Student challenges, strategies, and learning within the Data Mine Learning Community

Joseph A. Lyon
School of Engineering Education
Purdue University
West Lafayette, IN
lyonj@purdue.edu

Aparajita Jaiswal
Department of Computer and Information Technology
Purdue Univesity
West Lafayette, IN
jaiswal2@purdue.edu

Alejandra J. Magana
Department of Computer and Information Technology
Purdue University
West Lafayette, IN
admagana@purdue.edu

Ellen Gundlach
The Data Mine
Purdue University
West Lafayette, IN
gundlach@purdue.edu

Mark Daniel Ward
Statistics and The Data Mine
Purdue University
West Lafayette, IN
mdw@purdue.edu

***Abstract*— This full research paper looks at data science, which is an increasingly important topic of interest across all academic disciplines and industry sectors. As such, the need for curricula and learning designs for teaching data science has never been greater. One increasingly useful institutional initiative for building critical skills is living learning communities. This study reports on a data science learning community, called The Data Mine. As part of this learning community, students worked on similar tasks, took classes together, and participated in industry related assignments. This exploratory study aims to understand how students navigated learning challenges, what self-regulation strategies were adopted, and what students ultimately learned through participating in the Data Mine Learning Community. Students were grouped into achievement categories using a developed rubric on course assignments and their responses to open ended questions were evaluated for patterns among those in their achievement group. The results indicate that the highest performing students avoided coding errors and were highly focused on learning gains.**

Keywords—learning communities, higher education, data science

I. Introduction

Computational and data science are critical pieces to all levels of education in the 21st century across academic disciplines [1], [2]. The challenge for data science education is that it is inherently interdisciplinary, requiring skills from statistics, computer science, and domain knowledge to which it is applied [3]. One potential solution that allows for both disciplinary content and data science skills to be taught is the use of living learning communities.

Living learning communities are intentional educational interventions [4], [5] that allow students to live and learn alongside students of diverse interests and backgrounds in pursuit of some common goal or academic pursuit. Oftentimes, living learning communities integrate the classroom with activities that happen external to the classroom. In our study we investigated a learning community focused on the learning of data science skills called The Data Mine Learning Community (DMLC).

This exploratory study will use a mixed-methods approach to analyze student assignments and self-reported survey data from students who participated in the DMLC. The study focuses on answering the following research questions: (1) What types of learning challenges, self-regulation strategies, and learning outcomes do students report after completing an individualized data science assignment? (2) How does achievement on the individual assignment impact the types of learning challenges, self-regulation strategies, and learning outcomes that students reported?

II. Background

The background for this study consists of multiple bodies of literature. First is the literature around how data science is taught and learned within higher education contexts. Second is an understanding of the research and potential impact of using learning communities in education and its interest to the current study.

A. Data Science Education

One of the biggest issues facing data science education is that it has no obvious home within higher education [6]. Yet, it is needed in nearly every academic discipline, but it is often unclear to what degree and what is even meant by the term itself [7]. While there seems to be some overlap with traditional fields such as statistics, the term is much more interdisciplinary and has risen out of the need for professionals capable of working with the enormous amounts of data collected day to day [7].

Researchers and educators have tried many different strategies to incorporate data science into the curriculum, ranging from including it as a stand-alone introductory course at the undergraduate level [8] to developing data hackathons for data science learning [9]. While these interventions have often seen moderate to good success, curriculum space and extracurricular resources may not allow some educators the ability to use these developed interventions.

B. Living Learning Communities for Education

One opportunity for educators may be to look into living learning communities to help develop data science skills for students across disciplines. Living learning communities, or residential learning communities, are arrangements where students of similar interests live close together and often take similar coursework or programming [10]. Living learning communities are becoming more prominent within higher education settings [10].

This increase in prevalence is due to the numerous benefits that living learning communities can provide to students. Studies have shown increases in student engagement [11], increased sense of meaning and support [12], and ultimately both retention and academic performance [13]. Not only do these living learning communities have benefits on various student metrics, but also have been shown to broaden participation of underrepresented groups in fields such as STEM [14].

