Learning Communities Research and Practice

Volume 9 | Issue 1 Article 2

2021

Student Experiences within a Data Science Learning Community: A Communities of Practice Perspective

Aparajita Jaiswal
Purdue University, jaiswal2@purdue.edu

Alejandra J. Magana
Purdue University, admagana@purdue.edu

Joseph A. Lyon
Purdue University, lyonj@purdue.edu

Ellen Gundlach
Purdue University, gundlach@purdue.edu

Mark D. Ward
Purdue University - Main Campus, mdw@purdue.edu

Recommended Citation
Jaiswal, A. , Magana, A. J. , Lyon, J. A. , Gundlach, E. , Ward, M. D. (2021). Student Experiences within a Data Science Learning Community: A Communities of Practice Perspective. *Learning Communities Research and Practice*, 9(1), Article 2.
Available at: https://washingtoncenter.evergreen.edu/lcrpjournal/vol9/iss1/2

Student Experiences within a Data Science Learning Community: A Communities of Practice Perspective

Abstract

The discipline of data science is gaining attention due to a significant increase in the amount of data generated on a real-time basis. Despite the efforts from the institutions and national agencies, it has been witnessed that there have been substantial challenges in retaining and attracting students in the discipline of data science. Learning communities have been identified as effective mechanisms to improve student retention. However, fewer studies focus on student experiences within the context of a Data Science Learning Community. This study intends to investigate the student experiences within a Data Science Learning Community known as the The Data Mine (TDM) through the lens of Communities of Practice (CoP) and explore the contribution of the three tenets of CoP: *domain, community, and practice* in shaping student experiences. The study used a mixed-method case study design to understand the experiences of first-year students within TDM and identify the benefits and challenges associated with participating in the living-learning community. The findings of the study revealed that students considered participating in the learning community a valuable experience as it helped them collaborate with like-minded peers and mentors, also helped them develop data science skills.

Keywords

Residential Learning Community, Communities of Practice, First Year Students

Cover Page Footnote

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 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 work was also supported in part by the National Science Foundation Graduate Research Fellowship Program (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, recommendations expressed in this material are those of authors and do not necessarily reflect the views of the funding agencies.

Computational and data science-related academic programs face considerable challenges to attracting and retaining students, in part because of the need for students' positive attitudes toward learning (Felder et al., 1998; Hutchison et al., 2006; Irizarry, 2020; Koby & Orit, 2020; Surakka & Malmi, 2005). Furthermore, regardless of their majors, current professionals need to be empowered to contribute and benefit from the digital economy by taking active roles in our technology-driven society. Research suggests that changes in learning environments, along with teaching and advising methods, can result in improved retention of underrepresented minorities and women and students from all demographic groups (Seymour, 2002). Therefore, efforts must be directed toward improving students' learning experiences that promote a "practice perspective" (Roth & McGinn, 1997, p.92). From a practice perspective, the focus of learning is on participation in authentic experiences, where learning environments: (a) are personally meaningful to the learner, (b) relate to the real world, and (c) provide an opportunity to think in the modes of a particular discipline.

Living-learning communities are academic and social assimilation mechanisms that enable meaningful experiences for students (Hurtado et al., 2020; Stebleton & Jehangir, 2016) as they promote a practice perspective. Such experiences are integral to students' growth (Bobilya & Akey, 2002). Learning communities, thus, provide (a) a context that facilitates the acquisition of knowledge by immersing students into learning experiences (Jessup-Anger, 2015), and (b) collaboration processes and student active involvement, which results in social knowledge construction (Cross, 1998; Zhao & Kuh, 2004). The combination of situated experience and collective learning allows the learner to interact with the social context, creating meaningful learning experiences (Allal, 2001; Dewey, 1902; Jessup-Anger, 2015), and evolving into Communities of Practice (Lave & Wenger, 1991). Through regular and meaningful interactions, learners in a Community of Practice develop common interests sharing a concern or a passion for something they want to do better. By doing so, learners participate actively with other participants such as mentors or peers in the process of learning (Li et al., 2009). In addition to contributing to the collaborative knowledge construction, such social interactions contribute to the development of professional identity (Smith et al., 2017).

Research has revealed that learning communities are an effective institutional initiative that can improve student retention, surges in academic achievement, diminished faculty isolation, and increased curricular integration (Lenning & Ebbers, 1999). But there has been insufficient focus on how the learning community experience impacts student engagement (Hurtado et al., 2020) and cognitive effects associated with student success (Barefoot, 2000; Browne & Minnick, 2005). Therefore, it is crucial to (1) conduct further research on learning communities from the point of view of the student experience (Jaiswal, Lyon, Perera, et al., 2021; Magana et al., 2021), (2) utilize research methods that allow investigating students' opinions and experiences (Bauer & Kiger, 2017; Virtue et al., 2019), and based on the findings, (3) identify opportunities for long-term engagement (Virtue et al., 2019). Such opportunities should facilitate student interaction and learning in a collaborative environment (Halley et al., 2013).

Social engagement acts as a vehicle that promotes academic engagement and learning (Davis & Bost Laster, 2019). Failure to engage in a collaborative learning environment leads to student attrition (Tinto, 1987). Therefore, it is crucial to help students engage in social networks with other community members and help them become academically involved in the learning (Akili, 2021). The academic and social interaction among the students, in addition, helps them develop a sense of belonging and form identities (Carrino & Gerace, 2016), which jointly contribute to academic success and student retention (Flynn et al., 2016; Pike et al., 2008).

Through this study, we explore and investigate the three tenets of the Communities of Practice framework: *domain, community, and practice* (Wenger, 2004) that characterize the students' experiences enrolled in a learning community focused on enabling data science practices by engaging its members (students and mentors) in a Community of Practice herein called The Data Mine (TDM) at Purdue University. The goal of TDM is to form Communities of Practice allowing students from all disciplines, STEM or non-STEM backgrounds, to join the residential learning community and engage in data science practices under the guidance of expert faculties and mentors. This leads to the three research questions

- (RQ1) What were first-year students' experiences within the context of the living-learning community (domain) with regards to their mentor/mentee relationships (community) and their learning of data science concepts and skills (practice)?
- (RQ2) What benefits did the students self-report being a part of a residential learning community?
- (RQ3) What challenges did the students self-report and that need to be addressed to improve their learning experiences?

Background

Residential or Living-Learning Communities (LLC) gained popularity around the 1960s after the G.I. Bill was introduced (Smith, 2001). The introduction of the G.I. Bill led to a surge in student enrollment at Higher Education Institutions (Micomonaco, 2011). LLCs are known to positively impact student success academically and socially (Allal, 2001; Bobilya & Akey, 2002; Cross, 1998; Stassen, 2003). Research (e.g., Bobilya & Akey, 2002; Ujj, 2020) has revealed that residential learning communities improve student engagement by providing an environment where students of similar interests can live and learn together.

Literature investigating learning communities has reported positive outcomes when implemented among students pursuing majors in STEM domains (e.g., Carrino & Gerace, 2016; Freeman et al., 2008; Russell, 2017; Spanierman et al., 2013; Taylor et al., 2008). However, to the best of our knowledge, research on interdisciplinary learning communities with a central practice on data science has not been investigated. Recent works by (Irizarry, 2020; Koby & Orit, 2020) have revealed that the interdisciplinary nature of data science makes the discipline complex. Therefore data science concepts cannot be taught in a year; imparting data science knowledge requires time and effort (Irizarry, 2020). Furthermore, the report by National Research Council reveals that to provide discipline-based education, the traditional classroom education needs to be replaced or supplemented with research-oriented instructions that allow students to collaborate, socialize and work in groups to solve real problems (National Research Council, 2012).

Major institutions across the United States have implemented data science initiatives from the undergraduate level. For example, the Massachusetts Institute of Technology (MIT), New York University (NYU), University of California Berkeley, and the University of Michigan have started data science programs to impart data science knowledge (Donoho, 2017). The National Academies of Sciences (2018) has also highlighted the importance of data science education “to prepare their graduates for this new data-driven era, academic institutions should encourage the development of a basic understanding of data science in all undergraduates” (p.1). Therefore, efforts must be directed to improve student learning experiences, particularly in introductory classes (Heroux & Allen, 2016). This study uses the lens of Communities of Practice to understand the impact of the three prominent characteristics of Communities of Practice that is, domain, community, and practice (Lave & Wenger, 1991; Wenger, 2002, 2004) to comprehend the experiences of the novice learners enrolled in The Data Mine (TDM).

