

## Article

# Characterizing the Identity Formation and Sense of Belonging of the Students Enrolled in a Data Science Learning Community

Aparajita Jaiswal <sup>1,\*</sup>, Alejandra J. Magana <sup>1</sup>  and Mark D. Ward <sup>2</sup><sup>1</sup> Computer and Information Technology, Purdue University, West Lafayette, IN 47906, USA<sup>2</sup> Department of Statistics and The Data Mine, Purdue University, West Lafayette, IN 47906, USA

\* Correspondence: jaiswal2@purdue.edu

**Abstract:** Student attrition is a challenge experienced by higher education institutions. One of the key reasons for student attrition is the inability of students to develop an identity and a sense of belonging. This study aims to understand the role of a data science learning community in helping students to develop identity and a sense of belonging. The study used a mixed-methods approach to collect and analyze the data. The study used a pre-post survey to understand students' perceptions regarding identity formation and sense of belonging. The results of the survey study indicated that the perception of identity formation and sense of belonging increased as students progressed through the semester. To further investigate the reasons for the increase in the perception of identity formation and sense of belonging, in-depth semi-structured interviews were conducted. The results of the interview revealed that living in the data science learning community, participating in various events and interacting with faculty, staff and working with peers of similar interest helped students to develop a data science identity and leader identity, as well as experience sense of belonging.

**Keywords:** learning community; data science education; communities of practice; sense of belonging; identity formation; undergraduate education



**Citation:** Jaiswal, A.; Magana, A.J.; Ward, M.D. Characterizing the Identity Formation and Sense of Belonging of the Students Enrolled in a Data Science Learning Community. *Educ. Sci.* **2022**, *12*, 731. <https://doi.org/10.3390/educsci12100731>

Academic Editor: James Albright

Received: 23 September 2022

Accepted: 17 October 2022

Published: 21 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Student attrition among undergraduate students is a challenge the United States is trying to address. The retention of undergraduate students has become a complicated problem for the nation [1,2]. Only 41% of students graduate from college in four years, and 59% in six years [3]. Furthermore, the first year of undergraduate studies is the most challenging phase for all new undergraduates [4]. Studies [2,5–7] have revealed that the first-year college experience is critical in determining student retention. Transitioning from high school experiences to undergraduate life requires students to develop a self- and social identity. Students who fail to develop a self- or social identity experience identity conflict, creating a void and disengagement that results in college student attrition [4]. The feeling of isolation and disconnection that students experience during their undergraduate education jeopardizes student success and retention [8,9]. Students who experience a high level of disconnection socially and academically demonstrate a high probability of dropping out of college [10]. Psychosocial factors such as lack of socialization, belongingness and social support majorly contribute to student attrition. Studies by Whannell and Whannell [4] and O'Keeffe [7] revealed that students demonstrating a strong association with the institution have a lower probability of dropping out. In addition, there is a relationship between student retention and the feeling of disconnection [11]. Specifically, students who demonstrate a low level of connectedness or a high level of alienation have a higher probability of dropping out of the college [11]. Thus, higher education institutions must focus on refining the quality of the college experience for undergraduate students, specifically at the introductory level.

Prior studies have identified two factors that lead to student attrition: (a) the inability of students to develop identity [4], and (b) failure to experience a sense of belonging [5,8].

Identity refers to the perception of individuals or groups about “who they are, of what sort of people they are, and how they relate to others” [12]. Sense of belonging refers to students’ “psychological sense of identification and affiliation with the campus community” [13]. One crucial point is that identity and sense of belonging are inseparable constructs. A recent study by Hazari and colleagues [14] revealed that identity formation in recognition, interest, performance and competence are incomplete if students do not experience a sense of belonging. Since identities develop in a social setting and not in isolation, elements of a social setting such as context and peers are critical for identity formation. For instance, a study by Weidman and colleagues [15] aimed to characterize students’ identity formation through socialization. The study used a model of undergraduate socialization to understand the association between identity formation among college students and the impact of socialization in developing student identities. Weidman’s model sketches the identity of students in terms of the three forms of social interactions: (1) interpersonal social interaction that occurs with a peer and faculty, (2) intrapersonal relations that are related to learning such as attending classes or taking courses and (3) integration, referring to getting adjusted to the college life and campus activities [15].

Separating the aspects of social interaction in the form of socialization and sense of belonging from identity construction is difficult. Therefore, it is essential to study both constructs jointly to retain and attract students to higher education institutions. The socialization or sense of belonging allows the student to develop an association with peers, faculty and with the institution [16]. It also allows the student to accommodate the transition from high school to undergraduate life [14]. Factors such as the smooth transition of the student [4] and the development of a sense of belonging [8] reinforce the development of identity among the students [14,15]. In their study, Hoffman and colleagues [5] suggested that to inculcate a sense of belonging or a sense of affiliation among first-year undergraduate students, higher education institutions must develop first-year seminar courses or establish learning communities.

Studies have identified that learning communities can effectively promote engagement [17] and help students develop psychosocial skills such as identity and a sense of belonging [18]. However, if we specifically focus on STEM learning communities, it has been found that “psychosocial factors have been under-researched in STEM education in favor of more cognitive and behavioral factors” [19]. Based on our literature synthesis, we know that helping students develop psychosocial skills can improve student retention. Therefore, through this study, we aim to focus on characterizing two important psychosocial factors, identity formation and sense of belonging, among students enrolled in the data science learning community.

For this, the study pursues the following research questions:

**RQ 1 (Quant):** What are students’ perceptions regarding their identity formation and sense of belonging enrolled in a data science learning community?

**RQ 2 (Qual):** How do students describe their levels of a sense of belonging within a data science learning community?

**RQ 3 (Qual):** How do students’ participation and interaction in the data science learning community help them form their identity?

## 2. Theoretical Framework

Lave and Wenger [20] proposed the communities-of-practice framework based on their study conducted in Liberia, West Africa. According to Wenger and Wenger [21], a community of practice is not a group of friends or neighbors; rather, it is a group of individuals of similar interests who engage and work collaboratively to learn new skills in a specific context. Communities of practice are defined as “groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly” [21]. There are three important tenets of the communities-of-practice framework; they are *domain*, *community* and *practice* [20]. The *domain* refers to the area of the activity of interest shared by the community members. It represents the common interest

that brings the members of the community together [22]. The *community* characteristic of the communities-of-practice framework refers to the interaction between the community members. The members of the community share their interests and thoughts, collaborate and learn together as a group under the guidance of a mentor or guide. Wenger and Wenger [21] also stressed that the community needs the participants to interact. The *practice* tenet of the communities of practice refers to the shared repertoire of ideas, experiences, stories, tools and methods that community members create as they interact with each other [21]. The community members use these created artifacts to solve problems or challenges. Participation is integral to communities of practice. The active participation on the part of members in the context they are situated in inculcates a sense of community, allows the creation of learning artifacts through the act of practice and leads to identity formation and negotiations on finding the true meaning of experiences [20,23–25]. The negotiations and renegotiations on the meaning of learning experiences also help the identities of the member evolve as they progress towards becoming an expert learner through legitimate peripheral participation within the communities of practice [20,25].

