the higher degree of agency in code efficiency (P1) is likely due to the fact that most of the ScU students had taken an object-oriented programming course during their first semester, giving them a foundation in efficient computing prior to the course in question. In contrast, most of the MWU students were learning to code for the first time and consequently had likely not yet developed the same levels of programming skill. Similarly, the ScU students' emphasis on assumptions and limitations likely stemme from the fact that ScU students were evaluated on their justifications of the reasonability of their results, whereas the MWU were not evaluated on this point. Furthermore, all of the ScU essays were of a modeling type, which naturally lends itself to discussions of assumptions made within the model.

The three subcategories where MWU shows greater student agency (P2, I1, and MD3) also seem to result from differences in the two educational designs. As previously discussed, all MWU students are required to define their own investigation question, which likely explains the greater emphasis on the *Investigation question* subcategory (I1). ScU students, in contrast, were given the option to work off of a set of predefined questions, and many students chose this option

TABLE 2 Significance of differences in epistemic agency scores between MWU and ScU, evaluated using Fisher's exact test

Category	*p*-value (Fisher's exact test)
P1: Code efficiency and elegance	0.00442
P2: Code organization and self-description	0.00280
P3: New packages and coding tools	0.89420
I1: Investigation question	0.00002
I2: Investigative narrative	0.80332
I3: Discussion and interpretation of results	0.87839
MD1: Model development and data exploration	0.18509
MD2: Assumptions and limitations	0.04831
MD3: External sources	0.01576
C1: Code explanation and justification	0.60013
C2: Writing genre and polish	0.19365
C3: Graphics, images, and illustrations	0.71448

Note: Statistically significant p values are bolded ($p < 0.05$).

(resulting in a lower score in this category). We suspect the greater emphasis by MWU students on *Code organization and self-description* (P2) is due to the fact that MWU students work within the Jupyter Notebook framework from the first day of class and are presented with numerous assignments that model best-practices for notebook use and code documentation. ScU students, on the other hand, have less training in documenting code, and many ScU students reported this project as their first exposure to Jupyter notebooks. Finally, MWU students are explicitly required to include external sources in their essays (MD3) including citing any external dataset included in their work. At ScU, the use of external sources was encouraged but not required to pass.

These results help address our second research question: that is, they show that factors like differences in implementation and scaffolding can influence (or direct) student uptake of agency in certain categories. From one perspective, this is unsurprising: if one provides students with the option to build off of a suggested investigation question, there will naturally be fewer students who define their own, more novel questions. Similarly, if students have a strong background in code documentation or object-oriented programming, it seems reasonable to expect that this will be reflected in the choices they make (and the way they demonstrate agency) in their projects. And, criteria included in the respective grading rubrics clearly will affect student behaviors. However, these results imply that educators must be cognizant of these factors when implementing open-ended, project-based assignments like computational essays. If, for example, they wish to assess students on certain aspects of programming, inquiry, modeling, or communication, it is important to take into account student background, comfort, scaffolding, and grading criteria.

5.2 | Case studies of varying approaches to taking up disciplinary epistemic agency

We now present the narratives of three groups of students who took up agency in this project in different ways. These cases are meant to both provide both an illustration of key features of

TABLE 3 Summary of computational essay case study groups

Student pseudonym(s)	Student university and major(s)	Computational essay topic	Categories of demonstrated agency
Amy and Alexis	ScU; Electronics, Informatics, and Technology	Lightning safety in cars	External sources; writing genre and polish; graphics, images, and illustrations
Iggy	MWU; Computer Science	Baseball statistics	Investigation question; external sources
Margaret and Edward	ScU; Materials Science and Physics	Railgun dynamics	All categories *except* code efficiency and elegance, and new packages and coding tools

computational essays and showcase the varied approaches students took to this project. We summarize key features of these three cases in Table 3.

5.2.1 | Amy and Alexis focus on communication at the expense of computation

Amy and Alexis were engineering majors at the Scandinavian University, taking the physics course as part of their electronics, informatics and technology bachelor's degree. They based their project, which investigated lightning safety in cars, on one of the “seed” programs provided to ScU students. The program in question computationally simulated a lightning strike on flat ground, using certain electricity and magnetism concepts taught in the course, and one of the suggested investigation questions propose that students use the simulation to investigate whether a car was actually a safe place to shelter during a lightning strike.

In their interview, the pair reported that they had been quite curious about this topic, and so had spent extensive time reading and writing about it. Their essay clearly reflected this interest, featuring an extensive and detailed theoretical introduction including several eye-catching graphics illustrating different aspects of lightning safety in cars. However, when the pair reached the step of computationally modeling how lightning would strike a car, they struggled to find a way to implement a car-like structure within the simulation they were building on. Additionally, assignment guidelines stated that students could expect to spend approximately 10 hour on their project, and the pair realized that they had already reached that quota in the course of their background reading and writing. So, with the project deadline approaching, they left the provided code mostly untouched, pasted in a graphic from the textbook showing how they expected the simulation should look, and wrote a short summary and conclusion.