Theoretical Framework

For approaching our three research questions, we adopted the theoretical lens of Communities of Practice to understand the impact of the *domain, community, and practice* on student experiences within the learning community under investigation. Communities of Practice is a “learning partnership among people who find it useful to learn from and with each other about a particular domain. They use each other’s experience of practice as a learning resource” (Wenger et al., 2011, p. 9). Communities of Practice allow studying environments where the students of similar interests collaborate as a group under the guidance of a facilitator or mentor to develop skills and become experts in their domain (Wenger, 2004). The Communities of Practice framework was developed by Lave & Wenger (1991) by using an anthropological lens to understand how adult learners make meaning of their work. They studied Yucatec midwives in the American Indian community, and Vai and Gola tailor in West Africa (Mercieca, 2017). They concluded that individuals learn by socially engaging in the process of becoming a member of a community. Learning occurs within a context where individuals interact and share their knowledge, not when the teacher instructs the pupil in a classroom setting (Mercieca, 2017). According to Lave & Wenger (1991), “learning [is considered] not as a process of socially shared cognition that results, in the end, in the internalization of knowledge by individuals, but as a process of becoming a member of a sustained community” (p. 65). To recognize any community as Communities of Practice they must possess the following three characteristics *domain, community, and practice* (Lave & Wenger, 1991; Wenger, 2002, 2004).

The *domain* symbolizes “the area of knowledge that brings the community together, gives it its identity and defines the key issues that members need to address” (Wenger, 2004, para. 13). It allows members of the community to develop an identity as a group or individual based on the common interest and competence they share (Mercieca, 2017). Over time the group members develop expertise and become experts in their area of interest. The community is “the group of people for whom the domain is relevant, the quality of the relationships among members, and the definition of the boundary between the inside and the outside” (Wenger, 2004, para. 14). The *community* is responsible for the interaction that occurs among the members of the domain within the social context (McCann, 2003). It provides an opportunity for the members to interact, share their knowledge, and learn from one another as they interact (Mercieca, 2017). The community contains a mix of experts and novice learners. The expert learners act as a facilitator or mentor for the novice learners.

Prior research (e.g., Virtue et al., 2019; Zhao & Kuh, 2004) has investigated the benefits provided by learning communities. One of such benefits is that these environments allow the mentors and students to collaborate on intentional activities by promoting socialization and learning. Our study specifically focuses on the Corporate Partner cohort of TDM that is comprised of corporate mentors and first-year data science students. Therefore, in the context of our study, we defined *community* as the interaction between expert learners (corporate mentors) and novice learners (first-year data science students). The *practice* refers to the habit of indulging in the process of learning that requires the members to develop methods, tools, techniques, a body of knowledge, and artifacts (Smith et al., 2017). The members of the community engage in the process development of the tool, techniques, and artifacts and share among them (Smith et al., 2017; Wenger, 2004).

The Communities of Practice framework has framed numerous studies (e.g., Lindsay, 2001; Priest, 2012) as a lens for studying student learning within the context of learning communities. The study by Rehak (2018) used the Communities of Practice framework to investigate the identity formation and learning trajectories of novice undergraduate learners living in the learning community. The study's findings revealed that communities of practice allowed the community members to interact with each other and develop identities of expert learners as they progressed within the learning community. Priest et al., (2016) conducted a study to understand the experiences of participating in a learning community for first-year undergraduate students. The study used a lens of situated learning and communities of practice to explore the motivation level of the students, how they participated and interacted with other members of the community, and how their involvement in the learning community for a year impacted their future career decisions. The results of the study indicate that first-year students found joining the learning community a motivating experience. The learning community environment allowed them to collaborate with faculty and peers and engage in collaborative learning.

The communities of practice framework has also been used in the context of Faculty Learning Communities (FLC). The study conducted by Tinnell et al. (2019) focused on how members in the FLC developed methods to improve their pedagogical approaches for teaching engineering courses. The FLC was created on the principles of Communities of Practice. Participation in the learning community allowed the faculties to learn from one another. The study results demonstrated that participating in the FLC helped the faculties to study results demonstrated that participating in FLC helped the faculties improve their teaching strategies and improve their research group collaborations. Other studies have found similar benefits of participating in a living-learning community, such as building belongingness, leadership abilities, obtaining support, and building broader and more diverse values (Jessup-Anger et al., 2019; Spanierman et al., 2013).

Implications of the Theoretical Framework for the Study Design

Since prior studies (Priest et al., 2016; Tinnell et al., 2019) have revealed the benefits of the integration of learning communities with Communities of Practice, our goal is to use this lens to evaluate the benefits of the The Data Mine (TDM). TDM can be characterized as a learning community that fosters communities of practice by allowing the people of interest (students, mentors, faculties) to come together and engage and work collaboratively to learn new skills in a specific context (Wenger, 2011; Wenger & Wenger, 2015). The Corporate Partner cohort of The Data Mine (TDM) allowed the novice learners, irrespective of disciplines, to join the learning community and develop data science skills by interacting with the expert corporate mentors in the context of corporate projects. Students work on real-world data sets under the guidance and supervision of corporate mentors (Betz et al., 2020). Corporate mentors provide tools and algorithms that could be used to analyze the data sets and also work alongside the students to provide hands-on training for handling large and complex data sets. The engagement of the novice learners with the expert learners within the learning community helps the novice learners to develop the identity of expert learners through legitimate peripheral participation, as they move up in the Communities of Practice (Lave & Wenger, 1991).

The implications of adopting a Communities of Practice approach to our study relate to our goal of establishing a Community of Practice with The Data Mine (TDM). Specifically for this study, we intend to focus on the three key tenets of the Communities of Practice*: domain, community, and practice* (Lave & Wenger, 1991), to characterize The Data Mine (TDM). The *domain* refers to the perception of students regarding their learning of data science skills at TDM; the *community* refers to the perception of students about their mentor after a year-long mentor-mentee interaction, and *practice* refers to the perception of students about the corporate project. The study intends to understand how the three tenets of Communities of Practice can (1) help students of similar interests to come together, (2) develop effective social ties between the mentor and mentee, (3) shape data science identity, and (4) create meaningful learning experiences for the first-year students enrolled in a residential data science learning community. Within this characterization, the study identifies (a) the challenges and benefits that students experienced as part of the learning community, (b) the qualities of mentor that influenced the learners' perceived mentor/mentee relationship, and (c) the perceived learning benefits that students acquired while working on their corporate project.

The Data Mine (TDM)

The context of the study focuses on a data science learning community called The Data Mine (TDM), located at a large midwestern university. TDM is a unique model that brings together students from all majors across the university to live in a learning community focused on foundational skills and computational and data science tools. TDM is orthogonal to a traditional data science major. Data science is geared towards developing disciplinary expertise and (simultaneously) incorporating data science tools and methodologies into student coursework and research in a supportive environment. The program allows the students to live and learn under the same roof.

TDM was piloted with 100 undergraduate students during the 2018-2019 academic year. TDM has scaled up since 2019 to 600 undergraduate students annually. Students are already organized into 20 disciplinary learning communities and 23 corporate partnerships. Each variant has a disciplinary theme that enables students to build domain expertise while simultaneously learning computation and data science.

The design and implementation of the learning experiences within TDM followed guidelines from the cognitive apprenticeship framework. Cognitive apprenticeship is a framework for teaching cognitive skills through a traditional apprenticeship model (Collins et al., 1988). Specifically, TDM implementation was guided by the four principles of cognitive apprenticeship: content, sequencing, method, and sociology, as shown in Table 1.

Table 1. Principles of cognitive apprenticeship that guided the learning design of TDM

Principle	Description	Embodiment
Content	Dimension is the concepts to be learned, both domain and strategic knowledge.	·Computational and data science tools (R environment, Unix, SQL, XML). · Data practices: representing, extracting, manipulating, interpreting, transforming, and visualizing data.
Sequencing	Dimension is that concepts should slowly build on each other throughout the apprenticeship.	·Year-long introductory research course for all students. ·Two-semester sequence of disciplinary courses specific to each learning community.
Method	Dimension includes concepts such as modeling, coaching, and fading of material.	·Students paired with a corporate partner. ·Students participate in a research-based project.
Sociology	Dimension includes social dimensions of the apprenticeship, such as situatedness of the learning, and communities of practice.	·Students live together in residential learning communities. ·Students working together in interdisciplinary teams.

With a focus on corporate partnerships, the program provided a holistic approach for experiencing an authentic context with the learning community. Initially, the program helped students to develop disciplinary knowledge through shared coursework. The faculty and mentors worked together to introduce data science concepts to the students. The next step was to create student groups and assist them in identifying the appropriate project to gain real-world data science experience. The students worked in groups under the guidance of corporate mentors from reputable companies. The mentors helped the students to better understand the use of data science in a real-world environment, engaging them in case studies and problem-solving. The corporate mentors taught them specific computational skills required to address the project problem. They also provided valuable feedback to students to improve their projects and tracked their progress on the weekly basis.