### 3. The Data Science Learning Community

The learning community under investigation is The Data Mine (TDM). TDM is a residential learning community that allows students to live and learn under the same roof. Students of all disciplines can join the learning community and develop data science skills. Students get the opportunity to participate in a seminar course that lays the foundation of data science. Students also participate in other events, such as outside events and corporate projects. Outside events include talks delivered by experts from industry and academia on various topics such as data science as a career, research in data science, application of data science in industry, etc. The corporate partner projects are associated with the corporate partner (CRP) cohort of the learning community. More than 80 percent of the students belong to this cohort. The CRP cohort allows the students to work on real-world projects with a real company. Students work in teams; each team is headed by a Corporate Partner Teaching Assistant (CRP TA). The CRP Teaching Assistants are senior students in TDM; they act as a bridge between the corporate partner company and the student team. TDM also has data scientists to assist CRP teams if they get stuck on any problems. In addition, there are corporate partner mentors who are full-time employees of the company; they help provide student teams with the required resources and guidance on a project. Other than the CRP cohort, TDM also consists of three additional cohorts; they are the general cohort, research with a professor cohort and the academic learning community cohort. The general cohort comprises students who register for the one credit-hour TDM seminar course and participate in at least three outside events in the fall and spring semesters.

The research with a professor cohort of TDM allows students interested in research to work on interdisciplinary data-science projects. Currently, The Data Mine has four research with professor opportunities: Computational Investigation of Living Systems, Human Development and Family Studies; Public Health; Physics; and Vertically Integrated Projects. All four research projects have their own Teaching Assistants (TAs) who work with students and answer their questions. The last cohort of TDM is known as the academic learning community. This cohort comprises 14 academic cohorts, such as Actuarial Science, Agriculture, Analyzing Digital Gaming and Culture, Biology, Data in Health and Human Sciences, Data Visualization, etc. Students from any college can join an academic cohort and take their courses. Faculties hold office hours in the Hillenbrand Hall building. Some faculty members are Faculty Fellows who work with Resident Assistants to plan social events for the students. Students in these cohorts also get the opportunity to work on group projects and learn interdisciplinary applications of data science. In summary, TDM is a unique living–learning community that allows students to work in a collaborative environment and promotes socio-constructivist learning.

#### 4. Methods

This study uses a mixed-methods research design. A mixed-methods design has been considered an excellent approach for combining qualitative and quantitative findings to answer research questions. Mixed-methods design is a holistic approach that collects, analyzes and interprets the study's findings using qualitative and quantitative approaches [26]. Combining the two methods allows the researcher to explore the research question in depth. For this study, the sequential explanatory mixed-methods research design was used [27]. The study started with an initial quantitative survey followed by an in-depth qualitative study. The results from the quantitative study informed the design of the qualitative study. Specifically, results from the quantitative study guided us in designing the interview protocol for the qualitative study. Our sequential explanatory mixed-methods design prioritizes the qualitative phase of the research design. Therefore, this design was found appropriate for this study.

Prior studies by Jaiswal et al. [28] have used a sequential explanatory mixed-methods approach to analyze student reflections. The objective was to understand the reflection scores assigned to each cluster of students. Conducting the qualitative inquiry for the students' reflection helped to identify the emergent themes. Table 1 has been adapted from Ivankova et al. [27] below and represents the phases, sub-phases, procedures and products that would emerge at every step of the study.

**Table 1.** Illustration of Explanatory Mixed-Methods Approach.

Phase	Sub-Phase	Procedure	Product
PHASE 1 Quantitative Phase	Quantitative Data collection (RQ1)	A five-point Likert scale survey was used to collect data on students' perceived identity and sense of belonging.	Pre-test and post-test numeric data
	Quantitative data analysis	Data were analyzed using descriptive and inferential statistics	Measures of central tendency and spread of the pre-test and post-test data (mean and standard deviation) and statistics for significant differences ( <i>p</i> -values and test statistics) and visualizations.
Intermediate Phase	Connecting Qualitative and Quantitative phases	Survey results were used to create the interview protocol and purposefully select participants from the population for conducting semi-structured interviews	An interview protocol and participant selection for the interview
	Qualitative Data Collection Method (RQ2, RQ3)	Semi-structured interviews with students	Text Data (interview transcripts)
PHASE 2 Qualitative Phase	Qualitative data analysis	(i) To answer RQ2 and RQ3, thematic analysis was conducted to find emerging patterns in students' experiences	Codes and categories are organized into themes, along with representative quotes describing students' experiences.

#### 4.1. Context and Participants

The study was conducted in The Data Mine (TDM). Over six hundred undergraduate students join TDM each year to learn and develop data science skills. Therefore, as a part of this study, we intended to understand how participating in TDM helps students develop identity and experience a sense of belonging. A mixed-methods study was conducted, and participation in this study was voluntary. For the quantitative study, 25 undergraduate students who were in their first year participated in both pre-and post-survey. The pre-survey was conducted in September of 2020 and post-survey was conducted in December 2020.

Following this, the qualitative data was collected in the fall of 2021. For the qualitative study, 19 students were recruited, and the criteria for recruitment were students must have completed at least one year in TDM. Semi-structured interviews were conducted for all 19 students. It is also important to note that not all the students who were recruited for the interview participated in the survey study, but the students interviewed shared the same demographics as the students who participated in the survey study. In addition, prior to conducting the study, permission from the Institutional Review Board was obtained. In order to maintain the privacy of participants and the confidentiality of responses, all students were assigned pseudonyms, and any identifiers were removed. The data was accessible only to the members of the research team, and students' consent was obtained prior to conducting the study.