The pair described this decision as follows (translated from Norwegian):

> **Amy:** *We wanted to actually try to program it so that it would show how lightning would strike our car. How lightning would strike it and then go around and down instead of through, but we don't know if we had any, yeah...*
> **Alexis:** *It's only 10 hours you're supposed to use, we used more, but yeah...*

> **Amy:** *Yeah, and we had a lot of other things to do, so then it was like, okay, we can rather... And we found this picture in the textbook, so it was really this that we were trying to simulate, but I don't know. Maybe if we had had a little more time, or... then we could have tried to get a bit further.*

In the end, the pair fulfilled the necessary criteria to successfully pass the project. However, they also pursued a suggested investigation question instead of defining their own and made minimal modifications to their code, demonstrating low levels of agency (scoring a 0 or 1 out of 2) across several major categories (*Programming and Data Processing*, *Investigation*, and two-thirds of the subcategories of *Modeling and Data Analysis*). The students' description of their difficulties with (and ultimate abandonment of) the computational modeling parts of the project align with this characterization. At the same time, the students clearly spent a great deal of time and energy researching and writing about their chosen topic, and their scores in *Writing Genre and Polish*, *Graphics*, *Images and Illustrations*, and *External Sources* reflected this. This case, then, serves as an example of students who, due to constraints of time and background, used the computational essay project to reproduce and communicate existing textbook-based knowledge.

5.2.2 | Iggy defines a novel question but performs a surface-level analysis

Iggy was a second-year computer science student at the Midwestern University with an interest in data science. When the time came to choose a project, he initially struggled to settle on a topic, but eventually decided on an analysis relating to his interest in baseball:

> **Iggy:** *When we got to the project brainstorming phase, I'm not a very creative person, it was something that, it took a lot of days just sitting there and being, "No, no. That's not creative." But my biggest interest is baseball. My goal down the line for a career... [...] I'd like to do Sabermetrics, data analysis, looking at different stats, building algorithms, looking for patterns and such in a baseball context. Looking up how do we get the best players, or the best team based on a data-driven look than a scouting look. So I wanted to do something like that for my project.*

Based on this interest, he chose to analyze several of the most commonly used baseball statistics to see which was most correlated with a team's winning percentage. In his essay, after downloading and loading in several teams' datasets from baseball-reference.com, he calculated and plotted the correlation coefficient between different teams' win percentages and 10 different variables such as batting average, offensive strikeouts, and offensive walks. After summarizing these statistics, he concluded his essay with a summary of the various correlations and a brief discussion of the difference between correlation and causation.

Iggy's essay received exclusively 1s on all categories, with the exception of *Investigation Question* and *External Sources* (where it received 2s). Thus, this essay provided an example of an assignment in which the student had fulfilled all of the criteria, but only demonstrated agency in two areas: the question he chose to investigate, and the sources used to investigate it (including his chosen dataset). When interviewed, Iggy explicitly reflected on this point, describing how the most difficult parts of the project had been finding an appropriate question and dataset:

> **Interviewer:** *Okay. So you mentioned that the hardest part was choosing a topic. What made that so hard?*
> **Iggy:** *Kind of like I just said, aside from the fact that I'm really not a creative mind, being able to find data that we can use for our topic is one of the hardest things because we have an open internet with lots of different information and data on it, but to find exactly what we are looking for, something that we can modify to be what we are looking for can be a challenge, especially for what I was doing which was looking at inning by inning scores of baseball games.*

Once he had identified his question and dataset, Iggy deliberately chose a simple, surface-level analysis for the remainder of his project:

> **Interviewer:** *So what motivated you to go this route of just looking at, I think you said, ten different statistics, and looking at correlation and then making that the main focus of the project, rather than potentially other... How did you decide on that route?*
> **Iggy:** *To be honest, it was the simplest way to do it. I did not want to make anything harder than it had to be. So with this project, which was basically just trying to find which statistic is the strongest, there wasn't really any methods we learned in class I could think of, going through everything, that measures strictly correlation.*

Iggy's reference to not being a "creative mind", along with his explicit reference to looking for "the simplest way to do it" exemplify a student who chose to engage in a limited amount of knowledge production and communication, within self-defined boundaries.

5.2.3 | Margaret and Edward build a detailed model of railgun dynamics based on real-life examples

Margaret and Edward were second-year students at the Scandinavian University, majoring in materials science and physics respectively. For their project, the pair decided to build off of one of the provided simulations, which analyzed the motion of a projectile being launched out of a railgun. However, rather than using one of the suggested investigation questions, they instead chose to explore the effects of various phenomena they had learned about in a mechanics course they had taken the previous semester. As Edward explained during their interview

> **Edward:** *We did not really know how to formulate a good question in the start, but we knew we wanted to use the railgun as a model. Without being too unrealistic we wanted a realistic use of the railgun. We wanted to use the things we learned in mechanics of PhysMech [introductory mechanics] to model how it would move and forces would act upon the projectile. That's basically what we tried to do.*

To pursue this question, the pair began by researching actual railguns to try to find physically meaningful parameters for their simulation. They summarized their findings in the introduction of their essay, embedding a dramatic video clip of an actual railgun firing for illustration. They then built up a simple model of a railgun using well-documented and well-explained code, much of which was compartmentalized in functions. During their interview, they explicitly referenced this emphasis on good coding practices:

> **Margaret:** *We wanted coding that was efficient, even though sometimes I felt, at least, that we were not that good in programming and we could write a program that would be even more efficient, but our focus was on making the program as efficient as possible.*

Once they had created a simple working model, they refined it in several ways: implementing a function to allow them to find the appropriate angle necessary to launch the projectile a specified horizontal distance, implementing air resistance into their model, and exploring the effects of the Coriolis force on projectile trajectory. They illustrated their results with multiple plots, and explicitly discussed their efforts to provide useful illustrations for readers:

> **Edward:** *We wanted to present the visualization, the different distances it reached, because you could see the effect of the air resistance, which is quite big when the speed is such a high number, right?*

In their conclusion, they discussed the assumptions they made along the way, the limitations these assumptions put on the analysis, and provided an extensive list of works cited from a variety of sources.