Students in this context reflect on their experiences in TDM. By understanding students' experiences within learning communities, enabled by a cognitive apprenticeship model, this research aimed to better characterize the elements of a community of practice. The learning experience at TDM served as the *domain* as it allowed students and experts referred to as *members,* of similar interests to come together and develop data science skills. The opportunity for socialization and interaction between the members of TDM (in our case, mentor and mentee) within the learning community helped the members create a *community*. Mentors are experienced learners, as they served as a facilitator or guide to scaffold learners and provide them opportunities to succeed in the future. The *practice* referred to students working together and their mentors on real projects to become data science experts. As part of the curricular activities, students in TDM are expected to attend weekly or bi-weekly meetings with their corporate mentors to showcase their progress and complete bi-weekly reports. Team members were encouraged to take on individual roles, including the project manager, coding/data manager, and report manager. Students presented the deliverables of their projects in a research symposium. Students were evaluated based on timely completion of bi-weekly reports, the final poster (included the methods used to collect or analyze data, the analysis of the data, results, and conclusion), and were each given a grade by their corporate mentor. This environment allowed the students to interact constantly with peers, faculty, and mentors that helped them to reflect on their projects and articulate their findings in the research symposium.

Methods

Most of the research performed in the context of learning communities has taken a quantitative approach (Bauer & Kiger, 2017). Prior studies in the context of learning communities, as explained earlier, have identified the benefits of students' engagement within learning communities (Flynn et al., 2016; Rocconi, 2011; Zhao & Kuh, 2004). However, in-depth qualitative research is also needed to identify what elements of such learning communities benefit learners and in what ways (Virtue et al., 2019). Thus, to answer our research questions, this study implemented a case study method to conduct a detailed investigation to understand the phenomenon under the study (Meyer, 2016). This method's flexibility helped us organize the data collection and analysis methods to suit the research question appropriately, and the openness of the approach allowed us to align the research question with the relevant theoretical or conceptual framework (Meyer, 2016). The phenomenon under investigation is the student experiences within TDM, viewed through the lens of Communities of Practice. The bound of the case is one year of student involvement in TDM. The unit of analysis is the individual student involved in TDM. The data was collected in both qualitative and quantitative forms allowing the researchers to explore the student experiences from multiple data sources. The quantitative data were collected first, as students were asked to rate their perceptions concerning the learning experiences within TDM. Data science skills and practices were representative of the *domain*. The role and interaction with their mentors after one year of mentor-mentee relationship was representative of the *community*. Students' learning and experience in completing a year-long project experience was representative of *practice*.

Further qualitative data was collected in the form of written reflections asking the students to justify the ratings provided for the three quantitative measures. Since the study intended to understand students' experiences for the three tenets of Communities of Practice, we grouped the students into clusters using hierarchical clustering algorithms. The analysis was then followed by descriptive statistics within groups and an in-depth qualitative analysis of the student reflections for each cluster using thematic analysis.

Previous research performed by Virtue et al. (2019) and Nassaji (2015) revealed that the use of qualitative analysis had been limited in evaluating the long-term impact of learning communities. Therefore, by using qualitative methods for our study, we wanted to provide a holistic and deeper analysis of the student experiences for each student category. Permission from the Institutional Review Board (IRB) was obtained before conducting the study.

Participants

The participants for the study were 63 first-year students living in TDM. All the participants have completed their first two semesters in TDM and belonged to the corporate partner (CP) cohort. The CP cohort allowed all students to work on a corporate project under the guidance of a corporate mentor. The data collection process was de-identified; therefore, we do not have any information regarding student demographics.

Procedures and Data Collection Method

The data was collected in the form of written reflections. The CP cohort students lived in TDM and worked alongside their industry mentors on a Data Science Project. Students were asked to submit a written reflection as a part of the course of the assignment at the end of their first year. The objective of the reflection was to obtain an overall perspective on student experiences with context, mentor, and data science project (i.e., domain, community, and practice, respectively). The written reflection question had two important components: (1) a quantitative component referred to as experience rating presented in Table 2, and (2) a qualitative component that required students to justify their ratings and was referred to as reflection response, presented in Table 3.

Experience Rating

TDM required every student to submit a written reflection in the form of an experience rating at the end of their first year. The three rating questions are representative of three key tenets of the Communities of Practice theory as shown in Table 2. Students were required to rate their experiences on a scale from one to five.

Table 2. Key tenets of Communities of Practice (CoP) and Experience Rating Questions.

CoP Tenet	Experience Rating Questions
Domain	How would you rate your experience at The Data Mine? [5 would be an exceptional experience, 1 would be poor experience]
Community (Mentor/Mentee interaction)	How would you rate your Corporate Partner mentor? [5 would be an exceptional experience, 1 would be poor experience]
Practice	How would you rate your Corporate Partner project? [5 would be an exceptional experience, 1 would be poor experience]

Reflection Responses

The students were also asked to reflect on and rationalize their TDM ratings, as shown in Table 3. Written responses for all 63 students were analyzed by conducting thematic analysis. A total of 189 responses were analyzed qualitatively. The average length of one student response for each category was domain 144 words, community 116 words, and practice 135 words.

Table 3. Key tenets of Communities of Practice (CoP) and Reflection Prompts

CoP Tenet	Questions for Reflection Responses
Domain	Please justify your ratings for your experience at The Data Mine.
Community (Mentor/Mentee interaction)	Please justify your ratings for your Corporate Partner mentor.
Practice	Please justify your ratings for your Corporate Partner project.

Data Analysis Method

The data analysis procedure consisted of a combination of hierarchical clustering, descriptive statistics, and thematic analysis approaches. The data analysis was divided into two sections: quantitative data analysis and qualitative data analysis. To answer our first research question (RQ 1) the quantitative data analysis was performed. The quantitative data analysis involved clustering the experience ratings into three categories, followed by descriptive statistics. Further, to answer the research question two (RQ 2) and three (RQ 3) qualitative data analysis was conducted. To analyze the qualitative data in the form of reflective responses, the study used thematic analysis. The thematic analysis was used to take a deep dive into the data and identify emerging themes.

Methods for Quantitative Data Analysis

The first step for analysis was to conduct a hierarchical clustering to group 63 students based on their ratings for domain, community, and practice. Hierarchical clustering using the Ward method was used to cluster the students. Hierarchical Clustering is a widely used technique used in educational data mining. Prior studies Rodrigues et al., (2016) have used hierarchical clustering to detect patterns of student behavior while engaged in online learning. The study used the Ward method to examine the characteristics of student engagement on the online forum. (Li et al., 2009) used hierarchical clustering to profile students based on their behavior in the online learning environment to help create clickstream data. Recent studies (Antonenko et al., 2012; Jaiswal, Lyon, Zhang, et al., 2021; Magana et al., 2021; Medová & Bakusová, 2019) also demonstrated that hierarchical clustering methods such as Ward’s minimum variance can be used for small sample sizes such as 59 students in case of Antonenko et al., (2012) and 30 in-service teachers in the study by Medová & Bakusová (2019). Considering the descriptive nature of this current study, and a sample size of 63 students, the study found Ward hierarchical clustering, an appropriate methodology to group students based on the student experience ratings. The student response from all 63 students for all the three questions, refer to table 1, were clustered using hierarchical clustering. Further, the descriptive statistics for each cluster were reported in terms of mean and standard deviation in Table 5.

The study used the elbow method to identify the optimum number of clusters (Yuan & Yang, 2019). Identifying the optimal number of clusters is crucial before conducting the clustering process, as the optimal clusters explain the variance in the data. The elbow method revealed that the optimal number of clusters for this study was three clusters. Based on the optimal number of clusters obtained, the students were divided into three categories based on their scores. The three categories were: *moderate experience, good experience, and exceptional experience.* Table 3 represents the mean and standard deviation.

Methods for Qualitative Data Analysis

The qualitative data was in the form of written reflections; each student's reflections were evaluated using thematic analysis to identify the overarching themes. Thematic analysis was conducted for each tenet of the Communities of Practice framework. The thematic analysis process that was followed is outlined by Braun and Clarke (2006). The six steps that were followed for the thematic analysis follow (1) getting familiar with the data, (2) producing initial codes, (3) examining for themes, (4) reassessing the generated themes, (5) defining and identifying themes, and (6) creating the final report. Two researchers independently coded the data first. Later the two researchers met and created the final codebook. Based on the final codebook the data was re-coded, and themes were identified from the agreed-upon codes. The percentage agreement was calculated for establishing the inter-rater reliability for the two raters was 78 percent. The qualitative data was first categorized into codes (refer to Appendix). Codes that contributed more than 5% of the frequency were considered for generating themes. After this was done, the coded items were grouped into themes. The Appendix contains the definition and example of each theme.

Results

The results of the study are organized into three sections: the first section presents the descriptive statistics representing the overall student experiences. The second section presents the results from the cluster analysis representing the three levels of student experiences: *moderate, good, and exceptional* experiences. The third section describes in-depth quantitative and qualitative results for each experience category.

Overall student experiences

Descriptive statistics were used to analyze the three key tenets of Communities of Practice. Table 4 below represents the mean and standard deviation for the three tenets of Communities of Practice.

Table 4. Descriptive statistics for the three key tenets of Communities of Practice (CoP).