#### 4.2. Data Collection Method

The data were collected in two phases. The quantitative data was collected at the start and end of the fall 2020 semester. The quantitative data was collected in the form of a survey. The survey instrument contained questions related to identity and sense of belonging. The identity questions were derived from the questionnaire developed by Godwin [29], as the instrument was validated and also demonstrated a high reliability coefficient (Cronbach alpha greater than 0.80). The questions were adapted for the current study to understand TDM undergraduate students' perception of identity formation. In her study, Godwin [24] described three sub-constructs of identity formation: recognition, interest, and performance/competence. For this study, recognition refers to the extent to which an individual perceives oneself or how others (peers, parents, faculty) perceive the individual in the context of data science knowledge, skills and abilities. Interest refers to the extent to which an individual enjoys learning data science concepts. The performance/competence construct is described as the extent to which the individual is confident in applying data science techniques and tools [29]. Questions for sense of belonging were adopted from Chin et al. [30]. The overall reliability coefficient for the sense of belonging questions found in Chin et al. [30] was Cronbach alpha = 0.95. The sense of belonging construct consisted of three questions. The purpose of those questions was to understand the extent to which students feel accepted in TDM. Table 2 represents the sample survey questions used for understanding the perceptions of identity formation and sense of belonging; the complete survey is available in Appendix A of this paper.

**Table 2.** Sample Survey Questions.

Construct	Questions
Recognition	I see myself as a data science person
Interest	I enjoy learning data science concepts
Performance/Competence	I am confident that I can understand the data science concepts in class
Sense of Belonging	I feel a sense of belonging to my data science community

The qualitative data were collected in the fall of 2021. Students who had completed at least one year in TDM were recruited for the study. Semi-structured interviews were conducted that lasted between 60 to 90 min. Semi-structured interviews allow the researcher to ask open-ended one-to-one questions to the participants in order to understand the



phenomenon under study [31]. The interview protocol has undergone multiple iterations based on the feedback received by experts in the field of qualitative inquiry. The interviews were conducted using Microsoft Teams, and each interview session was recorded for conducting the analysis. Before the start of the interview, students provided their consent. The conversation of the interview focused on two criteria: (1) to understand how students perceive themselves after spending a substantial amount of time in TDM and what the benefits and challenges of participating in TDM were, and (2) opportunities provided to students in TDM in terms of mastering technical and non-technical skills, socializing, and growing as an individual. Post conducting the interviews, the interviews were transcribed and analyzed.

#### 4.3. Data Analysis Method

The quantitative data were analyzed using descriptive and inferential statistics. Descriptive statistics were used to describe the measures of central tendency for each construct. The *t*-test was used to test if there was a significant increase in students' perception from pre- to post-test. The quantitative data were analyzed using the R programming language. For analyzing the qualitative interview data, NVivo 12 software was used. We conducted an inductive thematic analysis, as the objective of the thematic analysis was to identify the patterns that demonstrated commonality or differences among the participants' experiences in TDM. The steps described by Braun and Clark [32] were followed for conducting the thematic analysis: (1) getting yourself familiar with the data, (2) generating the initial codes, (3) searching for the prospective themes, (4) reviewing the generated themes, (5) defining the themes and (6) writing the final report.

#### 4.4. Trustworthiness Considerations

Ensuring the trustworthiness of the qualitative data analysis is crucial. To ensure the reliability of the qualitative data, member-checking and peer-debriefing were conducted. For the member-checking, the narrative with students' quotes was sent to participants to confirm if the narrative resonates with their described experiences. Peer-debriefing was conducted with the help of a peer [33–35]. For the purpose of peer-debriefing, one of the authors met a peer with a robust research-method background and expertise in education research methods. The peer was not part of The Data Mine project. Regular meetings and extensive discussions were conducted around the research questions and research methods. For the data analysis, peers and one of the authors independently coded two interviews and discussed the results. The code book was revised based on the discussion and the coding was performed again based on the feedback. The final themes that emerged were discussed, and similar themes were merged.

### 5. Results

The results section consists of three main sections. The first section contains the results of the quantitative data analysis. The other two sections describe the results of the qualitative data analysis. The qualitative data analysis section contains the main themes followed by sub-themes for sense of belonging and identity formation.

#### 5.1. Quantitative Results

##### Overall Perceptions of Identity Formation and Sense of Belonging

The results of the descriptive analysis revealed that students demonstrated an increase in the perception of all three constructs of identity formation: recognition, interest, and performance/competence. The results from Table 3 for the paired *t*-test revealed that there was a statistically significant increase in the perception of recognition from pre-test ( $M = 3.73$ ,  $SD = 0.60$ ) to post-test ( $M = 3.98$ ,  $SD = 0.67$ ,  $t = 2.08$ ,  $p = 0.04$ ). It can be also observed from the post-test scores that students demonstrated a high degree of interest in data science, and felt competent while solving data science problems.

**Table 3.** Perceptions of Identity Formation.

	September N = 25		December N = 25		Gains	<i>t</i>	<i>p</i>
	Mean	SD	Mean	SD			
Recognition	3.73	0.6	3.98	0.67	0.25	2.08	0.04
Interest	4.48	0.5	4.56	0.59	0.08	1.11	0.27
Performance/Competence	4.02	0.64	4.13	0.83	0.11	0.85	0.39

From Table 4, we can observe that students demonstrated an increase in the perception of a sense of belonging from pre- to post-test. Although the results of the paired *t*-test revealed that the increase was not significant, it is still important to note that the mean for the post-test for a sense of belonging was high, i.e., 4.20 on a 5-point rating scale (1 being the lowest and 5 being the highest), meaning students felt accepted in TDM.

**Table 4.** Perceptions of Sense of Belonging.

	September N = 25		December N = 25		Gains	<i>t</i>	<i>p</i>
	Mean	SD	Mean	SD			
Sense of Belonging	4.13	0.65	4.20	0.71	0.07	0.75	0.46

## 5.2. Qualitative Results

### 5.2.1. Themes for Sense of Belonging

**Theme 1.** *Interacting with diverse peer groups, working on a corporate project and meeting like-minded people allowed the students to collaborate, socialize and learn from each other.*

**Sub-Theme (1a): Worked as a group on problems and learned from one another:** The students mentioned their interaction with peers while working on a problem and how that led to learning from one another. Marion mentions that he has learned a lot by interacting with his peers at The Data Mine.

*Uhm, I mean it. It's so huge. I mean, a lot of the learnings that I think I've had come from discussions that I have with my friends with the people around me with other Data Mine students. Also, I was learning from really, really, really bright students around me.*

**Sub-Theme (1b): Obtained multiple perspectives to solve the same problem through interaction:** The students described how working on multiple projects helped them obtain different perspectives. For example, Morgan called it fun to work with peers on seminar projects. She found it exciting to see how her peers approached the problem, what functions they used and how they solved the problem.