> **Edward:** *Then we talked a little about results and then about the real life implications that might actually happen.* [...] *Of course there's a lot of problems because of the forces on the...* [rails] *yeah, and also the temperature as well, which develops in such a fast cannon, of course. Also a lot of upsides because you can shoot it as many times as you want. You don't need heavy ammunition or... there's no chemistry involved, no explosions, right? It would be a lot easier if they get it to work. And then sources, of course, at the end. They didn't really say they want sources, so we just put the links to the videos and we didn't really talk about them, but it's there.*

This case exemplifies the students who engaged deeply with the process of knowledge production and communication, taking up agency across all four major categories included in our analysis. The students defined a novel question, based on a real-life context, and built a sophisticated multi-step model to investigate it. Their code was well explained, structured, and documented. Their essay explicitly attended to readability and polish, including multiple types of visualizations (videos and plots) to illustrate phenomena and results. As a result, their essay received scores of 2 on all categories, save for code efficiency and elegance (since several parts of the model used copy-and-pasted code) and new packages and coding tools (since they only used tools that had been taught in the course).

These three narratives complement our large-scale analysis of the computational essay artifacts, illuminating several features that were difficult to access using that dataset. First, the narratives suggest that students were aware that there were several different aspects of the computational essay assignment that they could choose to focus on. For example, Margaret and Edward made clear distinctions between their efforts to define an investigation question, produce efficient code, discuss and situate their results, and produce clear visualizations. Second, they illustrate how different groups of students used this freedom to engage in varying degrees of knowledge production and communication. For example, Amy and Alexis focused primarily on reproducing and communicating knowledge they had found in a textbook, whereas the other two featured groups made varying efforts to discover and communicate new results. Finally, these results provide evidence that at least some of the differences in agency we had noted were intentional on the part of the students. For example, both Amy and Alexis and Iggy chose to downplay several aspects of their investigations due to a combination of time constraints and simple lack of interest. In contrast, Margaret and Edward seemed to consciously make a decision to spend time and effort on many of the categories featured in our analysis.

6 | DISCUSSION AND IMPLICATIONS

Returning to our first research question, we argue that computational essays can provide significant opportunities for students to take up disciplinary epistemic agency in higher education STEM. They allow students to define novel, personally relevant questions; answer them with scientifically authentic models or analyses; run these models or analyses using code of vary degrees of efficiency and sophistication; and communicate their results in an authentic way. Our artifact analysis showed that each of these aspects was well-represented across the two studied institutions, although different students in our study naturally gravitated toward different aspects of computational essays, as can be seen from Figure 6. Our interview analysis suggests that the students themselves were cognizant of some or all of these aspects and made specific choices as to which aspects to focus on.

Addressing our second research question, our analysis also shows that factors like student background, assignment expectations, time pressure, and student interests can have a significant impact on the ways in which students take up disciplinary epistemic agency when writing computational essays. In terms of student background, students at ScU had experience with object-oriented programming which likely led to the greater emphasis on code efficiency; at MWU, students had a stronger background in code documentation and notebook use, leading to a greater emphasis on code organization and description. In terms of assignment expectations, students at MWU were explicitly expected to define novel research questions and consult (cite) external sources leading to significantly higher levels of demonstrated agency across these categories, whereas students at ScU were asked to interpret results and justify reasonability leading to a greater emphasis on model assumptions and limitations. Additionally, our case study narratives show how student interest, combined with factors like time pressure, can influence students' choices to take up agency in different parts of the essay writing process. Theoretically, these differences demonstrate how disciplinary epistemic agency is a product of a variety of different factors: student motivation and interest, student background, the educational environment, and the specifics of the task the student is working on. All of these factors influence the scientific practices available to the students, their opportunities to engage in inquiry, and the standards by which they evaluate the knowledge being created.

This study responds and adds to the existing literature on epistemic agency in science education in several key ways. First, and foremost, our results demonstrate that computation can be a powerful tool for fostering disciplinary epistemic agency in higher education science. Computational essays in particular can amplify several of the affordances of computation while also bringing in opportunities for agency in scientific communication. Moreover, our results show that these kinds of interventions can be implemented at scale: each of the two courses presented enrolled over 200 students, all of whom participated in the respective computational essay-focused course projects. These results are, in themselves, novel; as noted above, there are gaps in the literature around epistemic agency in relation to both computation and higher education STEM; our study begins to address this gap.

One implication from these results is that science instructors looking to incorporate more opportunities for student agency into their teaching might consider integrating computation into their courses. For example, in physics, certain modeling-based curricula like the *Matter & Interactions* curriculum (Chabay & Sherwood, 2007), *Projects and Practices in Physics* (Irving et al., 2017), and *C2STEM* (Hutchins et al., 2020) have been shown to provide students with enough training in computation to be able to construct, analyze, and present simple computation models in the course of a single semester. For instructors who have already taken this step, we propose that they can use our operationalization of disciplinary epistemic agency to consider which areas of their course could be "opened up" to provide room for student agency. As discussed, this will necessarily be a function of student background and interest. However, because all analyzed aspects of the computational essay writing process—coding, inquiry, modeling and data analysis, and presentation—are authentic to the discipline of science, even if instructors choose to focus on one or two of these aspects, they will still be providing students with opportunities to productively engage with authentic scientific practices.

