Self-Reflection & Socialization in Shaping Trajectories in Early Career Research: A LACOID-Based Text Mining Approach

Abstract

This machine learning assisted longitudinal study explores the multifaceted trajectories of early career bioscience researchers, focusing on the intricate interplay between self-reflection and socialization. The study capitalized on the extensive textual data generated from the Early Career Research Project's annual interviews and employed the Latent Code Identification (LACOID) method for data analysis and meaning construction.

The study aimed (1) to identify common factors impacting early career research trajectories and (2) to test the hypothesis that individual characteristics (e.g., gender, age, race) influence the comprehension of professional growth determinants. The results revealed a nuanced interplay between individual characteristics and the way bioscience Ph.D. students perceive influences on their early career research trajectories.

Keywords: machine learning, Latent Code Identification, text mining, trajectories in early career research, bioscience Ph.D. students

Introduction

The trajectory of early career researchers is multifaceted and shaped by intricate factors, with self-reflection and socialization playing pivotal roles. Spanning eight years, concluding in 2022, the Early Career Research (ECR) project tracked 336 bioscience Ph.D. students from 53 research universities across the United States to explore their developmental pathways (Feldon, Litson, Roksa & Griffin, 2023). The ECR project adopted mixed methods approaches that included longitudinal surveys, interviews, and participatory observation. The focus of this study is to analyze the rich, eight-year interview data from the ECR participants, seeking to shed new light on how self-reflection (involving critical examination of personal capabilities, aspirations, and motivations) and socialization (encompassing integration into the scientific community), collaboratively shape the trajectories of these early career researchers. By investigating this nuanced interplay, the study aims to deepen the understanding of the complex influences that guide the formation of researchers' professional identities, academic achievements, and personal growth.

Literature Review

The role of self-reflection and socialization in shaping early career researchers' trajectories is an emerging area of focus within academic literature. Using a theoretical framework incorporating both the Three-Career Model (Gläser & Laudel, 2008), which elucidates the multifaceted nature of academic careers (See Figure 1), and Bandura's (2000) social cognitive theory (See Figure 2), the brief literature review presented below aims to integrate these theories into a broader understanding of the topic.

[INSERT FIGURE 1 HERE]

[INSERT FIGURE 2 HERE]

One of the pioneering works that examine self-reflection is Nesbit's study on self-directed

leadership development (SDLD) (2012). The SDLD framework utilizes a set of meta-skills, such as managing emotional reactions to feedback and practicing self-reflection. This supports the importance of self-reflection in shaping an individual's professional pathway, contributing to the theoretical groundwork for the proposed study (Garcia & Roblin, 2008; Lent, Brown & Hackett, 2002; London, 2001).

Socialization, viewed through the lens of social cognitive theory, emphasizes the environment's role in shaping an individual's behavior. In this context, Byars-Winston and his colleagues' research (2010) provides evidence for the significance of social cognitive variables (such as academic self-efficacy and outcome expectations) in determining the academic interests and goals of undergraduate students in science and engineering fields. Other similar studies also underscore the importance of environmental factors (socialization) in shaping career trajectories (Byars-Winston & Rogers, 2019; Chemers, 2011; Garriott, et al., 2013).

Bandura and his colleagues' work further extends the understanding of social cognitive theory, indicating that children's perceived academic, social, and self-regulatory efficacy directly influence the types of occupational activities they consider for their careers. It indicates that an individual's self-efficacy beliefs can shape their career aspirations and trajectories.

While these studies provide significant insights, the existing literature does not explicitly address the combined effect of self-reflection and socialization in shaping trajectories in early career research. Most of these studies are focused either on self-reflection or socialization, leaving a gap in understanding the synergy between the two in the context of researchers' early career development. Furthermore, these theories have been largely applied to children and undergraduates, and less is known about their applicability to early career bioscience researchers.

In summary, while the current literature provides essential insights into the role of selfreflection and socialization, it still lacks a comprehensive view on how these two aspects interact and shape early career bioscience researchers' trajectories. This proposed study aims to bridge this gap, contributing to the theoretical and empirical understanding of early career development in the STEM research field.