Living learning communities can address the needs of the data science education community in that they largely allow this learning to occur in parallel with the existing disciplinary curricula. Our study is primarily interested in the learning challenges students face, their interest in the subject matter, the self-regulation strategies that they use, and the learning outcomes student experience while participating in a living learning community focused on teaching data science.

C. Sociocultural Learning

Living learning communities benefit from the importance of sociocultural factors in the learning process. Sociocultural learning theories postulate that learning is heavily impacted by social interactions, sharing understanding and new perspectives with each other, and allowing diverse views of the world to impact the way an individual thinks [15]. These benefits emerge from allowing students to interact with each other during their learning process, an important characteristic of any healthy living learning community where individuals are frequently interacting with each other in common learning situations.

Additionally, Vygotsky theorized that there is a zone of proximal development, which is bounded by the potential of what a student can learn on their own and also by what the student can learn with the help of a knowledgeable other [15]–[17]. Living learning communities do just that, by putting students into communities of individuals that can help each other share understanding, build new perspectives, and perhaps learn more than they could on their own. One other aspect of the DMLC is that students are given access to mentors and experts which may allow them to learn even more by potentially expanding each students' zone of proximal development [16].

III. METHODS

This study uses a sequential mixed-methods design to investigate student experiences within a living learning community [18].

A. The Data Mine Learning Community [DMLC]

The Data Mine Learning Community (DMLC) is a living learning community situated in a large Midwestern University. The DMLC allows the students of various disciplines with or without any prior knowledge of data science to join the learning community and develop data science skills [19]. The focus of DMLC is to create a data-driven workforce for the future. The vast amount of data being generated every moment has created an urge to develop a data driven workforce. The students in the DMLC learn data science skills by engaging in collaborative real-world projects. They work under the guidance of experienced mentors, staff, and faculty. The focus of DMLC is to help students develop an identity and foster socialization within the learning community.

B. Data Collection

Participants for this study were first-year undergraduate students from multiple academic disciplines, who were part of the DMLC. The 15 students for this study were selected on the basis of a voluntary survey conducted in December of 2020. Demographic information such as race/ethnicity or gender were not collected as part of the study.

As part of their first-year experience, students were asked to turn in one assignment of their choice for assessment in the class. In addition to turning in this assignment, students were asked to answer multiple survey questions along with the assignment for instructors to understand why it was selected and what the student learned as part of the process. The survey questions collected are listed below:

1. Why did you select this assignment?
2. What data science practices/skills did you think you learned best by completing this assignment?
3. What was the biggest challenge you encountered in solving this assignment?
4. What are the strategies that you adopted or would adopt to plan, monitor, and evaluate your assignment/project?

The student assignments were the basis for the quantitative analysis, while the free response survey questions were used as the basis for the qualitative analysis to understand the challenges, self-regulation strategies, and learning outcomes of the students.

C. Quantitative Analysis

Students turned in one assignment at the end of the semester for them to be assessed by the instructor. A rubric was developed that was able to assess student performance across a multitude of different assignments. Appendix I shows the rubric used by the researchers to evaluate student performance on the data science assignment. The rubric evaluated two criteria: (1) the coding style of the student on the programming portion of the assignment, and (2) deployment of disciplinary concepts used in the assignment. Together all student assignments were evaluated on these criteria by two researchers independently with a correlation between the two raters of 0.89.

The students (n=15) were then split into two groups, high performers (n==6, greater than the median) and moderate performers (n=9, equal or less than the median). These two groups then formed the basis for the comparison to understand the different challenges, strategies, and learning outcomes based on student achievement on the assignment.

D. Qualitative Analysis

Survey questions were then qualitatively analyzed for different codes throughout the participant sample. Each question was analyzed independently of the other questions. A thematic analysis process was used to identify the initial codes and converting them into eventual categories [20]. Once two researchers had finished coding the survey responses, they met together to discuss any differences between their codes and came to consensus on all student responses.

IV. RESULTS

The results indicate that students generally performed at a moderate to good level on their submitted assignments. The overall descriptive statistics for student scores based on the rubric (Appendix I) are shown in Table I. Additionally, Table I shows the statistics for the high performers (those above the median score) and for the moderate performers (those equal to or below the median score).