*I think it's a ton of fun. It's always fun to collaborate with your peers on project and just to see how they would approach it versus how I would approach it. I think it's a good time. Definitely, it's very interesting to see if there are different functions that they would use, or maybe just different methods or ways of approaching it. I think you learn a lot by working with other people.*

**Sub-Theme (1c): Worked in interdisciplinary teams for corporate projects and learned division of labor:** Students described how they worked in an interdisciplinary team for corporate projects. The teams comprised students from different majors or different levels of education, such as freshmen, sophomores, juniors or seniors. Orion described his experience of working with an interdisciplinary team.

*Our team had three or four juniors and two seniors and two grad students. Many of them were freshmen or sophomores. So, since this was such a big task, there were freshmen and sophomores, they were completely new to programming. For the project we need students having experience in creating mobile applications, creating APIs, create dashboard and they don't have any experience with that. So luckily since there was varying experience,*

*we had to have the senior members responsible for more of the more technically challenging tasks and we provided the smaller tasks for the freshmen and sophomores and also gave them opportunity to learn from the senior member.*

**Sub-Themes (1d): Lived in Hillenbrand and met people with similar interests:** Students shared their experiences of living and learning in Hillenbrand Hall. Jamie lived in Hillenbrand Hall, and she described how she would just grab a meal, sit in the lobby, and meet other students.

*We all live in Hillenbrand, So I would just grab a meal, sit and then like kind of meet with some people if I saw them. You know, we can just work on the project and then someone would help me like walking around and then you know that was like the environment. Every Friday we'd get together in the sub lobbies of each floor. Uhm, we just dealing bounce off ideas. Uh, on how to work on the project or just even corporate partners. So that was really interesting and how we could just kind of, you know, knock on a next door and like you could find someone who would be working on the same problem as you.*

**Theme 2.** Students felt a sense of belonging and acceptance in The Data Mine as they established connections with faculty, staff, and peers.

**Sub-Theme (2a): I feel included:** The students mentioned that they feel accepted and included. In addition, students mentioned that staff, faculty or peers made them feel included and accepted in TDM. Robbie found TDM very inclusive because she found it fun to live in Hillenbrand Hall with other TDM students. She also met the faculty and staff regularly; since they had an open-door policy, students used to drop by their office to greet them and interact with them, not just for academic reasons but even on a personal basis.

*Data Mine is very inclusive. We all got together or same floor. You know, I met Dr. W on Monday, Tuesday. Sometimes he'd be in the office at 10:00 PM. It is very easy to socialize with anyone in Data Mine. Just on the way to class I use to drop by Ms. E's office and her door was always open. I just talked to her about stuff not always you know related to Data Mine. It's pretty good.*

**Sub-Theme (2b): Faculty and Staff made me feel connected:** Students mentioned that they felt faculty and staff were helpful, supportive, friendly and made them feel connected to TDM. Morgan described all her interactions with faculty and staff at TDM as great. She found them to be very helpful and kind people. The most important aspect that Morgan mentioned was that the faculty and staff created a welcoming environment, irrespective of their background and competencies.

*I think all my interactions were great with faculty and staff at Data Mine. They were very kind people, very supportive and understanding of what your background is and they work really hard to make sure that The Data Mine is inclusive of everyone, regardless of major where you come from, what your grade is, they just do a great job in making it a good learning community.*

**Sub-Theme (2c): Felt connected with peers and made friends:** The students mentioned that their peers in TDM made them feel connected, and the interaction helped them to make good friends. They described their bonding with one another and sharing of similar interests. Frankie mentioned that she felt connected with her peers, as they met regularly for the project, worked together, experienced similar challenges and solved them.

*I would say my peers especially made me feel connected. I think I kind of bonded the most with them, especially because we were doing the projects together and we were meeting, you know with each other two or three times a week. So, we saw each other the most. So, I feel like they kind of made me feel the most connected and you know, we were with each other in the same boat and everything*



### 5.2.2. Themes for Identity Formation

**Theme 1. Data science identity:** *Students were in the process of developing their data science identity. They felt competent in applying their data science skills to solve real-world problems and pursue a data-oriented career in the future.*

**Sub-Theme (1a): I was a novice learner before joining TDM:** Most students mentioned that they had very limited to no knowledge of data science, statistics or coding competencies before joining TDM. Marion mentions that, before joining TDM, he did not know much about data science concepts as he did not have any technical background in high school. He thought joining TDM would be an exciting opportunity and help him learn data science skills.

*My idea of what the field of data science might look like was pretty blank. Um, like I didn't really know exactly what to expect, and I guess like I knew that I was going to be coming in learning some data science stuff I didn't really know what that was, I was not a technical student in high school, I didn't really have much technical background, so before coming in as a freshman I kind of just took it in and I just thought that the Data Mine be an opportunity for me to learn some data science skills.*

**Sub-Theme (1b): I am a data science person:** The majority of students described that they are in the process of developing data science skills—and they have started to feel competent in using them—but they have yet to identify themselves as data scientists, as they feel that they need to develop many more skills and gain experience. Therefore, they feel comfortable calling themselves a data science person. Morgan mentioned that she has developed data science skills and has grown in terms of coding skills. Therefore, she is comfortable calling herself a data science person. Morgan also believes that she needs more real-world experiences such as internships, learning new languages and developing hard skills to call herself a data scientist.

*But in terms of data science, I would definitely not call myself a data scientist by any means. I think I have a lot to learn, a lot more experience to have, so I would just say I'm someone that's interested and learning about coding. So, I would call myself something a data science person. I wouldn't call myself a data scientist until I've gotten more of maybe internship experience or just learning more languages or more skills or just actual hard skills, I guess.*

**Sub-Theme (1c): My peers and faculty recognize my data science skills:** Most students believe that faculty or peers recognize their data science skills. Quinn believes that peers and faculty recognize his skills. Faculty recognize his skills because they grade student projects and know how each student performs. In addition, he mentioned that one of the faculty members offered him opportunities for internships and jobs. This showcases that they know his abilities and potential. Peers also recognized his skills because he led a team as a TA; he was also competent in answering his students' queries and teaching them when required.