Purpose of Study

This preliminary study aims to use an advanced machine learning-based integrative analytical framework - Latent Code Identification (LACOID) – to classify huge volumes of textual data generated from the 8-year ERC project's annual interviews of the participating bioscience Ph.D. students, in order to

(1) identify and extract the self-perceived, common factors impacting these bioscience Ph.D. students' trajectories in early career research across the three critical milestones (i.e., enrollment in a bioscience Ph.D. program, completion of the Ph.D. program, & 1-3 years after Ph.D. graduation), and

(2) test the hypothesis if certain individual characteristics, such as prior graduate-level research experiences, gender, age, and race/ethnicity, result in significant differences in the way early career bioscience researchers comprehend the determinants of their professional growth and trajectories.

Method

In the analysis of the extensive textual data collected from the 8-year ERC project's annual interviews (i.e., Year 1, Year 4, and Year 7) with participating bioscience Ph.D. students, the Latent Code Identification (LACOID) method will be employed. This method incorporates natural language processing, text mining, and machine learning algorithms to identify patterns and classify textual data effectively and consistently. LACOID is designed to classify large volumes of textual data, while preserving the original context and nuanced meanings (González Canché, 2023). By mirroring line-by-line coding traditionally used in manual identification,

LACOID's algorithms will enable the rapid classification of self-perceived factors impacting the Ph.D. students' early career research trajectories. Moreover, the method's ability to process file naming conventions will facilitate the testing of hypotheses related to individual characteristics, such as prior graduate-level research experiences, gender, age, and race/ethnicity. The LACOID framework offers an innovative and efficient approach, capable of extracting insights that are pivotal to understanding the professional growth and trajectories of bioscience Ph.D. students (González Canché, 2023).

All data analyses were conducted using the free software program LACOID (González Canché, 2022).

Study Sample

The original data for the Early Career Research (ECR) project was collected under NSF grants 1760894 & 1431234 and is now available for public use (Feldon, et al., 2022). Participants included 336 bioscience Ph.D. students from 53 research universities across the United States. Data include yearly surveys and biweekly assessments sent to all participants who opted to participate and met criteria for remaining in the study each year, yearly scores on a sole author submitted research paper writing sample graded by 2 trained experts and averaged across expert ratings, and yearly interviews conducted with a subset of participants across eight years from 2015 to 2022.

Due to the complexity and scale of the original dataset, this preliminary study focused ONLY on three years of interview data (i.e., Year 1, Year 4, & Year7), signifying three critical milestones in the participants' early career research journey (i.e., enrollment in a bioscience Ph.D. program, completion of the Ph.D. program, & 1-3 years after Ph.D. graduation). Thus, the demographic composition of the 120 interviewees reveals a diverse group of early career

bioscience researchers. A significant majority (99.17%) have had some prior research experience, however, only 23.53% have engaged in graduate-level research. The birth years range from 1974 to 1993, with a median birth year of 1990. Approximately 35.83% of the participants were born before the year 1990, and 64.17% born after 1990. There is a higher representation of females (65.83%) compared to males (34.17%). In terms of racial categories, most of the interviewees identify as White (75.83%), followed by Black or African American (11.67%), Latino/Latina (10.83%), and Asian or Asian American (8.33%). A small percentage identify as American Indian or Alaska Native (1.67%), and Others (1.67%), with no representation from Native Hawaiian or Other Pacific Islanders.

Data Analysis Plan

Text preparation and mining procedures. Text preparation and mining procedures involve several steps to ensure effective analysis. This includes removing characters that don't form meaningful words, such as symbols and certain punctuation, and transforming all letters to lowercase to prevent duplicate words with different capitalization. Stop words, which don't contribute meaning to a text chunk, are also eliminated. A crucial part of the process is word lemmatization, which differs from stemming. Lemmatization produces longer vocabulary, exemplified by variations of the word "educate" (e.g., "educational," "educated," "educate," "education.") This method requires more computational power but is justified for text classification purposes (González Canché, 2023).