TABLE I. OVERALL DESCRIPTIVE STATISTICS

Grouping	Descriptive Statistics	
Overall Sample (n=15)	Mean	4.26
	Median	4.00
	St Dev	1.10
High Performers (n=6)	Mean	5.33
	St Dev	0.52
Moderate Performer (n=9)	Mean	3.56
	St Dev	0.73

The second group is considered moderate in that their average is still nearly at the midpoint for the rubric (4), while the high group is closer to the upper limit for the rubric (6). The following sections describe the specific categories that resulted from the qualitative analysis of the open-response questions of the survey. Four categories were identified in the topics of motivation, learning, challenges, and strategies used to overcome those challenges. Each of them include a list of identified codes, along with an overall description of the categories.

A. Motivation

Students discussed a variety of reasons why they chose the selected problem to be used for their course assessment.

TABLE II. CATEGORIES OF MOTIVATION

Category Name	Freq.(%)		Definition
Learned a lot/Specific Learning	T:	60	Reported learning a lot through the assignment, or mentioned specific topical learning because of the assignment.
	H:	83	
	M:	44	
Interesting Assignment/Fun	T:	40	Reported that the assignment was of interest or was fun.
	H:	17	
	M:	56	
Allowed Creativity	T:	20	Reported that the assignment allowed creativity or to create new code/solutions.
	H:	33	
	M:	11	
Brought New Perspective/Way of Thinking	T:	7	Reported that the assignment allowed them to have a new perspective about the world/discipline or allowed them to change the way they thought.
	H:	17	
	M:	0	
Real-world Data	T:	7	Reported that the assignment allowed them to use real-world data/practices.
	H:	17	
	M:	0	
Uninterested	T:	7	Lack of interest/no motivation for submitting assignment.
	H:	0	
	M:	11	

Many of these codes illuminate the characteristics of assignments that motivated students. Table II above lists the categories of motivation found in the responses to the survey question: (1) Why did you select this assignment?

The codes pertaining students' motivation ranged from the assignment allowing them to see the world differently, to assignments allowing creativity. These codes suggest a prescriptive framework for how instructors can instill motivation into their students through data science assignments, namely: (1) give real-world data, (2) make the problem sufficiently open-ended to allow for creativity, (3) embed the problem into disciplinary context for specific learning, and (4) try to create problems that create enough transfer to allow students to think about problems differently.

High performing students were more likely to focus on the learning gains as a motivator whereas the moderate group had a higher focus on how fun or interesting the topic was. The highest performers also mentioned unique motivators such as the assignment using real-world data was well as allowing them to bring a new perspective to their understanding.

B. Learning

In addition to the motivations for why students chose the assignment they did, students reported a wide variety of topics and concepts that were learned through this assignment while participating as part of the living learning community. Table III lists the categories found in response to the survey question: (2) What data science practices/skills did you think you learned best by completing this assignment?

One interesting finding from the codes of learning is the emergence of computational thinking concepts such as pattern recognition/generalization, algorithmic thinking, and decomposing large datasets [21]. However, there were other data science areas learned such as specific programming syntax and languages as well as how to work with real-world problem solving.

The moderate learner groups seemed to focus more on writing functions and learning the programming language, while the higher performing learners focused more on understanding and analyzing patterns within the data. Although the differences aren't large, they indicate that further research may be able to illuminate differences in student focus during the data science interventions.

TABLE III. CATEGORIES OF LEARNING

Category Name	Freq. (%)		Definition
Writing Functions	T: H: M:	40 17 56	Reported writing their own functions as a learning.
Using Specific Functions	T: H: M:	27 33 22	Reported that they learned about and how to use specific functions.
Analyze/Understand/Patterns in Datasets	T: H: M:	20 33 11	Reported that the assignment allowed them to analyze or see patterns in datasets.
Large Datasets	T: H: M:	13 17 11	Reported that they learned how to work with large datasets.
Data Visualization	T: H: M:	13 17 11	Reported that they learned about data visualization/plotting.
Loops/Logic/Indexing/Syntax	T: H: M:	13 17 11	Reported that the assignment taught them about specific syntax and the logic of coding.
Programming Languages	T: H: M:	7 0 11	Reported they the assignment allowed them to learn a new programming language.
Real-world Problems	T: H: M:	7 0 11	Reported that the assignment allowed them to learn about real-world problems.
Independent Analysis	T: H: M:	7 0 11	Reported that the assignment allowed them to learn because they worked independently.
Writing Functions	T: H: M:	7 0 11	Reported writing their own functions as a learning.