*Uhm, I think so because, faculty recognize my skills. I think that they do because, uh, we they like grade our projects so clearly, they know like that. I have the skill that I have, and I also think that the faculty specifically has recognized it because I've they've offered me like internships and jobs and they understand the level at which I work and like what my aspirations are, so I would definitely say they recognize my skill. Also, for my peers I teach them and answer their queries so I would definitely say they too recognize my skills.*

**Sub-Theme (1d): I plan to pursue a data-oriented career:** The students mentioned that they want to pursue a career related to data science, data analytics, or any career that emphasizes the application of data. Some students also mentioned that they want to go to graduate school to study courses related to data science, and then pursue a future career in a data-oriented field. Rio mentioned that she wants to enter the field of bioinformatics, specifically focused on computational drug design. Before joining TDM, she also mentioned,

she was unaware of this field. Previously, she wanted to go into biotechnology and become a research scientist as she did not know about the computational aspect of drug discovery.

*Uh, so as of now, I'd really like to enter the field of bioinformatics. Uh, maybe with like a focus on Computational aided drug design. But that has evolved through The Data Mine like I didn't know that was a field until I started with The Data Mine. I would say before I joined The Data Mine I wanted to work in biotechnology as like a research scientist as I didn't have any idea about the computational background of the like drug discovery things and all these things.*

**Sub-Themes (1e): I developed data science skills and feel competent in applying them:**

The students described how participating in various events in TDM helped them to develop data science skills and broaden their knowledge of data science. Morgan describes that, when she joined TDM, she came in as a novice with very little background in coding, and no data science skills. TDM helped her to develop her skills exponentially.

*I say the Data Mine has helped me with my data science skills like exponentially. I came in with very very little coding background and approximately zero data science skills . . . through the corporate partners project I learned a lot about machine learning ANOVA, Random Forest, just different statistical models. I also developed data collection skills Uh, we were trying our best to figure out a way to convert those qualitative data sets to more quantitative datasets just to get some sort of analysis.*

The students mentioned that they felt competent in applying their data science skills to real-world data sets, to their course work or during corporate projects. Finley mentioned that he felt competent when he was able to use his knowledge of data science to explore and analyze real-world data sets.

*I started applying my data science skills to projects outside the class. Because class projects, everyone else do, every student has to do, and more students end up doing that. But I think the point where I realized that wow, I have some ability to be able to actually analyze data, when I started using Kaggle.*

*Yeah, OK, you know, so I would read and a lot of datasets from Kaggle. Or do a data visualization with and then run some basic statistical analysis. The models like linear regression, K-means nearest neighbors, grid search, CV, stuff like that. So yeah, I would say that the point the switch over from me being or just a student to someone who actually applies will data science skills. Or you know to real world datasets and stuff like that, and that was a point where I would say that I was proficient.*

**Theme 2. Leader identity:** The Corporate Partner Teaching Assistants (CRP TAs) agreed that they were developing a leader identity and a data science identity, as they enjoyed their role and felt competent in solving student queries and managing a team.

**Sub-Theme (2a): I am a leader:** The CRP TAs identified themselves as growing leaders. They describe how leadership opportunities at TDM helped them to develop leadership skills. Quinn also mentioned that being a leader was a great and novel experience for him as he managed a team, facilitated meetings and motivated his team members.

*So, it was pretty novel experience for me. Uhm, I learned that I really like leading people. I think getting to like facilitate meetings and uhm, manage people and help people speak up when they're not speaking up enough or indicating. And I learned a lot about myself and the way I like to lead and kind of what makes a good team, such as like having good communication outside of meetings and during meetings and kind of inspiring people to want to work on what you're working on.*

**Sub-Theme (2b): Manage student team:** The corporate partner TAs described how they developed team management skills. Riley mentioned that she loves to be a TA and has developed team management skills. Riley and her co-TA work together to manage the team. They plan lab times, conduct ice-breaking activities, and get assignments done during the lab session. She makes sure that the team meets their deadlines in a timely fashion.

*I love it [team management] so much more because we get to plan things like lab times and you know, we do also a little icebreaking activity typically, every time which is good. Uh, and yeah, planning stuff like that just behind the scenes is what I like. And also, I like to say that I have leader qualities, so I also learnt like getting small assignments done, making sure the team has the deadlines in their calendar and notes are being written, and stuff like that so.*

**Sub-Theme (2c): Liaison between the student team and corporate partner company:** The corporate partner TAs described how they acted as liaisons between TDM and corporate partners. Marion mentioned that his role as TA is not focused on performing the work but on ensuring the work is performed. Therefore, he is responsible for managing the team, dividing the work into sprints, ensuring smooth execution of each sprint and, finally, acting as a bridge between the TDM student team and the corporate partners.

*I'd say the main difference as a TA is where we are removed from the physical development of the application in some sense, so our role is now less of a developer, and we are more of a leader where we are expected to keep everything on track. And so, the way we do that is we work within the agile methodology. Uhm, I'm responsible for conducting Sprint events and making sure that the development team, which is the group of students are, on track with their user stories and their task ownership. And then I'm also kind of like that bridge between the corporate partner mentor and the students. Uhm, like that a role is meant to be that communicator between both. So, to be kind of like a resource for each of those groups.*

**Sub-Themes (2d): Promoted student engagement through team bonding activities:** The corporate partner TAs described the steps they take to improve student engagement in their teams. For instance, they conduct team bonding activities such as ice-breaking activities, going out for lunch or other social activities with their team members. Rio mentioned that she loves to do the ice-breaking activity to help her team members interact.

*So generally, for like the first like 10 minutes or so, I'll put up like a question on the screen or like a list of questions on so for instance. It's like it could be any flavor Jellybean which flavor would you be? What's your favorite Donut? What's your favorite song? Uhm, but then also try to do like a game for the first 10 to 15 minutes where they can interact with each other and like because it's on a Friday afternoon is when our labs are so everyone is kind of dragging by the end of the week. So doing ice-breakers kind of gets the energy flowing . . . We also go outside and work in teams, which helps us get to know one another better.*

## 6. Discussion and Implications

The study followed a mixed-methods design and was intended to answer three research questions. The first research question (RQ 1) was intended to investigate students' perceptions regarding their identity formation and sense of belonging. The pre-post survey was conducted to examine students' perceptions at the beginning and end of the semester. The results of the quantitative study demonstrated that students' perceptions of their identity and sense of belonging improved as they progressed through the semester. Similar results were obtained in the study by Henderson and colleagues [36]; they demonstrated that the role of mentoring and the opportunity to socialize allowed students to develop an engineering identity and sense of belonging after completing the program. The study by Verdin and colleagues [37] has also confirmed that identity formation directly affects the students' sense of belonging. Living in a learning community and being surrounded by peers, faculty, and staff of similar interests could be a possible reason for the growth in identity formation and sense of belonging. The study by Omityoin and colleagues [38] discussed the impact of a community of peers on identity formation. The study revealed that community played a vital role in helping students develop their identity as they interacted with like-minded individuals. It also helped them to collaborate and work together in groups.