Machine learning, Gibbs sampling, and optimal code identification. Following the approach of Griffiths and Steyvers (2004), LDA models analyze text chunks by determining the probability distribution of various topics, referred to as codes. These probabilities illustrate how words and texts represent specific latent codes. The learning process identifies the likelihood of a

word being connected to a code and the probability of a word within a text (sentence or paragraph) being linked to a code (Feinerer, 2020; Silge & Robinson, 2017). This learning relies on the frequency with which words and texts have been relevant to a latent code. Initially, machine learning procedures randomly allocate word-to-code and word-to-text-to-code, then enhance these allocations based on previous learnings and methods like MCMC and Gibbs sampling (Grün & Hornik, 2011), using Bayesian statistics and statistical physics. Over thousands of iterations, this method refines the initial random assignments, reaching a suitable baseline accuracy after approximately 5000 iterations, discarding the first 500 (Raftery and Lewis, 1991).

Optimal code number identification. According to Griffiths and Steyvers (2004), the optimal number of codes in text mining can be determined using "perplexity," a measure of uncertainty in predicting a word in natural language processing (NLP). Lower perplexity values indicate better performance. In choosing the ideal number of codes, LACOID relies on four different metrics developed by various researchers. Although all aim to identify the optimal number of codes, they differ procedurally. Two metrics (i.e., Griffiths, 2004 & Deveaud, 2014) find the optimal number when the dissimilarity between topics reaches the maximum value of 1, implying clearer differentiation across text. In contrast, the other two metrics, proposed by Arun et al. (2010) and Cao et al. (2009), recognize the optimal number when the average dyadic correlation among topics is close to zero, indicating little overlap in words and frequency. Together, these metrics ensure that the selected codes will have few shared words, enabling precise LACOID analysis..

Hypotheses testing. The Chi-squared test reveals statistically significant associations between attributes and classified responses, indicating that these are not merely due to random

chance. This significance is shown by p-values smaller than three decimal places, computed by Monte Carlo simulations. In the relevant LACOID output, only blue and red rectangles, representing values more or less likely than expected respectively, drive these associations. The hollow solid rectangles do not diverge from expected values in a statistically meaningful way, thus not contributing to these associations.

Results

Year 1. Sixty-two interview transcripts were decomposed into 6,747 paragraphs (or text chunks), referred to as documents in text mining. Of these, 3,757 text chunks had less than 25 words. Since these texts have no meaningful linguistic variation, they were excluded from LACOID analysis. Thus, the text mining procedures included the 2, 990 paragraphs that have 25 or more words. Finally, all files yielded 22 groups based on individual characteristics for hypothesis testing.

Latent Dirichlet Allocation (LDA) & Gibbs Sampling analysis yielded 5-code solution as the optimal number of codes (See Figure 3.1 and Figure 3.2).

[INSERT FIGURES 3.1 & 3.2 HERE]

Due to the word limit, please see the detailed interpretation of the five codes in Table 1 below:

[INSERT TABLE 1 HERE]

Year 4. Eighty-two interview transcripts were decomposed into 8,998 paragraphs (or text chunks), referred to as documents in text mining. Of these, 3,728 text chunks had less than 25 words. Since these texts have no meaningful linguistic variation, they were excluded from LACOID analysis. Thus, the text mining procedures included the 5,270 paragraphs that have 25 or more words. Finally, all files yielded 21 groups based on individual characteristics for hypothesis testing.

Latent Dirichlet Allocation (LDA) & Gibbs Sampling analysis yielded 11-code solution as the optimal number of codes (See Figure 4.1 and Figure 4.2).

[INSERT FIGURES 4.1 & 4.2 HERE]

Due to the word limit, please see the detailed interpretation of the eleven codes in Table 2 below:

[INSERT TABLE 2 HERE]

Year 7. Sixty-five interview transcripts were decomposed into 8,483 paragraphs (or text chunks), referred to as documents in text mining. Of these, 3,508 text chunks had less than 25 words. Since these texts have no meaningful linguistic variation, they were excluded from LACOID analysis. Thus, the text mining procedures included the 4,975 paragraphs that have 25 or more words. Finally, all files yielded 19 groups based on individual characteristics for hypothesis testing.

Latent Dirichlet Allocation (LDA) & Gibbs Sampling analysis yielded 4-code solution as the optimal number of codes (See Figure 5.1 and Figure 5.2).

[INSERT FIGURES 5.1 & 5.2 HERE]

Due to the word limit, please see the detailed interpretation of the four codes in Table 3 below:

[INSERT TABLE 3 HERE]

Hypothesis testing for individual characteristics. As shown in Figures 6.1-6.3, chisquare tests across Year 1, Year 4, and Year 7 revealed intricate interplay between bioscience Ph.D. students' individual characteristics, such as prior graduate-level research experiences, gender, age, and race/ethnicity, and the way they comprehended what influenced their early career research trajectories.