Second, more broadly, our results suggest that there is ample room in postsecondary science education for interventions and educational environments designed around disciplinary epistemic agency, outside of laboratory courses. This result, too, is novel, as much of the existing literature on epistemic agency in science education focuses on epistemic agency as a function of either K-12 educational standards, choices made by teachers when designing learning environments, or laboratory work. For example, Ko and Krist (2019) provide a framework for how K-12 teachers can create curricula aligned with the Next Generation Science Standards that open up room for epistemic agency; Stroupe (2014) provides several case studies of different K-12 teacher approaches that fostered varying degrees of epistemic agency; Stroupe et al. (2018) analyzed the interactions between a professional research team and a middle school teacher and class when trying to design these types of learning environments; and Miller et al. (2018) call out specific questions and contradictions when trying to implement the NGSS in a way that supports student agency. Because the context of our study is both international and within higher education, in this study, we are moving beyond an NGSS-focused conceptualization of epistemic agency.

Moreover, our results complement the existing literature on epistemic agency in higher education by extending the current focus beyond laboratory environments (e.g., Holmes, 2020; McFadden & Fuselier, 2020). When combined with that research base, our results suggest that there are multiple areas of higher education science that could be reconceptualized to focus on fostering student agency. Open-ended, project-based work is, in many ways, more authentic to professional science than standard textbook problems or highly structured, "cookbook"-style activities, whether they be experimental or computational (Holmes, 2020; Holmes & Wieman, 2016; Wieman, 2015). This authenticity is based on the fact that projects like these allow students to go beyond the reproduction of established scientific knowledge. Rather, they allow students to make analyses, observations, and discoveries that may well be completely new, using the authentic tools of the discipline. Clearly, both computational and laboratory spaces in higher education are well-suited to this kind of educational redesign.

Third, our results speak to the fact that assignment design can and should be a key aspect of supporting student agency. This, again, is not well-addressed in the epistemic agency literature, since most studies focus on teacher decisions, curriculum or lesson design, and implementation of standards. Yet, this study raises a point of tension that must be acknowledged when designing assignments that are meant to support student agency: even as they enable student agency in some ways, the structure and expectations of such assignments will always constrain student agency in others. For example, in both versions of the computational essay assignment students were required to produce a certain type of artifact (the computational essay), using certain types of tools (code and text), while demonstrating mastery over certain techniques (coding and data analysis/physics modeling).

We, however, do not see these constraints as being incompatible with disciplinary epistemic agency. Although these constraints provide a set of boundaries, well-designed assignments can still leave students with significant freedom in what kinds of new knowledge they choose to produce, how they produce it, and how they communicate it within those boundaries. Returning to our conceptualization of disciplinary epistemic agency, we argue that the critical question is whether students are being given the opportunity, choice, and cognitive authority to produce new knowledge and judge the quality of that knowledge, or whether they are compelled to engage in knowledge *reproduction*. Thus, although students writing computational essays are required to use certain computational methods and produce a specific style of artifact, within those boundaries they are given significant authority to decide what knowledge they would like to produce, judge the quality of their results, and determine how best to communicate their findings.

We note that a key fea of computational essays is their multi-modal nature; that is, they blend together various representations of scientific knowledge, including text, code, equations, images, graphs and plots, data tables, and more. We argue that this multi-modal nature is critical to fostering disciplinary epistemic agency because it gives students the potential to choose which areas they wish to focus on. When combined with a set of assignment expectations and assessment criteria that give students multiple viable paths to success, this multi-modality can be an important design principle for agency-focused assignments.

There are, of course, some clear limitations to this study. First, there is the fact that we have presented data from only two courses, from two institutions, from a single semester. A greater variety of data would certainly help us to strengthen and refine the claims we have made here, especially regarding the connections between educational context and disciplinary epistemic agency demonstrated by students. Second, many of our conclusions about student agency are based on our own judgments as practicing scientists and data analysts. They thus represent a top-down view of the phenomenon under study—a phenomenon that is, at its core, individual to learners. This top-down view is to some degree coherent with our theoretical framework, since one of the key aspects of disciplinary epistemic agency is authenticity to scientific disciplines. However, we must be cognizant that students come from a variety of backgrounds, and what may appear to us as, for example, a mediocre project may in fact have been very challenging (and required a great deal of agency) for the student who produce it.

Third, much of our analysis comes from the evaluation of student artifacts, which naturally has the limitation of only showing a finished product. We thus had limited insight into the students' thought processes as they pursued the majority of these projects. We tried to offset this limitation, to some degree, by interviewing subsets of students to ask about their thought processes. However, even these interviews are retrospective and do not reflect the rich messiness of the inquiry process. Furthermore, our dataset for this study includes artifacts and interviews from both individual students and pairs, with little distinction made between the two. We argue that the choice to combine these data was justified because our research questions and methods focus on the affordances of computational essay assignments for supporting student agency.