C. *Challenges*

Students reported multiple challenges from the data science assignment of their choosing, with a large range from programming related issues to conceptual difficulty with the material. Table IV lists the different codes regarding challenges identified in response to the survey question: (3) What was the biggest challenge you encountered in solving this assignment?

While some of these challenges were directly related to the learning goals for the assignment (concepts given or understanding the logic needed to solve the problem), others were challenges from peripheral requirements of the assignment (formatting). One challenge encountered was a lack of examples given by the instructional team.

The high performing students seemed to struggle more with understanding the syntax and logic behind the programs they were creating while the moderate performing students seemed to struggle with coding errors. This difference may highlight either prior programming experience or self-efficacy gaps between the two groups [22], [23].

D. *Strategies*

Finally, there were categories of strategy that students used to overcome challenges as part of their data science assignment. Table V overviews the different codes related to strategies that students reported in response to the final survey question: (4) What are the strategies that you adopted or would adopt to plan, monitor, and evaluate your assignment/project?

While some students used metacognitive strategies such as reflection, allowing failure, and breaking down the problem, some students discussed extrinsic strategies such as looking for help from other examples, the instructor, or previous course materials. One external resource notably missing from this list of strategies would be other students within the learning community.

Most notably, it was the moderate performers who were much more likely to look to the TA or teacher for help when they were experiencing challenges, whereas the biggest strategy for many of the highest performers was focused on time management during the problem solving process.

TABLE IV. CATEGORIES OF CHALLENGE

Category Name	Freq. (%)		Definition
Creating Code/Coding Errors	T: H: M:	40 17 56	Challenges were in writing code or errors from the code.
Understanding Syntax/Logic	T: H: M:	33 50 22	Challenges were in understanding the specific syntax or logic of programming the assignment.
Understanding the Problem	T: H: M:	13 17 11	Challenges were in understanding the actual problem or concepts that were given.
Formatting Errors	T: H: M:	13 17 11	Challenges were in how to format the assignment.
Lack of Examples	T: H: M:	7 0 11	Challenges were in there being a lack of examples of how to do the assignment.

V. DISCUSSION

The discussion is divided into two discrete sections. In the first section, a discussion of the overall results as they relate to research question one is presented. From there, the differences that may be indicated from the results as it relates to research question two.

A. *Overall results*

When students discussed their motivation for the assignment they chose to submit for grading, it is not surprising that how much they learned was the most frequent response. It is interesting that many students mentioned how interesting or fun the work was in choosing which assignment to submit for grading. This result does support the growing trend in the literature for the need for interesting and engaging learning approaches, some of which are currently growing trends in the educational literature as game-based learning [24], [25] and inquiry-based learning [26], [27]. Continued research into bringing these pedagogical practices into data science education is encouraged by our results.

TABLE V. CATEGORIES OF STRATEGY

Category Name	Freq. (%)	Definition
Reviewing Materials	T: 40 H: 33 M: 44	Discusses reviewing the problem, materials, or notes as a strategy.
Other Examples/Practice	T: 33 H: 33 M: 33	Discusses using other examples or practicing to overcome challenges they are experiencing.
TA/Teacher	T: 27 H: 0 M: 44	Discusses reaching out to the teaching assistant or teacher as a strategy.
Time Management	T: 27 H: 50 M: 11	Discusses time management/break or planning ahead as a strategy.
Testing Code	T: 13 H: 0 M: 22	Discusses testing/validating the code as a strategy.
Reflection	T: 7 H: 17 M: 0	Discusses using reflection as a practice to overcome challenges.
Breaking down the problem	T: 7 H: 0 M: 11	Discusses breaking down the problem to overcome challenges.
Allowing Failure	T: 7 H: 17 M: 0	Discusses allowing failure or to reach the point of failure before looking for additional help.