CoP Tenet	Experience Rating Questions	Mean	SD
Domain	How would you rate your experience at The Data Mine? [5 would be an exceptional experience, 1 would be poor experience]	4.15	0.75
Community	How would you rate your Corporate Partner mentor? [5 would be exceptional experience, 1 would be poor experience]	4.65	0.75

Practice	How would you rate your Corporate Partner project? [5 would be an exceptional experience, 1 would be poor experience]	4.20	0.75

Overall average scores for the three tenets of Communities of Practice were interpreted as follows: mean scores from 0.00 to 2.50 were classified as poor experiences, between 2.51 to 3.95 were classified as fair experiences, between 3.96 to 4.50 were classified as average experiences and 4.51 and above were referred to as excellent experiences. From Table 4 above, we can observe that scores for the *domain* and *practice* were perceived as average experiences, and the score for the interaction with the mentor (*community*) was perceived as excellent.

Three Levels of the Student Experience

Hierarchical clustering was used to further explore students' experiences and categorize the overall student experiences into similar patterns. The output of Ward's minimum variance approach resulted in three clusters:

Cluster 1: is the group of students that had an *exceptional experience* in The Data Mine

Cluster 2: is the group of students that had a *good experience* in the The Data Mine

Cluster 3: is the group of students that had a *moderate experience* in the The Data Mine

Table 5 below represents the scores for each tenet of Communities of Practice. The *exceptional experience* cluster students demonstrated the highest scores regarding their experience with excellent perceptions regarding their experiences with all the three tenets of Communities of Practice. The *good experience* cluster students demonstrated average scores for *domain and practice* and excellent perceived experience for the *community*. The *moderate experience* cluster students demonstrated fair scores for *domain and practice* and average scores for the perceived *community* experience.

Table 5. Moderate, Good, and Exceptional experience clusters and descriptive statistics

	Moderate Experience (N= 20)		Good Experience (N=23)		Exceptional Experience (N= 20)	
CoP Tenet	Mean	SD	Mean	SD	Mean	SD
Domain	3.35	0.65	4.45	0.45	4.55	0.50
Community	4.15	1.10	4.60	0.35	5.00	0.00
Practice	3.55	0.80	4.10	0.35	5.00	0.00

Figure 1 below represents the box plots for the three-student experience cluster for each tenet of Communities of Practice. The box plots in the figure reveal that students in the *moderate and good experience* categories faced challenges with the *practice* tenet, whereas students in the *exceptional experience* category had rewarding experiences. With regards to *domain* tenet, students in the *good* and *exceptional experience* categories had a mostly positive experience, but students in the *moderate experience* category faced challenges. Lastly, with the *community* tenet, students in all the *good* and *exceptional experience* categories demonstrated a rewarding experience, but students in the *moderate experience* category faced some minimal challenges.

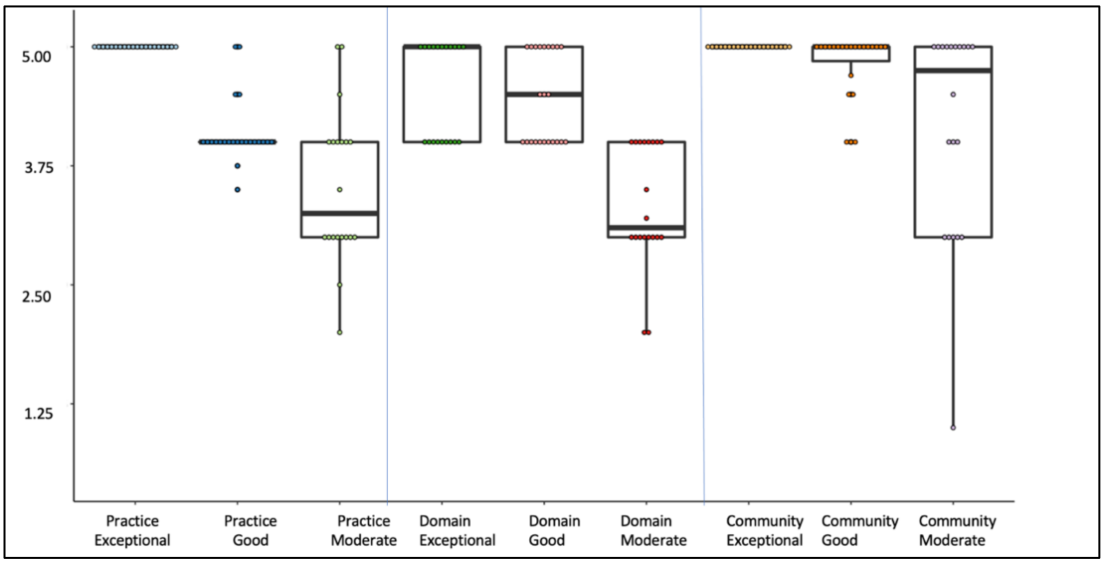

Figure 1: Boxplots representing the three Clusters and three tenets of Communities of Practice

Experience of students in Moderate-Experience cluster

The descriptive statistics in Table 5 suggest that *moderate-experience* students demonstrated a fair perception of *domain* (*M*= 3.35, *SD* = 0.65), and *practice* (*M*= 3.55, *SD* = 0.80), but an average perception with *community* (*M*= 4.15, *SD* = 1.10). Results from the thematic analysis performed on descriptions provided by students in the *moderate-experience* group demonstrate that students had a fair learning experience overall but also faced substantial challenges for the *domain, community, and practice* tenets. The challenge faced with the *domain* tenet was related to the lack of prior knowledge; around 20% of the students reported this as a challenge. Other challenges reported with the domain were lack of structure (15% of students), the inadequacy of learning material (25% of students), and personal issues (35% of students). The challenge encountered with the *community* tenet consisted of mentors lacking communication skills (20% of students). Lastly, the challenges posed by the *practice* tenet included students that lacked technical skills (25% of students), lack of structure (15% of students), and clarity concerns (35% of students) raised by students while working on the project. Table 6 represents the main patterns regarding the benefits and challenges that emerged for *the moderate-experience group's domain, community, and practice*. Table 6 also presents some representative student quotes.

Table 6. Results of the thematic analysis for the Moderate-Experience Group

CoP Tenet	Main Pattern	Quotes
Domain	Participation in the learning community was overall a fair experience as that allowed students to collaborate and develop data science skills, but students also reported challenges such as lack of prior knowledge, inadequate learning material, and lack of structure.	*I felt like I learned a lot about data analysis and data science methods. However, the lack of structure caused me to lots of work perhaps more hands-on mentorship could have saved me. I also felt lots of stress trying to deliver week after week. The seminar material also felt like busywork.*
Community	Students found the mentor-mentee interaction valuable as mentors were supportive, committed, and knowledgeable, they experienced issues with communication and guidance.	*I would rate my Corporate Partner mentor a four out of five. This is because he was a good mentor and provided us with lots of context as to how the data, we worked with, was collected and what we should be looking at. However, he occasionally did not have a general path we should follow as at certain times we were stumped without some background research. At certain times, he also seemed to have no idea where this project was going making it difficult to figure out what he is asking from us in the process. Overall, he was a great mentor in general.*
Practice	Students found the project experience valuable and motivating, as they used data science skills to solve real problems. But the inadequacy of technical skills on the part of students, lack of structure, and clarity posed some challenges.	*I would rate our project as a 4/5. The concepts were interesting to explore but there was a general lack of direction at times. I was unaware of these applications before joining.*

From the main patterns described in Table 6, the overall experience for this group can be described by the following Theme: For the students in the *moderate-experience* group, participation in the learning community was a blend of benefits and challenges that impacted student's learning experience. The context allowed them to collaborate with peers, learn under the guidance of supportive and knowledgeable mentors, and work on data science projects to develop data science skills. The factors that affected the learning experiences were, students lacked technical/prior coding knowledge, experienced personal issues, had communication issues with the mentors, and lacked adequate guidance on the project.

Experience of students in Good-Experience cluster

The descriptive statistics in Table 3 reveal that students in the *good-experience* group classified their experience with scores for the *domain* (*M*= 4.45, *SD* = 0.45) and *practice* (*M*= 4.10, *SD* = 0.35) as average experience. Whereas experience with the *communit*y (*M*= 4.60, *SD* = 0.35) as an excellent experience. The results from the thematic analysis performed on descriptions provided by the *good-experience* group led to three themes indicative of each of the tenets of Communities of Practice. The themes for the *good-experience* group suggest that students had an overall great experience and expressed minimal challenges for the *domain* and a fair number of challenges for the *practice* tenets. The challenge faced with the *domain* was related to the slow start of the semester reported by 26% of students. Lastly, for the *practice,* tenet students reported as the main challenge a lack of structure (17% of students) and difficulty in procuring the project data (22% of students). Table 7 represents the main patterns regarding the benefits and challenges that emerged for *domain, community, and practice* for the good-experience group. Table 7 also presents representative student quotes.