[INSERT FIGURES 6.1-6.3 HERE]

Specifically for example, for Year 1, participants who are Black, female, born in the 1990's and had no prior graduate-level research experiences were significantly more likely to be associated with Code 1 (Personal Aspirations, Career Goals, & Early Influences) and Code 3 (Guidance, Interpersonal Relationships, & Various Social Factors), and less likely to voice Code 5 (Inquiries, Reflections, & Self-Clarifications). While their counterparts who are Black, male, born in the 1990's and had no prior graduate-level research experiences were significantly more likely to be associated with Code 1 (Personal Aspirations, Career Goals, & Early Influences) and Code 5 (Inquiries, Reflections, & Self-Clarifications). While their counterparts who are Black, male, born in the 1990's and had no prior graduate-level research experiences were significantly more likely to be associated with Code 1 (Personal Aspirations, Career Goals, & Early Influences) and Code 5 (Inquiries, Reflections, & Self-Clarifications), and less likely to voice Code 2 (Expectations, Opportunities, and Research Engagement).

Significance of the Study

The study's significance lies in three key dimensions. First, its systematic analysis of qualitative interview data from the ECR project offers a humane and epistemological perspective, providing insights that quantitative analysis cannot achieve. This approach reveals nuanced experiences and emotions, enriching our understanding of early career researchers' experiences. Second, the longitudinal perspective tracks the trends in early career research trajectories across three milestones of bioscience Ph.D. students. This temporal view is vital for understanding the evolving needs and experiences of these students, potentially informing educational policies. Third, the innovative use of the Latent Code Identification (LACOID) method represents a cutting-edge fusion of qualitative insights with machine learning. This not only speeds up the analytical process but also enhances reliability, marking a significant advancement in research methodology.

References

- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child development*, 72(1), 187-206.
- Byars-Winston, A., Estrada, Y., Howard, C., Davis, D., & Zalapa, J. (2010). Influence of social cognitive and ethnic variables on academic goals of underrepresented students in science and engineering: a multiple-groups analysis. *Journal of counseling psychology*, *57*(2), 205.
- Byars-Winston, A., & Rogers, J. G. (2019). Testing intersectionality of race/ethnicity× gender in a social–cognitive career theory model with science identity. *Journal of Counseling Psychology*, *66*(1), 30.
- Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues*, 67(3), 469-491.
- Feldon, D. F., Litson, K., Roksa, J., & Griffin, K. (2023). Trajectories in Early Career Research: Data Repository. DOI 10.17605/OSF.IO/HYMUS.
- Garcia, L. M., & Roblin, N. P. (2008). Innovation, research and professional development in higher education: Learning from our own experience. *Teaching and teacher education*, 24(1), 104-116.
- Garriott, P. O., Flores, L. Y., & Martens, M. P. (2013). Predicting the math/science career goals of low-income prospective first-generation college students. *Journal of counseling psychology*, 60(2), 200.
- González Canché, M. S. (2023). Latent code identification (LACOID): A machine learningbased integrative framework [and open-source software] to classify big textual data, rebuild contextualized/unaltered meanings, and avoid aggregation bias. *International Journal of Qualitative Methods*, 22, 16094069221144940.
- Lab, F., Litson, K., Jeong, S., Feldon, D., Searle, D., Brockbank, R., ... Zhang, F. (2022, September 22). Early Career Research (ECR) project. Retrieved from osf.io/ah6y3.
- Lent, R. W., Brown, S. D., & Hackett, G. (2002). Social cognitive career theory. *Career choice and development*, 4(1), 255-311.
- London, M. (2001). *Leadership development: Paths to self-insight and professional growth.* Psychology Press.