Additionally, the biggest challenges that students encountered during the submitted assignments was coding errors as well as understanding the logic of the code. This is to be expected in that the educational system is often lacking in programming and computational thinking curriculum [1], [28]. It is encouraging from our results that the biggest challenges were not the data practices themselves, but often coding errors and issues with the programming. One pivot that the DMLC program could make to lessen these challenges would be to partner with the computer science department to help develop programming skills, or to put the problems into disciplinary contexts which has been shown to help improve programming skills while teaching focused content [29], [30].

Finally, in terms of strategy of overcoming challenges, one glaring omission from the results is that students did not report reaching out to others in their learning community. One encouraging result is that higher order strategies emerged from a small percentage of students, such as reflecting on their struggles which is a basis for metacognition[31], and allowing themselves to get to a failure point for their own learning which has been shown to increase learning gains [32], [33]. These results may indicate that incorporating these topics into data science interventions may prove very effective, although a larger sample size and further research is needed to evaluate the efficacy of doing so.

B. Differences by achievement level

There were a few notable differences between the high and moderate achievers in terms of learnings, strategies, challenges, and motivations. While our exploratory study has too small of a sample size to do meaningful statistical differentiation, results here may indicate areas for further study in the literature.

In terms of the motivation for turning in the assignment that they did, the highest achievers seemed to focus on the learning gains, whereas the moderate achievers seemed to focus more on how interesting the assignment was. This difference is likely the reason that the students ended up being either high or moderate achievers in the first place, as those who felt like they learned the most were also more likely to perform better on the assignment itself.

As for the challenges, the moderate performers were more likely to mention coding errors and struggles with writing the actual code. While more research on the previous experiences would be needed to completely understand this difference, the difference may lie in which students felt comfortable with the assignment compared to others, which can have a significant effect on actual performance [34], [35]. Future research should monitor student previous experiences and self-efficacy of students.

One striking difference between the two groups is that the high performers mentioned time management strategies more frequently than the moderate performers. In contrast, the moderate performers mentioned asking the teaching assistant or teacher questions as a strategy for overcoming challenges. This could partly be because the higher performers were well integrated into the learning community, or because they felt that they already had the tools necessary to complete the assignment [35]. In discussing time management, as well as things like reflection and allowing failure, the higher performers seem to embody some practices of indicative of expertise [36], [37].

VI. CONCLUSIONS AND LIMITATIONS

While exploratory in nature, there are a few conclusions to draw from the current study. Our study investigate the reported challenges, motivations, learnings, and strategies that students used on a data science assignment while involved in the Data Mine Learning Community. The results indicate that high performers felt that they learned a lot, that understanding the underlying logic was challenging, and that time management was a successful strategy for learning. The other group, the moderate performers, felt that their assignments were interesting, but struggled with coding errors and felt the need to reach out to teaching assistants and teachers when struggling.

Our study has limitations in the small sample size and that the data is primarily self-reported data from students themselves. Additionally, our study is limited by the lack of information on student demographics or prevous expereinces. Our plan is that our future research will be able to take some of these exploratory findings to develop future research questions, such as: (1) what is the role of self-efficacy of students in The Data Mine Learning Community?, (2) what impact does The Data Mine Learning Community have on building expert practices?, (3) what pedagogical practices futher build a sense of student belongingness and community within The Data Mine Learning Community?, and (4) what pedagogical practices may limit the computing programming barriers to success in TThe Data Mine Learning community?

ACKNOWLEDGMENT

The research reported in this paper was supported in part by the National Science Foundation under the awards DMS- 1246818 and EEC-1449238 and by the Lilly Endowment Charting the Future Phase I Planning Grant, through the Purdue Office of the Provost. This work was also supported in part by the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE-1842166). M.D. Ward's research is also supported by National Science Foundation (NSF) grants CCF-0939370, and OAC-2005632, by the Foundation for Food and Agriculture Research (FFAR) grant 534662, by the National Institute of Food and Agriculture (NIFA) grants 2019-67032-29077 and 2020- 70003-32299, by the Society of Actuaries grant 19111857, by Cummins Inc. grant 20067847, by Sandia National Laboratories grant 2207382, and by Gro Master. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] President's Information Technology Advisory Committee, "Computational science: Ensuring America's competitiveness." Arlington, VA, 2005.