Table 7. Results of the thematic analysis for the Good-Experience Group

CoP Tenet	Main Pattern	Quotes
Domain	Students had a great learning experience at the Data Mine, as that helped students develop data science skills, professional skills, and work collaboratively with peers on real-world problems, students reported some minor challenges as the project start was slow.	*I think, despite not being able to do any work the first semester, that it was still a valuable experience. The project made good use of what I had learned in previous semesters in TDM at Purdue University, and I got to experience the kind of data-driven issues that a company (one in my field of study, no less) deals with on a regular basis. It showed students that real data is never quite as sanitized as we might deal with in TDM, and it demonstrated how important data can be even to companies that are not "data science" companies.*

Community	Students found mentors as good leaders as they were supportive, knowledgeable, and committed.	*I would rate my Corporate Partner mentors a 5 overall as well. [MENTOR 1] and [MENTOR 2] were fantastic mentors. They provided very great guidance and criticism where it was needed, and helped lead me to insights I would not have seen myself with my data analyses. They were also quite flexible with everything going on, and were overall a pleasure to work with. Even when COVID complicated things in the latter part of the semester, they ensured that work for our project could continue. I believe they are truly what made my experience in the Corporate Partners program so valuable, and I would not have found the program as fulfilling without their help.*
Practice	Working on a real project was a motivating and valuable experience for students, as students acquired new knowledge, but lack of structure and data issues brought up some challenges.	*I liked the open-ended nature of the project that we were given, which was identifying potential reasons and solutions for a shortage of [COMPANY] personnel. However, that same open-endedness did at times leave me wondering what to do and where to go as there were no real defined steps to follow. Overall, I felt it was a good experience but I would have liked a little more guidance in my work.*

From the main patterns described in Table 7, the overall experience for this group can be described by the following Theme: For students in the *good-experience* group, participation in the learning community illustrated higher benefits and fewer challenges. Students found the learning experience valuable. It allowed them to interact with like-minded peers, work with knowledgeable and committed mentors, and feel motivated to work on real-world projects. Students raised some concerns regarding the initial pace of the project, the structure of the project, and data procurement.

Experience of students in the Exceptional-Experience cluster

The descriptive statistics suggest that *exceptional-experience* group had an excellent experience with all the three tenets of Communities of Practice: domain (M= 4.55, SD = 0.50), community (M= 5.00, SD = 0.00) and practice (M= 5.00, SD = 0.00). The results from the thematic analysis performed on descriptions provided by the *exceptional-experience* group led to three themes indicative of each of the tenets of Communities of Practice. The themes for the exceptional experience group demonstrate that students had overall an excellent experience and did not face any challenges while interacting with mentors (*community*) or working on the project (*practice*). Around 10% of students expressed challenges related to the *domain* as they found the start of the semester a bit unsteady as the first weeks were a bit confusing. Table 9 represents the main patterns regarding the benefits and challenges that emerged for *domain, community, and practice* for the exceptional-experience group. Table 8 also presents representative student quotes.

Table 8. Results of the thematic analysis for the Exceptional-Experience Group

CoP Tenet	Main Pattern	Quotes
Domain	Students in the learning community had a great learning experience, developed professional, data science skills, working on a real and unique project by collaborating with students of similar interests but few students found the first weeks a bit unsteady and confusing.	*This is honestly a wonderful program, and I think our project specifically is the reason why. Our project was part of something that could help students all over campus that are just like us, and if [project] becomes a bigger deal on campus I can say that I had something to do with it. This experience was priceless.* *I would rate my experience as a 5/5. I had an exceptional experience working in TDM and the Corporate Partners program. TDM taught me a lot about R and programming, which I was able to use working with my corporate partner. The program allowed me to gain work experience and it served as a transition from the classroom to the workforce.* *I liked that I had the opportunity to work with cutting-edge development in NLP, an experience that would be extremely rare among lower class undergraduates. However, it was confusing at first as to what is expected each week.*
Community	Students found mentors supportive, knowledgeable, and actively committed towards student's success	*[MENTOR] was an exceptional mentor. Her goal was to make us learn and succeed at our tasks. To facilitate this, she was available on Slack most of the time, and even included office hours to discuss issues that we were having. She is also an approachable person and is overall a great person to work with.*
Practice	Students found project experience valuable as they worked in teams on real projects, used data science skills, and acquired new knowledge.	*The project was challenging and truly encapsulated the definition of a project. This project is something that can be accomplished by the end of the school year, only if there is a team working on it. The amount of knowledge gained from this project really sets us above the knowledge curve amongst our fellow peers, with respect to deep learning and group work. I am truly grateful for participating in this project.*

From the main patterns described in Table 8, the overall experience for this group can be described by the following Theme: For the students in the *exceptional-experience* group, participation in the learning community primarily resulted in benefits and reported negligible challenges. Students found the learning experience great and valuable. It allowed them to work on real problems, use their prior coding knowledge, and learn new skills by working closely with knowledgeable, active, and committed mentors. The only challenge reported was that students found the first few weeks were a little confusing in deliverable expectations.

Discussion

TDM focused on creating a Communities of Practice that allowed the members to interact, socialize, and learn in the context of corporate data science projects. The regular interaction among the members within the Communities of Practice instilled a sense of belonging and helped them to develop identities (Carrino & Gerace, 2016; Lave & Wenger, 1991). The results of the study indicate that TDM had an overall positive impact on the students' experiences. The perception of students with regards to the *domain, community, and practice* characteristics of this Community of Practice ranged from fair to exceptional, whereas no students reported having negative or poor experiences. The reason that students did not report negative experiences can be explained from the perspective of communities of practice. TDM allowed students to come together, interact by immersing themselves in "authentic learning" experiences under the guidance of their knowledgeable mentors (Shaffer & Resnick, 1999, p.197). The mutual engagement among the students and mentors helped the novice learners to participate in this community by developing data science skills and become mature learners (Lave & Wenger, 1991). Furthermore, the constant interaction among the members of the community made the learning experiences valuable (Lave, 2004; Lave & Wenger, 1991; Wenger, 2004; Wenger & Wenger, 2015). This perceived value among the most novice of the learners allowed all students to report relatively high scores in terms of *domain, community,* and *practice.*

Patterns from the thematic analysis revealed that overall, students reported multiple positive experiences for all the three elements of Communities of Practice. Specifically, the students in the *moderate-experience* category reported both benefits and challenges for all the tenets of Communities of Practice. The benefits focused on the learning experience that helped students to develop data science skills. They also found that the mentors were supportive and that working on the project was a valuable experience. The challenges for the *moderate-experience* group were the blend of issues that students faced within the learning community and their personal issues. For example, students reported challenges such as inadequate learning material, difficulty communicating with mentors, personal issues such as course load, and lack of prior programming knowledge. This lack of prior coding knowledge continues to be an issue amongst university students engaging in the computational sciences and can often be linked to overpacked curriculum and from disciplinary instructors' lack of coding experience or lack of belief in the importance of programming (Magana & Coutinho, 2017). However, it appears that learning within a community of practice may provide the necessary support to overcome many of these hurdles that learning programming can provide.

In contrast, the benefits increased for students in the moderate-experience and the exceptional-experience groups, and challenges were almost negligible. The theme of the *exceptional experience* group identified that the learning experience was precious as it allowed students to collaborate with others of similar interests. They also found the mentors very helpful and the project interesting. Furthermore, none of the students in the *exceptional-experience* group reported issues such as lack of technical knowledge, or prior knowledge of coding, or any personal issues. They found themselves motivated and learning experiences valuable that helped them develop data science skills. One possible explanation for the observed differences between the three groups can be attributed to the prior data science knowledge students in the *good* and *exceptional-experience* groups had, whereas students in the *moderate-experience* group lacked the knowledge. One of the possible effects of the lack of prior and technical skills on the part of the *moderate-experience* group was that it made it challenging for them to understand the guidance provided by the mentor. The mentor served as a knowledgeable other to help students learn (Hausfather, 1996; Vygotsky et al., 1978). However, the knowledge students could develop through the guidance of their mentor is a function of their previous knowledge (Hausfather, 1996; Vygotsky et al., 1978). Students in the *moderate-experience* group experience limited development of their expertise in the area of data science, as they lacked prior knowledge, whereas students in the *good* and *exceptional-experience* groups were able to build on their existing knowledge with the help of mentors and engage in constructivist learning.

A second pattern observed across the three groups was that students considered the relationship with their mentors valuable. The results from the descriptive statistics in Table 5 suggest that experience with the mentor ranged from average to excellent. Students in the *moderate-experience* group found their experience with the mentors as average, whereas students in the *good*- and *exceptional-experience* groups reported their mentor experiences as excellent. The thematic analysis provided further details as only 20% of students in the *moderate-experience* group reported communication and guidance issues with their mentors, whereas the other two groups did not report any significant issues. This result aligns well with the literature, which has shown that mentoring can be important in improving both academic achievement and attitudes; that is to say, without a strong mentoring relationship, student attitudes will likely not see as positive results (Jekielek et al., 2002). From the perspective of communities of practice, the social setting such as TDM allows the mentor and mentee to work together and create a shared repertoire of ideas, artifacts, and tools that could be used by other members of the community (Holland, 2018). In this case, students and mentors worked on the corporate projects, and they developed learning artifacts such as codes, data science tools to solve real problems.