The quantitative results of the current study confirmed that students were in the process of identity formation and were experiencing a sense of belonging. However, the factors that were contributing to identity formation or an increase in sense of belonging were unclear. Therefore, a qualitative study was conducted to answer research questions 2 and 3. The qualitative analysis results helped us identify the reasons that led to the growth in identity formation and sense of belonging. It was observed that students demonstrated an increased perception of their sense of belonging as they progressed in TDM. This reveals that students felt accepted, valued, and included in TDM. A sense of belonging is very closely connected to the interaction of students with their peers and faculty/staff/mentors in an institutional setting [39]. Studies by Osterman [40] and Schneiderman [41] have revealed that, if students have a good level of interaction with their peers and faculty, they demonstrate a high sense of belonging. A sense of belonging is also closely associated with identity formation, as students cannot form identity if they do not experience a sense of belonging [14]. The results show that TDM provided various opportunities for students to connect and interact with faculty/staff/mentors and peers.

Since identity is a socially constructed entity, interaction and collaboration play an important role in shaping individual identity. In TDM, students received multiple opportunities to interact and socialize with students, peers and mentors which helped them to recognize their potential, identify their interests and acknowledge their performance and competence in the area of data science. The theme of data science identity revealed that most students, at the start, had limited-to-no knowledge of data science disciplines. After joining TDM, participating in various events and collaborating with peers on problem-solving tasks helped all nineteen students develop competencies and grow in data science skills and knowledge. Since students were developing confidence and self-efficacy in approaching the data science problems, they planned to pursue data-oriented careers in the future. It is essential to note that identities are dynamic [42–44]; they keep changing as the individual progresses. Individuals can have multiple identities as they keep acquiring new skills and knowledge. Identity answers the questions: Who am I? How do others see me? How did I become what I am today? How do I see myself in the future? [45,46]. Responding to these questions is difficult, as multiple factors play an essential role in shaping an individual's perceptions of identity [45,46]. For example, the results of the qualitative analysis of our study revealed that students serving as TAs for corporate partner projects demonstrated a dual identity. The corporate partner TAs demonstrated a data science identity and an emerging leader identity. Focusing on the identity formation of the twelve corporate partners TAs, we found that they developed data science skills but also identified themselves as leaders, as they were constantly engaged in managing students and student teams. The emerging leader identity resulted from the following developmental factors: faculty/staff recognized leadership skills, peers/team members acknowledged leadership skills of TAs and student TAs found their role meaningful as they contributed to the growth and success of TDM. Therefore, we are required to focus on identity formation, especially in STEM, beyond the constructs of interest and performance, recognition and competence. Hence, we would like to discuss the study results from the perspective of an interactionist lens [43]. Kim and Sinatra [43] have also argued that we have approached a level where we need to interpret the STEM identity-formation process differently. The interactionist lens has been used in sociology to study identity formation [44]. This lens helps to account for the impact various social factors, such as the influence of environment/context, peers, engagement in activities, etc., have on the identity development of an individual [44].

Moreover, drawing on the theoretical framework of communities of practice, we find that participation is integral to the identity-formation process and experiencing a sense of belonging [20]. Participation allows the learner to interact and socialize, which helps them to master new skills and move from being a novice learner to expert members of the community. We saw a similar trend in TDM students starting as novice learners in TDM, and the activities and participation opportunities helped them to transform into expert learners and leaders. It also important to note that the three tenets of communities

of practice can be operationalized in the context of TDM, as TDM served as a *domain* by bringing people of similar interests together, creating a *community* that promoted interaction and collaboration and engaged students in *practice* by allowing them to work on various projects and assignments. The constant interaction with the peers, faculty and staff helped the students to learn new skills, recognize their potential as data science persons and team leaders and experience a sense of belonging with the people and with the context of TDM.

#### *Implications for Data Science Education*

The study results also have implications for data science education. The study suggests that undergraduate students must be engaged in data science education from their first year onwards [47], as the results of the study demonstrated that it takes time for students to develop and master data science skills. The study also demonstrated that applying data science knowledge to a real-world context helped the students develop self-efficacy. Therefore, data science educators must find venues that allow the students to apply their disciplinary knowledge to real-world problem-solving. Data science education must be imparted in an active project-based learning environment that allows students to work alongside their peers and learn from one another, as it helps to master complex concepts faster [48,49]. Lastly, the discipline of data science is multifaceted; therefore, just helping students to develop data science competencies is not enough. The data science educators must encourage students and provide them opportunities to develop research, communication, presentation and teamwork skills [50,51].

#### **7. Conclusions, Limitations and Future work**

The study's findings revealed that students in TDM developed a data science identity and experienced a sense of belonging. It was also noted from the results that the CRP TAs demonstrated a leader identity, as they were engaged in managing the corporate project teams and acted as liaisons between the TDM team and corporate partners. It is also important to note that, after spending more than one year in TDM, each student demonstrated growth in their data science skills, professional skills and communication skills as well as developed identities. TDM provided multiple opportunities to interact and collaborate that helped the students to develop identities and experience a sense of belonging.

The study addresses the research need of conducting studies to identify the psychosocial factors influencing student experiences in STEM learning communities. The study also contributes to the body of literature on data science education. Prior studies have revealed that the interdisciplinary nature of data science poses a challenge for educators to impart data science education. Therefore, higher education institutions must establish an intentional structure such as learning communities to impart data science education. In the case of TDM, the learning community served as an effective mechanism to bring people with different backgrounds but similar interests together and engage them in data science activities that help develop data science and psychosocial skills. The study also revealed that engaging students in an active learning project-based environment helped them to master complex data science concepts efficiently and learn an interdisciplinary application of data science. Lastly, the study also emphasizes that the higher education institution must foster a healthy social environment to help students develop multiple positive identities. The study has some limitations; the sample size for the quantitative study is small; therefore, a follow-up survey study with a larger sample size would be required to perform statistical analysis in the future. In addition, the majority of students in this study are from the corporate partner cohort; therefore, to understand the impact of other academic units in TDM on student experiences, future studies on student experiences from other academic units are required. For future work, we will focus on understanding how the skills developed at TDM are helping students to advance in their career. Furthermore, future work will be conducted to understand how TDM is helping underrepresented students grow into experienced members of the learning community.