[2] [NRC], Next Generation Science Standards: For States, By States. Washington, D.C.: The National Academies Press, 2013.

[3] M. Koby and H. Orit, "Ten Challenges of Data Science Education," Commun. ACM, 2020.

[4] O. T. Lenning, D. M. Hill, K. P. Saunders, A. Solan, and A. Stokes, Powferful learning communities: A guide to developing student, faculty, and professional learning communities to improve student success and organizational effectiveness. Stylus Publishing, 2013.

[5] G. R. Pike, "The effects of residential learning communities and traditional residential living arrangements on educational gains during the first year of college.," J. Coll. Stud. Dev., vol. 40, no. 3, p. 269, 1999.

[6] W. Finzer, "The Data Science Education Dilemma," Technol. Innov. Stat. Educ., vol. 7, no. 2, 2013.

[7] R. A. Irizarry, "The Role of Academia in Data Science Education," Harvard Data Sci. Rev., pp. 1–8, 2020.

[8] J. Saltz and R. Heckman, "Big Data Science Education: A Case Study of a Project-Focused Introductory Course.," Themes Sci. Technol. Educ., vol. 8, no. 2, pp. 85–94, 2015.

[9] C. Anslow, J. Brosz, F. Maurer, and M. Boyes, "Datathons: An experience report of data hackathons for data science education," SIGCSE 2016 - Proc. 47th ACM Tech. Symp. Comput. Sci. Educ., pp. 615–620, 2016.

[10] M. S. Dunn and L. A. Dean, "Together We Can Live and Learn," Sch. A J. Leis. Stud. Recreat. Educ., vol. 28, no. 1, pp. 11–23, 2013.

[11] C.-M. Zhao and G. D. Kuh, "Learning communities and student engagement," Res. High. Educ., vol. 45, no. 2, pp. 115–138, 2004.

[12] L. B. Spanierman et al., "Living learning communities and students' sense of community and belonging," J. Stud. Aff. Res. Pract., vol. 50, no. 3, pp. 308–325, 2013.

[13] J. Hotchkiss, R. E. Moore, and M. M. Pitts, "Freshman learning communities, college performance, and retention," Educ. Econ., vol. 14, no. 2, pp. 197–210, 2006.

[14] J. Maltby, C. Brooks, M. Horton, and H. Morgan, "Long Term Benefits for Women in a Science, Technology, Engineering, and Mathematics Living-Learning Community.," Learn. Communities Res. Pract., vol. 4, no. 1, p. 2, 2016.

[15] S. J. Hausfather, "Vygotsky and Schooling: Creating a Social Context for Learning," Action Teach. Educ., vol. 18, no. 2, pp. 1–10, 1996.

[16] L. S. Vygotsky, Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press, 1978.

[17] J. Bruner, "Celebrating divergence: Piaget and Vygotsky," Hum. Dev., vol. 40, no. 2, pp. 63–73, 1997.

[18] N. V. Ivankova, J. W. Creswell, and S. L. Stick, "Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice," Field methods, vol. 18, no. 1, pp. 3–20, 2006.

[19] A. J. Magana, A. Jaiswal, A. Madamanchi, L. C. Parker, E. Gundlach, and M. D. Ward, "Characterizing the psychosocial effects of participating in a year-long residential research-oriented learning community," Curr. Psychol., 2021.

[20] V. Braun and V. Clarke, "Using thematic analysis in psychology," Qual. Res. Psychol., vol. 3, no. 2, pp. 77–101, 2006.

[21] C. C. Selby and J. Woollard, "Computational Thinking : The Developing Definition," Proc. 2014 Conf. Spec. Interes. Gr. Comput. Sci. Educ., pp. 5–8, 2013.

[22] M. A. Hutchison, D. K. Follman, M. Sumpter, and G. M. Bodner, "Factors influencing the self-efficacy beliefs of first-year engineering students.," J. Eng. Educ., vol. 95, no. 1, pp. 39–47, 2006.