Recent studies (e.g., Bottoms et al., 2020; Holland, 2018) have revealed that prolonged mentor-mentee interaction promotes the mentor and mentee to develop an informal relationship out of the social context and exchange knowledge. Since the students worked with mentors for a year, many of them developed an informal relationship as they were engaged in learning and knowledge sharing. The observed relationship-building can be explained under the lens of brokering (Bottoms et al., 2020). Brokering allows the members from two different domains (in our case, peer, faculty, and corporate mentors and undergraduate students) to come together to create a shared repertoire of knowledge and understanding. The learning community allowed the mentors and mentees to cross their learning boundaries, learn together, and reach the consensus to share and learn together. That is, it provided an opportunity to the mentors and mentees to negotiate and renegotiate the meaning of learning through constant interaction to the point of developing a collaborative relationship (Bottoms et al., 2020). The active nature of the learning community promoted the interaction between the mentors and mentees and helped each other develop their own identities within the Community of Practice (Lave & Wenger, 1991).

Finally, our results indicate that mentoring experiences are effective as long as students have the prerequisite knowledge to grow and learn. The theories of Vygotsky indicate that knowledgeable others are critical to helping students learn and develop within their zone of proximal development (Hausfather, 1996; Vygotsky et al., 1978). Our results show that students who had the prerequisite knowledge (good and exceptional-experience students) reported more personal learning and growth due to their learning community and mentee experience. Instructors should seek out mentors and learning opportunities for students that align with their current level of knowledge. Additionally, mentors can be even more beneficial for minority groups (such as females in data science) to help improve student engagement and self-efficacy to be successful in the field (Alvarado et al., 2012).

Implications for Teaching and Learning

The results of this study have multiple implications for how institutions could implement data science instruction inside and outside of the classroom. First, given that learning is a highly social process, students should be put in situations where they are allowed to collaborate similarly to that of a learning community. Research has shown time after time that collaboration, such as that among members of a learning community, produces beneficial results for the students learning as well as other soft skills such as communication (Friedman & Alexander, 2007; Zhao & Kuh, 2004). This is especially true in virtual and distance learning, where face-to-face interaction within the classroom is difficult, if not impossible (Swan, 2002).

To improve the experiences of the *moderate-experience* group, TDM needs to impart the basic knowledge to the students with no prior experience and then gradually move them to real projects. One of the approaches could be the further application of cognitive apprenticeship theory in imparting content knowledge (Brown et al., 1989; Collins, 2006). Additional scaffolding could provide learners and mentors with tools to help the most novice of students the ability to build up their programming knowledge during the project so that it becomes less of a challenge and impact to their experience.

Another challenge of the *moderate-experience* group was the reported communication and guidance received from the mentor. To improve this mentor-mentee relationship, instructors can increase the structure of the meetings and make sure the meetings are driven by the needs of the mentees (Jekielek et al., 2002). Additionally, our results indicate that instructors should look to make projects and courses as tied to real-world problems and context as possible. This helps students learn that their learning is tied to previous experiences they had (Dewey, 1986). It also increases motivation for students in that they will see the value of the material they are learning (Ryan & Deci, 2000).

Other educational frameworks within engineering education have shown that realistic contexts for problems are beneficial for students (Hjalmarson et al., 2006; Lyon, 2020, 2021). Not only is this realistic and applied context good for learning, but it has been shown to help increase participation among the gender gap in data science and computing (Alvarado et al., 2012). Research also suggests that the additional layer of living-learning communities may help to further reduce the persistent gender gap in STEM by providing additional support (Pace et al., 2008).

Learning communities, specifically in a data science context, were able to take this one step further and not just provide realistic contexts but very real-world practice. Even though the study focused on a data science learning community, the results of the study are also applicable to learning communities in general. Any learning community can function as a Community of Practice by (1) identifying a relevant *domain* that brings the people of similar interest together, (2) providing engagement opportunities to collaborate and form a *community*, and (3) focusing on creating a "socially constructed" learning (Zhao & Kuh, 2004; p.117), that allows the members to learn from one another and develop tools, stories, and experiences as a part of the *practice* (Lave & Wenger, 1991). The engagement of the members within the Communities of Practice will not just promote a sense of belonging and identity formation but also help to retain students (Weidman et al.,

2014) and make learning happen in a socially constructed environment (Zhao & Kuh, 2004).

Conclusion, Limitations, And Future Work

This study helped to understand students' experiences living in a residential learning community under the Communities of Practice perspective. In addition, the emphasis of the learning community on the data science education context is a novel addition to the body of knowledge. The study's findings demonstrated that the learning community's situated nature allowed the students to collaborate, interact with mentors and peers, and provided the opportunity to work on real-world projects sponsored by corporate partners. The interaction among the members in a data science context allowed the students to develop data science skills from both experts and peers. While not every student had exceptional experiences, very few students had negative feedback regarding their experiences in the learning community. The results of this study, along with others in the literature, indicate that learning communities can be a very beneficial construct for student learning, engagement, and motivation. We expect that the results of this study will continue to help educators design and implement learning community interventions in their institutions.

Limitations of the study are those associated with qualitative research in that the study occurred in a specific context and is a particular case of a living-learning community. In addition, sample bias and sample size resulted in our findings not being statistically representative. Furthermore, since the study focused on understanding the student experiences, we can only make claims about the students' experiences and not on their actual learning gains. However, by following a qualitative approach, our study provided us with the flexibility to gather insights from a student's perspective, allowing us to understand their opinions.

For future work, we plan to increase our sample size to conduct a mixed-method study. We will also perform a quantitative study, allowing us to compare two programs, one program within a learning community and another program in a traditional major, to understand the impact of the two settings on the students' sense of belonging and identity formation.

References

Akili, W. (2021). *Engaging Students in Learning through Cooperative Learning Strategies*. Paper presented at 2010 North Midwest Section, Minnesota State University, Mankato, Minnesota. 10.18260/1-2-1114-36451

Allal, L. (2001). *Situated cognition and learning: From conceptual frameworks to classroom investigations*. https://doi.org/10.5169/seals-786642

Alvarado, C., Dodds, Z., & Libeskind-Hadas, R. (2012). Increasing women's participation in computing at Harvey Mudd College. *All HMC Faculty Publications and Research*. https://scholarship.claremont.edu/hmc_fac_pub/889

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, *60*(3), 383–398. https://doi.org/10.1007/s11423-012-9235-8

Barefoot, B. O. (2000). The first-year experience: Are we making it any better? *About Campus*, *4*(6), 12–18.

Bauer, R. N., & Kiger, S. (2017). Nursing living-learning communities and student retention: A qualitative study. *Nursing Education Perspectives*, *38*(2), 75–79.

Betz, M., Gundlach, E., Hillery, E., Rickus, J., & Ward, M. D. (2020). The next wave: We will all be data scientists. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *13*(6), 544–547.

Bobilya, A. J., & Akey, L. D. (2002). An evaluation of adventure education components in a residential learning community. *Journal of Experiential Education*, *25*(2), 296–304. https://doi.org/10.1177/105382590202500208

Bottoms, S. I., Pegg, J., Adams, A., Risser, H. S., & Wu, K. (2020). Mentoring within communities of practice. In *The Wiley International Handbook of Mentoring* (pp. 141–166). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119142973.ch10

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Browne, M. N., & Minnick, K. J. (2005). The unnecessary tension between learning communities and intellectual growth. *College Student Journal*, *39*(4), 775–784.

Carrino, S. S., & Gerace, W. J. (2016). Why STEM learning communities work: The development of psychosocial learning factors through social interaction. *Learning Communities Research & Practice*, *4*(1), 3.

Collins, A. (2006). Cognitive apprenticeship. In *The Cambridge handbook of: The learning sciences* (pp. 47–60). Cambridge University Press.

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children*, *8*(1), 2–10.

National Research Council, (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. National Academies Press.

Cross, K. P. (1998). Why learning communities? Why now? *About Campus*, *3*(3), 4–11. https://doi.org/10.1177/108648229800300303

Davis, M., & Bost Laster, B. (2019). Acquiring social capital: Conclusions from a social science summer bridge community. *Learning Communities Research and Practice*, *7*(2), 4.

Dewey, J. (1902). School as Social Center. *The Elementary School Teacher*, *3*(1), 73-.

Dewey, J. (1986) Experience and Education. *The Educational Forum*, *50*(3), 241-252, DOI: 10.1080/00131728609335764

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766.

Felder, R. M., Felder, G. N., & Dietz, E. J. (1998). A longitudinal study of engineering student performance and retention. V. Comparisons with traditionally-taught students. *Journal of Engineering Education*, *87*(4), 469–480.

Flynn, M. A., Everett, J. W., & Whittinghill, D. (2016). The impact of a living learning community on first-year engineering students. *European Journal of Engineering Education*, *41*(3), 331–341.