**Author Contributions:** Conceptualization, A.J., A.J.M. and M.D.W.; methodology, A.J., A.J.M. and M.D.W.; software, A.J.; validation, A.J. and A.J.M.; formal analysis, A.J.; investigation, A.J., A.J.M. and M.D.W.; resources, M.D.W.; data curation, A.J.; writing—original draft preparation, A.J.; writing—review and editing, A.J., A.J.M. and M.D.W.; visualization, A.J.; supervision, A.J.M. and M.D.W.; project administration, A.J., A.J.M. and M.D.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported in part by the Lilly Endowment Charting the Future Phase I Planning Grant, through the Purdue Office of the Provost. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

**Institutional Review Board Statement:** The study was approved by the Institutional Review Board of Purdue University.

**Informed Consent Statement:** Informed consent was obtained from the participants prior to conduct of the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available to protect the privacy and confidentiality of the respondents.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Survey Questions for Identity Formation.

Construct	Questions
Recognition	My parents see me as a data science person
Recognition	Instructors see me as a data science person
Recognition	My peers see me as a data science person
Recognition	I see myself as a data science person
Recognition	Data science is good field for a person like me
Interest	I am interested in learning more about data science
Interest	I enjoy learning data science concepts
Interest	I find fulfillment in applying data science concepts
Performance/Competence	I am confident that I can understand the data science concepts in class
Performance/Competence	I am confident that I can understand the data science concepts used in assignments
Performance/Competence	I can do well on data science assignments
Performance/Competence	I understand concepts I have studied in the data science course
Performance/Competence	Others ask me for help with data science concepts or assignments

Indicate whether you: (1—Strongly Disagree, 2—Disagree, 3—Neutral, 4—Agree and 5—Strongly Agree).

**Table A2.** Survey Questions for Sense of Belonging.

Construct	Questions
Sense of Belonging	I feel a sense of belonging to my data science community (Data Mine Learning Community or the Data Science Certificate)
Sense of Belonging	I am a member of my data science community (Data Mine Learning Community or the Data Science Certificate)
Sense of Belonging	I see myself as part of my data science community (Data Mine Learning Community or the Data Science Certificate)

Indicate whether you: (1—Strongly Disagree, 2—Disagree, 3—Neutral, 4—Agree and 5—Strongly Agree).



## References

- Lyons, J.B. A Study of Undergraduate Mentoring and the Factors Affecting Students' Disengagement from Mentoring. Ph.D. Thesis, St. John Fisher University, Rochester, NY, USA, 2012.
- Tinto, V. Research and Practice of Student Retention: What Next? *J. Coll. Stud. Retent. Res. Theory Pract.* **2006**, *8*, 1–19. [CrossRef]
- Hess, A.J. Graduating in 4 Years or Less Helps Keep College Costs Down—But Just 41% of Students Do. Available online: <https://www.cnbc.com/2019/06/19/just-41percent-of-college-students-graduate-in-four-years.html> (accessed on 22 February 2022).
- Whannell, R.; Whannell, P. Identity theory as a theoretical framework to understand attrition for university students in transition. *Stud. Success* **2015**, *6*, 43–53. [CrossRef]
- Hoffman, M.; Richmond, J.; Morrow, J.; Salomone, K. Investigating “sense of belonging” in first-year college students. *J. Coll. Stud. Retent. Res. Theory Pract.* **2002**, *4*, 227–256. [CrossRef]
- Tinto, V. *Leaving College: Rethinking the Causes and Cures of Student Attrition*; University of Chicago Press: Chicago, IL, USA, 1993.
- Lanphier, T.S.; Carini, R.M. A Process Evaluation of a Learning Community Program: Implemented as Designed? *Educ. Sci.* **2022**, *12*, 60. [CrossRef]
- O’Keeffe, P. A Sense of Belonging: Improving Student Retention. *Coll. Stud. J.* **2013**, *47*, 605–613.
- Singh, A. Investigating the Sense of Belonging of International Students through a Predictive Model. Ph.D. Thesis, Iowa State University, Ames, IA, USA, 2018; p. 12331544.
- Williford, A.M.; Wadley, J.Y. *How Institutional Research Can Create and Synthesize Retention and Attrition Information. Professional File Number 108, Fall 2008*; Association for Institutional Research: 2008. Available online: <https://eric.ed.gov/?id=ED504412> (accessed on 5 October 2020).
- Liu, R. *Alienation and First-Year Student Retention. Professional File. Number 116, Spring*; Association for Institutional Research: 2010. Available online: <https://eric.ed.gov/?id=ED511456> (accessed on 5 October 2020).
- Abrams, D.; Hogg, M.A. *Social Identifications: A Social Psychology of Intergroup Relations and Group Processes*; Routledge: London, UK, 2006.
- Hausmann, L.R.; Ye, F.; Schofield, J.W.; Woods, R.L. Sense of belonging and persistence in White and African American first-year students. *Res. High. Educ.* **2009**, *50*, 649–669. [CrossRef]
- Hazari, Z.; Chari, D.; Potvin, G.; Brewe, E. The context dependence of physics identity: Examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates. *J. Res. Sci. Teach.* **2020**, *57*, 1583–1607. [CrossRef]
- Weidman, J.C.; DeAngelo, L.; Bethea, K.A. The Weidman model of undergraduate socialization: Continuity and change. In Proceedings of the Annual Meeting of the American Educational Research Association, Philadelphia, PA, USA, 3–7 April 2014.
- Morrow, J.; Ackermann, M. Intention to persist and retention of first-year students: The importance of motivation and sense of belonging. *Coll. Stud. J.* **2012**, *46*, 483–491.
- Fosnacht, K.; Graham, P. Is a HIP Always a HIP? The Case of Learning Communities. *J. Stud. Aff. Res. Pract.* **2022**, *59*, 59–72. [CrossRef]
- Magana, A.J.; Jaiswal, A.; Madamanchi, A.; Parker, L.C.; Gundlach, E.; Ward, M.D. Characterizing the psychosocial effects of participating in a year-long residential research-oriented learning community. *Curr. Psychol.* **2021**, 1–18. [CrossRef]
- Carrino, S.S.; Gerace, W.J. Why STEM Learning Communities Work: The Development of Psychosocial Learning Factors through Social Interaction. *Learn. Communities Res. Pract.* **2016**, *4*, 3.
- Lave, J.; Wenger, E. *Situated Learning: Legitimate Peripheral Participation*; Learning in Doing; Cambridge University Press: Cambridge, UK; New York, NY, USA, 1991; ISBN 978-0-521-41308-4.
- Wenger, E.; Wenger, B. Communities of Practice: A Brief Introduction. 2015. Available online: <http://wenger-trayner.com/wp-content/uploads/2015/04/07-Brief-introduction-to-communities-of-practice.pdf> (accessed on 5 October 2020).
- Wenger, E. Knowledge management as a doughnut: Shaping your knowledge strategy through communities of practice. *Ivey Bus. J.* **2004**, *68*, 1–8.
- Lindsay, L.N. *Transformation of Learners in a Community of Practice Occupational Therapy Fieldwork Environment*; University of Georgia: Athens, GA, USA, 2001.
- Jaiswal, A.; Lyon, J.A.; Perera, V.; Magana, A.J.; Gundlach, E.; Ward, M.D. Work in Progress: Evaluating Student Experiences in a Residential Learning Community: A Situated Learning Perspective. In Proceedings of the 2021 ASEE Virtual Annual Conference Content Access, Virtual, 26–29 July 2021.
- Jaiswal, A.; Lyon, J.A.; Magana, A.J.; Ellen, G.; Ward, M.D. Student Experiences within a Data Science Learning Community: A Communities of Practice Perspective. *Learn. Community Res. Pract.* **2021**, *9*, 2.
- Li, L.; Worch, E.; Zhou, Y.; Aguiton, R. How and why digital generation teachers use technology in the classroom: An explanatory sequential mixed methods study. *Int. J. Scholarsh. Teach. Learn.* **2015**, *9*, n2. [CrossRef]
- Ivankova, N.V.; Creswell, J.W.; Stick, S.L. Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods* **2006**, *18*, 3–20. [CrossRef]
- Jaiswal, A.; Lyon, J.A.; Zhang, Y.; Magana, A.J. Supporting student reflective practices through modelling-based learning assignments. *Eur. J. Eng. Educ.* **2021**, *46*, 987–1006. [CrossRef]
- Godwin, A. The development of a measure of engineering identity. In Proceedings of the ASEE Annual Conference & Exposition, New Orleans, LA, USA, 26 June–28 August 2016.