Accordingly, our analysis focused primarily on the computational essays themselves, with interviews serving to triangulate and unpack key features noted in the artifact analysis. However, we acknowledge that the experience of working with a partner on this kind of project certainly differs from that of working individually, and this undoubtedly affects the ways students take up disciplinary epistemic agency when writing computational essays. Our present analysis was not designed to capture these kinds of distinctions; however, future work on this project will explore the student experience of writing these types of essays, which will allow us to address this important element of student agency.

Finally, we must acknowledge the effects of self-selection bias in our dataset. We were only able to collect artifacts from a subset of students taking each course. It is possible, even likely, that the students who consented to let us use their work were those who were most secure or proud of the work they had done. Thus, we suspect our results might overestimate the spectrum of agency present in the courses as a whole.

Despite these limitations, we find these results encouraging. Broadly speaking, the results from this project suggest that computation can be an important tool in the ongoing effort to make science learning more authentic to the scientific disciplines and grant students more agency in their learning.

7 | CONCLUSION

It has long been known that the structures within schools set the expectations for student and teacher roles, which determine what kinds of knowledge are valuable, who holds that knowledge, and how it is constructed (Stroupe, 2014; Warren & Rosebery, 1995). In this regard, we argue that computational essays, when used for open-ended, inquiry-based, student-driven projects, may act as a structure to help students take up disciplinary epistemic agency in higher education science. Courses and institutions that wish to use this structure will almost certainly need to make adjustments to the current paradigm of teaching in higher education STEM, such as a decreased focus on students achieving one singular outcome or right answer. However, given the potential benefits of allowing students to experience authentic scientific inquiry, using authentic scientific methods, and learning an authentic mode of scientific communication, these seem like relatively minor tradeoffs—especially if computational essays are used as a single assignment within a larger course.

Although we have presented two examples of computational essay implementations, we are greatly interested in seei more. Computational essays have been discussed, at least theoretically, for the last two decades (diSessa, 2000; Somers, 2018; Wolfram, 2017). Now, however, we are finally seeing what it looks like when this genre begins to make its way into educational contexts. Because professional scientists frequently use computational essays for communication of research methods and results, we predict that their use in education will only increase in upcoming years. We see computational essays as a solid foundation for bringing in more principles of ambitious science teaching into the higher education sphere, and are excited about the possibility that computational essays may act as a site for this type of change.

ACKNOWLEDGMENTS

We would like to thank Sebastian Gregorius Winther-Larsen, Marcos (Danny) Caballero, David Stroupe, and Crina Damşa for their support and feedback in this project. This study was funded by Norges Forskningsråd Project No. 288125 "International Partnership for Computing in

Science Education" and by the Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education (DIKU) which supports the Center for Computing in Science Education.

ORCID

Tor Ole B. Odden https://orcid.org/0000-0003-1635-9491
Devin W. Silvia https://orcid.org/0000-0002-4109-9313
Anders Malthe-Sørenssen https://orcid.org/0000-0001-8138-3995

ENDNOTES

[1] A larger-scale analysis of the interviews is beyond the scope of this article, but is planned for a subsequent publication.

[2] We note, however, that this analysis only includes data from consenting students, which could also create a self-selection bias (i.e., students who failed the project may have been less likely to consent to data collection).

REFERENCES

American Association For The Advancement of Science. (2009). *Vision and change in undergraduate biology education: A call to action*. American Association For The Advancement of Science.

Apkarian, N., Henderson, C., Stains, M., Raker, J., Johnson, E., & Dancy, M. (2021). What really impacts the use of active learning in undergraduate STEM education? Results from a national survey of chemistry, mathematics, and physics instructors. *PLoS One*, *16*(2), e0247544. https://doi.org/10.1371/journal.pone.0247544

Association of American Universities. (2017). Progress toward achieving systemic change: A five-year status report on the AAU Undergraduate STEM Education Initiative.

Berland, L. K. (2011). Explaining variation in how classroom communities adapt the practice of scientific argumentation. *Journal of the Learning Sciences*, *20*(4), 625–664. https://doi.org/10.1080/10508406.2011.591718

Berland, L. K., & Crucet, K. (2016). Epistemological tradeoffs: Accounting for context when evaluating epistemological sophistication of student engagement in scientific practices. *Science Education*, *100*, 5–29. https://doi.org/10.1002/sce.21196

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, *95*(2), 191–216. https://doi.org/10.1002/sce.20420

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2015). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, *53*(7), 1082–1112. https://doi.org/10.1002/tea.21257

Blakeslee, A. M. (1997). Activity, context, interaction, and authority: Learning to write scientific papers in situ. *Journal of Business and Technical Communication*, *11*(2), 125–169.