Freeman, K. E., Alston, S. T., & Winborne, D. G. (2008). Do learning communities enhance the quality of students' learning and motivation in STEM? *The Journal of Negro Education*, 227–240.

Friedman, D. B., & Alexander, J. S. (2007). Investigating a first-year seminar as an anchor course in learning communities. *Journal of The First-Year Experience & Students in Transition*, *19*(1), 63–74.

Halley, J., Heiserman, C., Felix, V., & Eshleman, A. (2013). Students teaching students: A method for collaborative learning. *Learning Communities Research & Practice*, *1*(3), 7.

Hausfather, S. J. (1996). Vygotsky and Schooling: Creating a Social Context for Learning. *Action in Teacher Education, 18*(2). DOI: 10.1080/01626620.1996.10462828Heroux,

Heroux, M., & Allen, G. (2016). *Computational Science And Engineering Software Sustainability And Productivity (CSESSP) Challenges Workshop Report*. https://www.nitrd.gov/pubs/CSESSPWorkshopReport.pdf.

Hjalmarson, M., Diefes-Dux, H. A., Bowman, K., & Zawojewski, J. S. (2006). Quantifying aluminum crystal size part 2: The model-development sequence. *Journal of STEM Education: Innovations and Research*, *7*, 67–73.

Holland, E. (2018). Mentoring communities of practice: What's in it for the mentor? *International Journal of Mentoring and Coaching in Education*, *7*. https://doi.org/10.1108/IJMCE-04-2017-0034

Hurtado, S. S., Gonyea, R. M., Graham, P. A., & Fosnacht, K. (2020). The relationship between residential learning communities and student engagement. *Learning Communities Research and Practice*, *8*(1), 1–18.

Hutchison, M. A., Follman, D. K., Sumpter, M., & Bodner, G. M. (2006). Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education*, *95*(1), 39–47.

Irizarry, R. A. (2020). The role of academia in data science education. *Harvard Data Science Review*, *2*(1). https://doi.org/10.1162/99608f92.dd363929

Jaiswal, A., Lyon, J. A., Perera, V., Magana, A. J., Gundlach, E., & Ward, M. D. (2021). Work in progress: Evaluating student experiences in a residential learning community: A situated learning perspective. *2021 ASEE Virtual Annual Conference Content Access*.

Jaiswal, A., Lyon, J. A., Zhang, Y., & Magana, A. J. (2021). Supporting student reflective practices through modelling-based learning assignments. *European Journal of Engineering Education*, 1–20.

Jekielek, S. M., Moore, K. A., Hair, E. C., & Scarupa, H. J. (2002). Mentoring: A promising strategy for youth development. *Child Trends Research Brief*, *2*, 1–8.

Jessup-Anger, J., Armstrong, M., Kerrick, E., & Siddiqui, N. (2019). Exploring students' perceptions of their experiences in a social justice living–learning community. *Journal of Student Affairs Research and Practice, 56*(2), 194–206.

Jessup-Anger, J. E. (2015). Theoretical foundations of learning communities. *New Directions for Student Services*, *2015*(149), 17–27. https://doi.org/10.1002/ss.20114

Koby, M., & Orit, H. (2020). *Ten Challenges of Data Science Education*. https://cacm.acm.org/blogs/blog-cacm/246219-ten-challenges-of-data-science-education/fulltext

Lave, J. (2004). *Situating learning in communities of practice* (pp. 63–82). American Psychological Association. https://doi.org/10.1037/10096-003

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lenning, O. T., & Ebbers, L. H. (1999). The powerful potential of learning communities: Improving education for the future. *ASHE-ERIC Higher Education Report*, *26*(6). ERIC.

Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., & Graham, I. D. (2009). Evolution of Wenger's concept of community of practice. *Implementation Science : IS*, *4*(1), 11–11. https://doi.org/10.1186/1748-5908-4-11

Lindsay, L. N. (2001). *Transformation of learners in a community of practice occupational therapy fieldwork environment.* University of Georgia.

Lyon, J. A., & Magana, A. J. (2020). A review of mathematical modeling in engineering education. *The International Journal of Engineering Education*, *36*(1), 101–116.

Lyon, J. A., & Magana, A. J. (2021). The use of engineering model-building activities to elicit computational thinking: A design-based research study. *Journal of Engineering Education*. https://doi.org/10.1002/jee.20372

Magana, A. J., Jaiswal, A., Madamanchi, A., Parker, L. C., Gundlach, E., & Ward, M. D. (2021). Characterizing the psychosocial effects of participating in a year-long residential research-oriented learning community. *Current Psychology*, 1–18.

Magana, A. J., & Silva Coutinho, G. (2017). Modeling and simulation practices for a computational thinking-enabled engineering workforce. *Computer Applications in Engineering Education*, *25*(1), 62–78.

McCann, E. P. (2003). *Passion and persistence: A study of communities of practice in schools*. University of Wisconsin–Madison.

Medová, J., & Bakusová, J. (2019). Application of hierarchical cluster analysis in educational research: Distinguishing between transmissive and constructivist oriented mathematics teachers. *Statistika-Statistics and Economy Journal*, *99*(2) 142-150.

Mercieca, B. (2017). What is a community of practice? In *Communities of practice* (pp. 3–25). Springer.

Meyer, C. B. (2016). A Case in Case Study Methodology: *Field Methods*. https://doi.org/10.1177/1525822X0101300402

Micomonaco, J. (2011). *Living-learning communities as an intervention to improve disciplinary retention and learning outcomes in engineering education*. Michigan State University. Higher, Adult, and Lifelong Education.

Nassaji, H. (2015). *Qualitative and descriptive research: Data type versus data analysis*. Sage Publications Sage UK: London, England.

National Academies of Sciences. (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press.

Pace, D., Witucki, L., & Blumreich, K. (2008). Benefiting female students in science, math, and engineering: The nuts and bolts of establishing a WISE (Women in Science and Engineering) learning community. *NASPA Journal*, *45*(3), 373–383.

Pike, G., Kuh, G., & McCormick, A. (2008). Learning community participation and educational outcomes: Direct, indirect and contingent relationships. *Association for the Study of Higher Education Conference in Jacksonville, FL*.

Priest, K. L. (2012). *Constructing Leadership Identities through Participation in a Leadership Living-Learning Community* [PhD Thesis]. Virginia Tech.

Priest, K. L., Saucier, D. A., & Eiselein, G. (2016). Exploring Students' Experiences in First-Year Learning Communities From a Situated Learning Perspective. *International Journal of Teaching and Learning in Higher Education*, *28*(3), 361–371.

Rehak, A. M. (2018). *Becoming and Being an Inspire Student: Learning, Participation, and Engagement in a Living-Learning Community* [PhD Thesis]. Indiana University.

Rocconi, L. M. (2011). The impact of learning communities on first year students' growth and development in college. *Research in Higher Education*, *52*(2), 178–193.

Rodrigues, R. L., Ramos, J. L. C., Silva, J. C. S., & Gomes, A. S. (2016). Discovery engagement patterns MOOCs through cluster analysis. *IEEE Latin America Transactions*, *14*(9), 4129–4135. https://doi.org/10.1109/TLA.2016.7785943

Roth, W.-M., & McGinn, M. K. (1997). Graphing: Cognitive ability or practice? *Science Education*, *81*(1), 91–106. https://doi.org/10.1002/(SICI)1098-237X(199701)81:1<91::AID-SCE5>3.0.CO;2-X

Russell, L. (2017). Can learning communities boost success of women and minorities in STEM? Evidence from the Massachusetts Institute of Technology. *Economics of Education Review*, *61*, 98–111.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Seymour, E. (2002). Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology. *Science Education*, *86*(1), 79–105.

Shaffer, D. W., & Resnick, M. (1999). " Thick" authenticity: New media and authentic learning. *Journal of Interactive Learning Research*, *10*(2), 195–216.

Smith, B. L. (2001). The challenge of learning communities as a growing national movement. *Peer Review 4*(1).

Smith, S. U., Hayes, S., & Shea, P. (2017). A critical review of the use of Wenger's community of practice (CoP) theoretical framework in online and blended learning research, 2000-2014. *Online Learning*, *21*(1), 209–237.

Spanierman, L. B., Soble, J. R., Mayfield, J. B., Neville, H. A., Aber, M., Khuri, L., & De La Rosa, B. (2013). Living learning communities and students' sense of community and belonging. *Journal of Student Affairs Research and Practice*, *50*(3), 308–325.

Stassen, M. L. (2003). Student outcomes: The impact of varying living-learning community models. *Research in Higher Education*, *44*(5), 581–613.

Stebleton, M. J., & Jehangir, R. (2016). Creating communities of engaged learners: An analysis of a first-year inquiry seminar. *Learning Communities Research and Practice*, *2*(5), 1–13.

Surakka, S., & Malmi, L. (2005). Need assessment of computer science and engineering graduates. *Computer Science Education*, *15*(2), 103–121.