30. Chin, W.W.; Salisbury, W.D.; Pearson, A.W.; Stollak, M.J. Perceived cohesion in small groups: Adapting and testing the perceived cohesion scale in a small-group setting. *Small Group Res.* **1999**, *30*, 751–766. [CrossRef]
31. Mueller, R.A. Episodic narrative interview: Capturing stories of experience with a methods fusion. *Int. J. Qual. Methods* **2019**, *18*, 1609406919866044. [CrossRef]
32. Clarke, V.; Braun, V. Thematic Analysis. In *Encyclopedia of Critical Psychology*; Teo, T., Ed.; Springer: New York, NY, USA, 2014. [CrossRef]
33. Guba, E.G.; Lincoln, Y.S. *Fourth Generation Evaluation*; Sage Publications: Newbury Park, CA, USA, 1989.
34. Lincoln, Y.S.; Guba, E.G. *Naturalist Inquiry*; Sage: Beverly Hills, CA, USA, 1985.
35. Spall, S. Peer debriefing in qualitative research: Emerging operational models. *Qual. Inq.* **1998**, *4*, 280–292. [CrossRef]
36. Henderson, J.; Snodgrass Rangel, V.; Holly, J., Jr.; Greer, R.; Manuel, M. Enhancing Engineering Identity Among Boys of Color. *J. Pre-Coll. Eng. Educ. Res. J-PEER* **2021**, *11*, 2. [CrossRef]
37. Verdín, D.; Godwin, A.; Kirn, A.; Benson, L.; Potvin, G. Understanding How Engineering Identity and Belongingness Predict Grit for First-Generation College Students. 2018. Available online: <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1077&context=enegs> (accessed on 5 October 2020).
38. Omitoyin, J.A.; Revelo, R.A.; Bilgin, B.; Darabi, H.; Nazempour, R. Low-Income, High-Achieving Students and Their Engineering Identity Development After One Year of Engineering School. In Proceedings of the 2021 ASEE Virtual Annual Conference Content Access, Virtual, 26–29 July 2021.
39. Goodenew, C. Classroom belonging among early adolescent students. *J. Early Adolesc.* **1993**, *13*, 21–43. [CrossRef]
40. Osterman, K.F. Teacher practice and students' sense of belonging. In *International Research Handbook on Values Education and Student Wellbeing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 239–260.
41. Schneiderman, J. Improving Sense of Belonging among Freshmen at a Large, Urban High School through Service Learning. Ph.D. Thesis, The Johns Hopkins University, Baltimore, MD, USA, 2020; p. 185.
42. Godwin, A.; Potvin, G. Pushing and pulling Sara: A case study of the contrasting influences of high school and university experiences on engineering agency, identity, and participation. *J. Res. Sci. Teach.* **2017**, *54*, 439–462. [CrossRef]
43. Kim, A.Y.; Sinatra, G.M. Science identity development: An interactionist approach. *Int. J. STEM Educ.* **2018**, *5*, 51. [CrossRef]
44. Serpe, R.T.; Stryker, S. The symbolic interactionist perspective and identity theory. In *Handbook of Identity Theory and Research*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 225–248.
45. Avraamidou, L. Stories we live, identities we build: How are elementary teachers' science identities shaped by their lived experiences? *Cult. Stud. Sci. Educ.* **2019**, *14*, 33–59. [CrossRef]
46. Avraamidou, L. Science identity as a landscape of becoming: Rethinking recognition and emotions through an intersectionality lens. *Cult. Stud. Sci. Educ.* **2020**, *15*, 323–345. [CrossRef]
47. National Academies of Sciences. *Data Science for Undergraduates: Opportunities and Options*; National Academies Press: Washington, DC, USA, 2018.
48. Ellingsen, P.; Tonholm, T.; Johansen, F.R.; Andersson, G. Learning from Problem-Based Projects in Cross-Disciplinary Student Teams. *Educ. Sci.* **2021**, *11*, 259. [CrossRef]
49. Jaiswal, A.; Karabiyik, T.; Thomas, P.; Magana, A.J. Characterizing Team Orientations and Academic Performance in Cooperative Project-Based Learning Environments. *Educ. Sci.* **2021**, *11*, 520. [CrossRef]
50. Koby, M.; Orit, H. Ten Challenges of Data Science Education. 2020. Available online: <https://cacm.acm.org/blogs/blog-cacm/246219-ten-challenges-of-data-science-education/fulltext> (accessed on 26 February 2021).
51. Hardin, J. Expectations and Skills for Undergraduate Students Doing Research in Statistics and Data Science. *Amstat News*, 1 September 2017. Available online: <https://magazine.amstat.org/blog/2017/09/01/undergraduateexpectations/> (accessed on 26 February 2021).