Blum, R. (1971). *Computers in undergraduate science education: Conference proceedings*. Illinois Institute of Technology. Retrieved from. https://files.eric.ed.gov/fulltext/ED072648.pdf

Brownell, S. E., Kloser, M. J., Fukami, T., & Shavelson, R. (2012). Undergraduate biology lab courses: Comparing the impact of traditionally based "cookbook" and authentic research-based courses on student lab experiences. *Journal of College Science Teaching*, *41*(4), 36–45. Retrieved from. http://www.jstor.org/stable/43748539

Caballero, M. D., & Merner, L. (2018). Prevalence and nature of computational instruction in undergraduate physics programs across the United States. *Physical Review Physics Education Research*, *14*, 020129. https://doi.org/10.1103/PhysRevPhysEducRes.14.020129

Calabrese Barton, A., & Tan, E. (2010). We be burnin'! Agency, identity, and science learning. *Journal of the Learning Sciences*, *19*(2), 187–229. https://doi.org/10.1080/10508400903530044

Center for Computing in Science Education. (2019). *Computational essays from the University of Oslo*. Center for Computing in Science Education. Retrieved from. https://uio-ccse.github.io/computational-essay-showroom/index

Chabay, R., & Sherwood, B. (2007). Matter & Interactions. In E. F. Redish & P. J. Cooney (Eds.), *Reviews in PER volume 1: Research-based reform of university physics* (Vol. 1). American Association of Physics Teachers. Retrieved from. http://www.per-central.org/document/ServeFile.cfm?ID=4989

Chinn, C., & Malhotra, B. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218. https://doi.org/10.1002/sce.10001

Chonacky, N., & Winch, D. (2008). Integrating computation into the undergraduate curriculum: A vision and guidelines for future developments. *American Journal of Physics*, *76*, 327–333. https://doi.org/10.1119/1.2837811

Cooper, M. (1994). Cooperative chemistry laboratories. *Journal of Chemical Education*, *71*(4), 307. https://doi.org/10.1021/ed071p307

Cooper, M., & Klymkowsky, M. (2013). Chemistry, life, the universe, and everything: A new approach to general chemistry, and a model for curriculum reform. *Journal of Chemical Education*, *90*(9), 1116–1122. https://doi.org/10.1021/ed300456y

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). An expanded typology for classifying mixed methods research into designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Sage. Retrieved from. https://us.corwin.com/sites/default/files/upm-binaries/19291_Chapter_7.pdf

Damşa, C. I., Kirschner, P. A., Andriessen, J. E. B., Erkens, G., & Sins, P. H. M. (2010). Shared epistemic agency: An empirical study of an emergent construct. *Journal of the Learning Sciences*, *19*(2), 143–186. https://doi.org/10.1080/10508401003708381

Denning, P. J., & Tedre, M. (2019). *Computational thinking*. MIT Press.

diSessa, A. A. (2000). *Changing minds: Computers, learning, and literacy*. Mit Press.

Dounas-frazer, D. R., Bogart, K. L. V. D., Stetzer, M. R., & Lewandowski, H. J. (2016). Investigating the role of model-based reasoning while troubleshooting an electric circuit. *Physical Review Physics Education Research*, *12*, 010137. https://doi.org/10.1103/PhysRevPhysEducRes.12.010137

Dounas-Frazer, D. R., & Lewandowski, H. J. (2016). Nothing works the first time: An expert experimental physics epistemology. Physics Education Research Conference, June 4, 2016. https://doi.org/10.1103/PhysRevPhysEducRes.12.020129

Dounas-Frazer, D. R., & Lewandowski, H. J. (2018). The modelling framework for experimental physics: Description, development, and applications. *European Journal of Physics*, *39*(6), 064005. https://doi.org/10.1088/1361-6404/aae3ce

Ellis, M., & Lang, M.. (1965). Instruction by Design, A Report on the Conference on New Instructional Materials in Physics (University of Washington, Seattle, June 21–August 21, 1965). Retrieved from. http://files.eric.ed.gov/fulltext/ED043479.pdf.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. the problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 543–549. https://doi.org/10.1016/0895-4356(90)90158-l

Gutiérrez, K. D., & Calabrese Barton, A. (2015). The possibilities and limits of the structure–agency dialectic in advancing science for all. *Journal of Research in Science Teaching*, *52*(4), 574–583. https://doi.org/10.1002/tea.21229

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The Definitive Guide to Measuring The Extent of Agreement Among Raters* (4th ed.). Advanced Analytics, LLC.

Hardy, L., Dixon, C., & Hsi, S. (2020). From data collectors to data producers: Shifting Students' relationship to data. *Journal of the Learning Sciences*, *29*(1), 104–126. https://doi.org/10.1080/10508406.2019.1678164

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952–984. https://doi.org/10.1002/tea.20439

Henderson, C., & Dancy, M. (2009). Impact of physics education research on the teaching of introductory quantitative physics in the United States. *Physical Review Special Topics - Physics Education Research*, *5*(2), 020102. https://doi.org/10.1103/PhysRevSTPER.5.020107

Henderson, C., Dancy, M., & Niewiadomska-Bugaj, M. (2012). Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Physical Review Special Topics - Physics Education Research*, *8*(2), 020104. https://doi.org/10.1103/PhysRevSTPER.8.020104

Holmes, N. G. (2020). Why traditional labs fail… and what we can do about it. In J. J. Mintzes & E. M. Walter (Eds.), *Active learning in college science: The case for evidence-based practice* (pp. 271–290). Springer. https://doi.org/10.1007/978-3-030-33600-4

Holmes, N. G., Keep, B., & Wieman, C. E. (2020). Developing scientific decision making by structuring and supporting student agency. *Physical Review Physics Education Research*, *16*(1), 10109. https://doi.org/10.1103/PhysRevPhysEducRes.16.010109

Holmes, N. G., & Wieman, C. E. (2016). Examining and contrasting the cognitive activities engaged in undergraduate research experiences and lab courses. *Physical Review Physics Education Research*, *12*(2), 020103. https://doi.org/10.1103/PhysRevPhysEducRes.12.020103