[23] N. Lytle et al., "Use, modify, create: Comparing computational thinking lesson progressions for STEM classes," Annu. Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE, pp. 395–401, 2019.

[24] A. Steinmaurer, J. Pirker, and C. Gütl, "Scool - Game-based learning in computer science class A case study in secondary education," Int. J. Eng. Pedagog., vol. 9, no. 2, pp. 35–50, 2019.

[25] Q. T. Le, A. Pedro, and C. S. Park, "A Social Virtual Reality Based Construction Safety Education System for Experiential Learning," J. Intell. Robot. Syst. Theory Appl., vol. 79, no. 3–4, pp. 487–506, 2015.

[26] M. A. Al Mamun, G. Lawrie, and T. Wright, "Instructional design of scaffolded online learning modules for self-directed and inquiry-based learning environments," Comput. Educ., vol. 144, no. September 2019, p. 103695, 2020.

[27] D. Ketcheson, "Teaching numerical methods with IPython notebooks and inquiry-based learning," Proc. 13th Python Sci. Conf., pp. 19–24, 2014.

[28] National Research Council, "Report of a Workshop on the Pedagogical Aspects of Computational Thinking," National Academies Press, Washington, D.C., 2011.

[29] A. J. Magana, M. L. Falk, and M. J. Reese Jr., "Introducing Discipline-Based Computing in Undergraduate Engineering Education," ACM Trans. Comput. Educ., vol. 13, no. 4, pp. 16:1-16:22, 2013.

[30] H. W. Fennell, J. A. Lyon, A. Madamanchi, and A. J. Magana, "Toward computational apprenticeship: Bringing a constructivist agenda to computational pedagogy," J. Eng. Educ., no. February, pp. 1–7, 2020.

[31] J. H. (Stanford U. Flavell, "Metacognition and Cognitive Monitoring A New Area of Cognitive — Developmental Inquiry," Am. Psychol., vol. 34, no. 10, pp. 906–911, 1979.

[32] M. Kapur and K. Bielaczyc, "Designing for Productive Failure," J. Learn. Sci., vol. 21, no. 1, pp. 45–83, 2012.

[33] M. Kapur, "Productive failure," Cogn. Instr., vol. 26, no. 3, pp. 379–424, 2008.

[34] T. Busch, "Gender Differences in Self-Efficacy and Attitudes toward Computers," J. Educ. Comput. Res., vol. 12, no. 2, pp. 147–158, 1995.

[35] G. M. Hutchison, M. A., Follman, Deborah K, Sumpter, Melissa, Bodner, "Factors Influencing the Self-Efficacy Beliefs of First-Year Engineering Students.," J. Eng. Educ., vol. 95, no. 1, pp. 39–47, 2006.

[36] N. Steenhof, N. N. Woods, and M. Mylopoulos, "Exploring why we learn from productive failure: insights from the cognitive and learning sciences," Adv. Heal. Sci. Educ., vol. 25, no. 5, pp. 1099–1106, 2020.

[37] T. A. Litzinger, L. R. Lattuca, R. G. Hadgraft, and W. C. Newstetter, "Engineering Education and the Development of Expertise," vol. 100, no. 1, pp. 123–150, 2011.

Appendix I. Rubric used to assess student assignments.

Rubric Category:	Low (1)	Good (2)	Excellent (3)
Coding Style Measures the extent to which the code is presented in a manner that is clearly readable by others. Is the code indented, commented and are variable and function names chosen to enhance readability?	- Code has no to few comments. - Code is not properly indented. - Variable and function names are declared but do not make sense	- Code is adequately commented. - Code is properly indented and variable and function names are well chosen. - Code could be made more readable in one or more ways by additional commenting or by more logically organizing its structure.	- Code is well commented. - Code is properly indented and variable and function names are well chosen. - Code is well structured.
Deployment of Disciplinary Concepts Evaluates whether the student can use the solution to approach a disciplinary problem. Can the student use their code to address the disciplinary issue or to solve a related problem?	- Some solution/explanation provided.	- A solution is provided that would adequately address the issue or problem, but it is presented in a way that is unclear, or improperly documented. (E.g. graphs without axes, no written description when requested.)	- A solution is provided that is correct, clear and well documented.