Swan, K. (2002). *Building Learning Communities in Online Courses: The importance of interaction*. https://doi.org/10.1080/1463631022000005016

Taylor, O. L., McGowan, J., & Alston, S. T. (2008). The effect of learning communities on achievement in STEM fields for African Americans across four campuses. *The Journal of Negro Education*, 190–202.

Tinnell, T. L., Ralston, P. A., Tretter, T. R., & Mills, M. E. (2019). Sustaining pedagogical change via faculty learning community. *International Journal of STEM Education*, *6*(1), 1–16.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition.* ERIC.

Ujj, C. (2020). *Impact of Living-Learning Communities on Second-Year Students* [PhD Thesis]. University of Georgia.

Virtue, E. E., Maddox, G., & Pfaff, K. (2019). The lasting effects of learning communities. *Learning Communities Research and Practice*, *7*(2), 6.

Vygotsky, L. S., Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.

Weidman, J. C., DeAngelo, L., & Bethea, K. A. (2014). Understanding student identity from a socialization perspective. *New Directions for Higher Education*, *2014*(166), 43–51.

Wenger, E. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business School Press.

Wenger, E. (2004). Knowledge management as a doughnut: Shaping your knowledge strategy through communities of practice. *Ivey Business Journal*, *68*(3), n-a.

Wenger, E. (2011). *Communities of practice: A brief introduction*.

Wenger, E., Trayner, B., & De Laat, M. (2011). Promoting and assessing value creation in communities and networks: A conceptual framework. *The Netherlands: Ruud de Moor Centrum*, *20*, 2010–2011.

Wenger, E., & Wenger, B. (2015). *Communities of practice: A brief introduction*.

Yuan, C., & Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. *J—Multidisciplinary Scientific Journal*, *2*(2), 226–235.

Zhao, C.-M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education*, *45*(2), 115–138.

Appendix A
Codes and definitions for the Domain (Benefits)

Table 10. Themes of benefit for the Domain

Benefits	Definition	Moderate	Good	Exceptional
Good Corporate Partner	Refers to any instance where students mentioned some benefits/positive statement related to corporate partner	10%	22%	0%
Data Science Skills	Students talks about acquisition or developing technical skills, computational skills such as R, Python, etc.	25%	43%	50%
Real Application	Students mention their application of knowledge on real world project or real-world problem solving	5%	30%	15%
Supportive Staff/Faculty	Students mention helping/supportive/caring/considerate Staff/Faculty at the learning Community	5%	9%	5%
Collaborative Knowledge Construction	Students mention how with peer/mentors/people of the same interest group they worked together to solve real world problems	10%	30%	40%
Metacognitive Skills	Students demonstrate a good understand of their thought process or discuss their learning from mistakes	5%	9%	5%
Great Learning Experience	Students mention learning experience at the learning community as great, exceptional, good, interesting all these instances are classified as 'great learning experience'	10%	88%	90%
Fair Learning Experience	Students mention learning experience at the learning community as fair, average, nice or decent classified as fair learning experience'	90%	12%	10%
Professional Skills	Students mention any instances of developing skills such as project management, time management, leadership, helping in career transition	5%	22%	20%
Unique Project	Students mention the project uniqueness or call the project unique or how the work they did was very different from all other projects	0%	0%	15%
Socialization	Students demonstrate a sense of belonging with context, peers, faculty as a part of the Data Science group	10%	0%	25%

Appendix B
Codes and definitions for the Domain (Challenges)

Table 11. Themes of Challenge for the Domain

Challenges	Definition	Moderate	Good	Exceptional
Gained No New Knowledge	Students mention that participating in project or learning community did not help them to acquire any new knowledge	10%	0%	0%
Student Lacks Prior Knowledge/experience	Students mention that they lack prior knowledge/experience of coding, technical knowledge, computational skills	20%	0%	0%
Personal Issues	Students mention that due to personal issues such as busy schedule, food, ability to cope up affected the student's experience in the learning community	35%	4%	0%
Busy Work	The students mention 'busy work' or aspects of busywork	10%	0%	0%
Lack of Structure	The student mentions the lack of organization, structure, directions, or clarity	15%	9%	0%
Lack of interaction	The student mentions a lack of communication, meetings, or interactions	5%	0%	0%
Delay in Getting Data	Students mention that data was received late, or project outcome was affected due to delay in getting data	5%	0%	0%
Inadequate of learning Material/tools	The student mentions that learning materials lacked solved examples, or provided less guidance	25%	0%	5%
Unsteady Start	Student mentions start of the project/semester/corporate partner experience as slow, rough, unstable, rocky, late, confusing has been classified as unsteady start	0%	26%	10%
Lack of Guidance Initially	The student mentions that the guidance was less in the first semester or initially or at the start of the project	0%	0%	5%

Appendix C
Codes and definitions for the community (Mentor/Mentee Interaction_Benefits)

Table 12. Themes of benefit for Community (Mentor Characteristics)

Benefits	Definition	Moderate	Good	Exceptional
Supportive	Students mentioned mentor as supportive, helping, caring	70%	65%	85%
Committed	Students mentioned mentor as passionate, engaged with students, committed are classified as 'committed'	20%	22%	30%
Good Leader	Students mention a mentor as a good guide, leader, provided proper direction, led meetings, etc. are classified as a good leader.	15%	57%	25%
Knowledgeable	Students mentioned mentor possess good disciplinary knowledge, good technical, computational knowledge	20%	30%	45%
Candid	Students mentioned mentor as straight forward or frank in providing feedback	5%	4%	0%
Role Model	The student mentioned mentor as an inspirational person	5%	0%	0%
Scaffolded Learners	Instructors allowed students to explore, come up with ideas, and guided them to translate ideas into solutions.	0%	9%	15%
Active	Students mentioned mentor as proactive, actively participating in the project and discussions	0%	9%	15%

Appendix D
Codes and definitions for the community (Mentor/Mentee Interaction Challenges)

Table 13. Themes of challenges for mentor characteristics.

Challenges	Definition	Moderate	Good	Exceptional
Only One Mentor Effective	Students mentioned one mentor as effective and the other as ineffective	5%	4%	0%
Lack of guidance	Students mentioned there was a lack of guidance on the part of the mentor	5%	0%	0%
Lack of Clarity	Students mentioned a lack of clarity in providing direction or instruction on the part of mentors	5%	4%	0%
Lack of technical knowledge	Students mentioned a lack of real-world knowledge on the part of the mentor	5%	4%	0%
Lack of communication	Students mentioned that mentors lacked communication skills	20%	0%	0%
Lack of teaching experience	Students mentioned that mentors lacked teaching experience	0%	4%	0%

Appendix E
Codes and definitions for the Practice (Benefits)

Table 14. Themes of benefit from the Practice

Benefits	Definition	Moderate	Good	Exceptional
Good Mentors	Students mention mentors are good, helpful, supportive	5%	0%	5%
Valuable Experience	Students describe their experience with the project as great, exceptional, valuable, interesting rewarding	60%	48%	75%
Data Science Skill	Students describe about acquisition or developing technical skills, computational skills such as R, Python, etc. while working on Project	10%	9%	40%
Real Application	Students mention their application of knowledge on real world project or real-world problem-solving concerning Project	15%	35%	40%
Teamwork	Students describe their working as teams and talk about team cohesion or effectiveness	5%	4%	20%
Acquired New Knowledge	Students mention the acquisition of any kind of disciplinary or computational knowledge while working on the project	5%	22%	25%
Data Science Identity	Students demonstrate the development of disciplinary identity while working on the project	0%	4%	0%
Interdisciplinary	Students describes the project experiences integration/application of various fields of study	0%	4%	5%
Prior Knowledge	Students describes the use of prior knowledge while solving project problems	0%	4%	5%
Unique Project	Students describes the project as unique	0%	9%	15%
Motivated	Students demonstrates the drive for the project	15%	17%	20%
Professional skills	Student describes acquisition of skills such as interpersonal skills, time management, leadership, project management while working on the project	10%	4%	5%

Appendix F
Codes and definitions for the Practice (Challenges)

Table 15. Themes of challenge from the Practice

Challenges	Definition	Moderate	Good	Exceptional
Unsteady Start	Student mentions start of the project as unsteady, slow, rough, unstable, rocky, late, confusing has been classified as unsteady start	5%	9%	0%
Project lack real application	Student mentions that project lacked real application	10%	4%	0%
Lack of technical skills (Students)	Students describe the lack of technical/coding/computing skills required for the project	25%	5%	0%
Lack of structure	Student mentions about the lack of organization, structure, directions or clarity with respect to the project	15%	17%	0%
Gained no new skills	Students mention that participating in project did not help them to acquire any new skill	5%	0%	0%
Data Issues	Student mentions any form of issues with data such as delay in getting data, lack of real data with regards to project	5%	22%	5%
Lack of team cohesion	Students reports issues in teamwork, team effectiveness or team cohesion	0%	4%	0%
Lack of clarity/direction	Students mentions any lack of clarity or direction with regards to Project	35%	9%	0%