Holmes, N. G., & Wieman, C. E. (2018). Introductory physics labs: We can do better. *Physics Today*, *71*(1), 38–45. https://doi.org/10.1063/pt.3.3816

Hora, M. T. (2015). Toward a descriptive science of teaching: How the TDOP illuminates the multidimensional nature of active learning in postsecondary classrooms. *Science Education*, *99*(5), 783–818. https://doi.org/10.1002/sce.21175

Hurtado, S., Eagan, K., Pryor, J. H., Whang, H., & Tran, S. (2012). *Undergraduate teaching faculty 2010: The 2010-2011 HERI faculty survey*. UCLA Higher Education Research Institute. Retrieved from. https://play.google.com/store/books/details?id=WUVaPQAACAAJ

Hutchins, N. M., Biswas, G., Maróti, M., Lédeczi, Á., Grover, S., Wolf, R., Blair, K. P., Chin, D., Conlin, L., Basu, S., & McElhaney, K. (2020). C2STEM: a system for synergistic learning of physics and computational thinking. *Journal of Science Education and Technology*, *29*(1), 83–100. https://doi.org/10.1007/s10956-019-09804-9

Irving, P. W., Mcpadden, D., & Caballero, M. D. (2020). Communities of practice as a curriculum design theory in an introductory physics class for engineers. *Physical Review Physics Education Research*, *16*(2), 20143. https://doi.org/10.1103/PhysRevPhysEducRes.16.020143

Irving, P. W., Obsniuk, M. J., & Caballero, M. D. (2017). P3: a practice focused learning environment. *European Journal of Physics*, *38*, 055701. https://doi.org/10.1088/1361-6404/aa7529

Kane, J. M. (2015). The structure-agency dialectic in contested science spaces: "Do earthworms eat apples?". *Journal of Research in Science Teaching*, *52*(4), 461–473. https://doi.org/10.1002/tea.21206

Kloser, M. J., Brownell, S. E., Chiariello, N. R., & Fukami, T. (2011). Integrating teaching and research in undergraduate biology laboratory education. *PLoS Biology*, *9*(11), e1001174. https://doi.org/10.1371/journal.pbio.1001174

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Jupyter Development Team (2016). Jupyter notebooks—A publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (Vol. 16, pp. 87–90). IOS Press. https://doi.org/10.3233/978-1-61499-649-1-87

Ko, M. L. M., & Krist, C. (2019). Opening up curricula to redistribute epistemic agency: A framework for supporting science teaching. *Science Education*, *103*(4), 979–1010. https://doi.org/10.1002/sce.21511

Lehrer, R., & Schauble, L. (2007). Scientific thinking and science literacy. In *Handbook of child psychology, Vol. IV. Child psychology in practice*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0405

Livingstone, D. N. (2010). *Putting science in its place: Geographies of scientific knowledge*. University of Chicago Press.

Lombardi, D., Shipley, T. F., & Astronomy Team, Biology Team, Chemistry Team, Engineering Team, Geography Team, Geoscience Team, & Physics Team. (2021). The curious construct of active learning. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, *22*(1), 8–43. https://doi.org/10.1177/1529100620973974

Magana, A. J., Falk, M. L., Vieira, C., & Reese, M. J. (2016). A case study of undergraduate engineering students' computational literacy and self-beliefs about computing in the context of authentic practices. *Computers in Human Behavior*, *61*, 427–442. https://doi.org/10.1016/j.chb.2016.03.025

Manduca, C. A., Iverson, E. R., Luxenberg, M., Macdonald, R. H., McConnell, D. A., Mogk, D. W., & Tewksbury, B. J. (2017). Improving undergraduate STEM education: The efficacy of discipline-based professional development. *Science Advances*, *3*(2), e1600193. https://doi.org/10.1126/sciadv.1600193

McFadden, J. R., & Fuselier, L. (2020). Graduate teaching assistants: Sharing epistemic agency with non-science majors in the biology laboratory. *Disciplinary and Interdisciplinary Science Education Research*, *2*(1), 1–19. https://doi.org/10.1186/s43031-020-00024-5

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*(7), 1053–1075. https://doi.org/10.1002/tea.21459

Moskovitz, C., & Kellogg, D. (2005). Primary science communication in the first-year writing course. *College Composition and Communication*, *57*(2), 307–334.

NGSS Lead States. (2013). *Next generation science standards: For states, By States*. The National Academies Press.

Norwegian Ministry of Education and Research. (2017). *Meld. St. 16 (2016–2017) Report to the Storting: Quality Culture in Higher Education*. Norwegian Ministry of Education and Research. Retrieved from. https://www.regjeringen.no/en/dokumenter/meld.-st.-16-20162017/id2536007/

Odden, T. O. B., & Burk, J. (2020). Computational essays in the physics classroom. *Physics Teacher*, *58*(4), 252–255. https://doi.org/10.1119/1.5145471

Odden, T. O. B., Lockwood, E., & Caballero, M. D. (2019). Physics computational literacy: An exploratory case study using computational essays. *Physical Review Physics Education Research*, *15*(2), 20152. https://doi.org/10.1103/PhysRevPhysEducRes.15.020152

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

Olson, S., & Riordan, D. G. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the President*. Executive Office of the President. Retrieved from. http://files.eric.ed.gov/fulltext/ED541511.pdf

Rudolph, J. L. (2005). Epistemology for the masses: The origins of "the scientific method" in American schools. *History of Education Quarterly*, *45*(3), 341–376. https://doi.org/10.1111/j.1748-5959.2005.tb00039.x/full

Rule, A., Tabard, A., & Hollan, J. (2018). Exploration and explanation in computational notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*. https://doi.org/10.1145/3173574.3173606

Sandi-Urena, S., Cooper, M., & Stevens, R. (2012). Effect of cooperative problem-based lab instruction on metacognition and problem-solving skills. *Journal of Chemical Education*, *89*(6), 700–706. https://doi.org/10.1021/ed1011844

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Open Court.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of Agency for Children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences*, *1*(1), 37–68.

Sikorski, T.-R., & Hammer, D. (2017). Looking for coherence in science curriculum. *Science Education*, *101*(6), 929–943. https://doi.org/10.1002/sce.21299

Silvia, D., O'shea, B., & Danielak, B. (2019). A learner-centered approach to teaching computational modeling, data analysis, and programming. In *Computational Science—ICCS 2019. ICCS 2019. Lecture Notes in Computer Science* (Vol. 11540, pp. 374–388). Springer. https://doi.org/10.1007/978-3-030-22750-0_30

Somers, J. (2018). The scientific paper is obsolete. Here's what's next. *Atlantic*, *4*, 1–19.

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Eagan, M. K., Jr., Esson, J. M., Knight, J. K., Laski, F. A., Levis-Fitzgerald, M., Lee, C. J., Lo, S. M., McDonnell, L. M., McKay, T. A., Michelotti, N., Musgrove, A., Palmer, M. S., Plank, K. M., … Young, A. M. (2018). Anatomy of STEM teaching in north American universities. *Science*, *359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892

Stroupe, D. (2014). Examining classroom science practice communities: How teachers and students negotiate epistemic agency and learn science-as-practice. *Science Education*, *98*(3), 487–516. https://doi.org/10.1002/sce.21112

Stroupe, D., Caballero, M. D., & White, P. (2018). Fostering students' epistemic agency through the co-configuration of moth research. *Science Education*, *102*(6), 1176–1200. https://doi.org/10.1002/sce.21469

Teasdale, R., Viskupic, K., Bartley, J. K., McConnell, D., Manduca, C., Bruckner, M., Farthing, D., & Iverson, E. (2017). A multidimensional assessment of reformed teaching practice in geoscience classrooms. *Geosphere*, *13*(2), 608–627. https://doi.org/10.1130/GES01479.1

Thompson, K. V., Chmielewski, J., Gaines, M. S., Hrycyna, C. A., & LaCourse, W. R. (2013). Competency-based reforms of the undergraduate biology curriculum: Integrating the physical and biological sciences. *CBE Life Sciences Education*, *12*(2), 162–169. https://doi.org/10.1187/cbe.12-09-0143

Varelas, M., Tucker-Raymond, E., & Richards, K. (2015). A structure-agency perspective on young children's engagement in school science: Carlos's performance and narrative. *Journal of Research in Science Teaching*, *52*(4), 516–529. https://doi.org/10.1002/tea.21211

Viskupic, K., Ryker, K., Teasdale, R., Manduca, C., Iverson, E., Farthing, D., Bruckner, M. Z., & McFadden, R. (2019). Classroom observations indicate the positive impacts of discipline-based professional development. *Journal for STEM Education Research*, *2*(2), 201–228. https://doi.org/10.1007/s41979-019-00015-w

Warren, B., & Rosebery, A. S. (1995). Equity in the future tense: Redefining relationships among teachers, students, and science in linguistic minority classrooms. In W. G. Secada, E. Fennema, & L. Byrd (Eds.), *New directions for equity in mathematics education* (pp. 298–328). Cambridge University Press.

Watkins, J., Coffey, J. E., Redish, E. F., & Cooke, T. J. (2012). Disciplinary authenticity: Enriching the reforms of introductory physics courses for life-science students. *Physical Review Special Topics - Physics Education Research*, *8*, 010112. https://doi.org/10.1103/PhysRevSTPER.8.010112

Wieman, C. (2015). Comparative cognitive task analyses of experimental science and instructional laboratory courses. *Physics Teacher*, *349*(53), 349–351. https://doi.org/10.1119/1.4928349

Wilson, J. M., & Redish, E. F. (1989). Using computers in teaching physics. *Physics Today*, *42*(1), 34–41. https://doi.org/10.1063/1.881202

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, *96*(5), 878–903. https://doi.org/10.1002/sce.21027

Wolfram, S. (2017). What is a computational essay? Retrieved from http://blog.stephenwolfram.com/2017/11/what-is-a-computational-essay/.

Young, N. T., Allen, G., Aiken, J. M., Henderson, R., & Caballero, M. D. (2019). Identifying features predictive of faculty integrating computation into physics courses. *Physical Review Physics Education Research*, *15*, 010114. https://doi.org/10.1103/PhysRevPhysEducRes.15.010114

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Odden, T. O. B., Silvia, D. W., & Malthe-Sørenssen, A. (2022). Using computational essays to foster disciplinary epistemic agency in undergraduate science. *Journal of Research in Science Teaching*, 1–41. https://doi.org/10.1002/tea.21821