The Three-Career Model in Academia (Laudel & Gläser, 2008)



Figure 1. The Three-Career Model in Academia (Laudel & Gläser, 2008)



Figure 2. The Socialization Model Based on Bandura's (2000) Social Cognitive Theory



Figure 3.1. The Metrics Plot for Year 1 Optimal Number of Codes Solutions



Figure 3.2. The Distance Map for Year 1 5-Code Solution



Figure 4.1. The Metrics Plot for Year 4 Optimal Number of Codes Solutions



Figure 4.2. The Distance Map for Year 4 11-Code Solution



Figure 5.1. The Metrics Plot for Year 7 Optimal Number of Codes Solutions



Figure 5.2. The Distance Map for Year 7 4-Code Solution



Figure 6.1. The Chi-Square Test Results for the relationship between Year 1 Interviewee Characteristics and Codes Association



Figure 6.2. The Chi-Square Test Results for the relationship between Year 4 Interviewee Characteristics and Codes Association



Figure 6.3. The Chi-Square Test Results for the relationship between Year 7 Interviewee Characteristics and Codes Association

Test of independence, X-squared = 120.3, p.val =< 0

Code #	Code Name	Code Description	Example Quotes
V1	Personal Aspirations and Early Influences	Code V1 is indicative of personal aspirations, career goals, and early influences that shaped individuals' choices and decisions in choosing and entering a Ph.D. program/a research project.	Interviewee : Yeah, that's a really a good question because I mean I did molecular biology as my undergrad, and then I went to my master's where it was translational science, so it was much more clinically related. Now I'm kinda going back to like more basic science side. I mean, that was kinda based more off of my research background, because I've been doing like a lot of like basic science research, but it also has like a clinical element.
V2	Expectations, Opportunities, and Research Engagement	Code V2 embodies themes related to expectations, opportunities, and engagement in research projects.	Interviewee: Well, I don't know if this would be considered a research skill. In a way it is, but definitely presenting orally is something I need to work on, just giving better talks, and relaying scientists' information a little bit more clearly 'cause sometimes I still have some trouble with that. 'Cause I know what I'm trying to say, but it doesn't really come out the way that I'm trying to get it out. I just need to work on oral presentation skills 'cause talking at conferences is definitely something that we're gonna have to be doing throughout graduate school.
V3	Guidance, Relationships, and Decision Factors	Code V3 encapsulates the themes related to guidance, interpersonal relationships, and various factors influencing decisions.	Interviewee: Well, I'm not going to lie. I figured I would be able to get a decent education wherever I went. My application wasn't the strongest, in my opinion. I only had like a [number A] GPA and I know a lot of people who are applying definitely have higher GPAs, and though my research experience was really good, I didn't see myself as a strong candidate, so to be honest, I applied pretty much exclusively to where I could see myself living, so I applied to [institution B] and then I applied to [UNIVID 28], because that is where that professor who prompted me to join a Ph.D. program moved to, and I got denied by all the other programs I applied to, so that is why I am here

Table 1. Interpretation of Year 1 5-Code Solution Based on LACOID Text Mining Results

V4	Program Characteristics, Teaching, and Career Goals	Code V4 revolves around the characteristics of academic programs, teaching experiences, and the alignment with career goals.
V5	Inquiries, Reflections, and Self-Clarification Remarks	Code V5 is characterized by inquiries, reflections, and self- clarification remarks typically made by the interviewer.

right now. So, there wasn't really a big draw. I l didn't look at the qualifications of the programs or what kind of research they were doing. I can get inspired by pretty much any type of science, so I chose to apply based on the location and where I thought I might be able to get in.

Interviewee: I just found it really interesting. That's really all. The program I'm in is microbiology, immunology, and cancer. I'm looking at immunology in the context of cancer, and hopefully coming up with immunotherapeutic treatments for cancer. I just find [inaudible 00:03:07] extremely interesting and [inaudible 00:03:09] maybe a practical application of that, toward treating cancer, appeals to me.

Interviewer: Mm-hmm. Okay. What are these expectations based on? Is this your prior undergraduate research or handbook descriptions or maybe what you were told by peers?

Code #	Code Name	Code Description	Example Quotes	
V1	Navigating Professional Relationships and Personal Choices	Code V1 reflects bioscience Ph.D. students' experiences with professional relationships, personal choices, and decisions that shape their paths. Themes include PI interactions, cultural and gender dynamics, personal boundaries, academic path, growth reflections, and challenges.	Interviewee: Never in lab. More in faculty. I did not wear my engagement ring around faculty until the very end, after I defended. My faculty didn't even know I had, at the time, boyfriend of eight years. I kept that very private. My lab knew. My lab is invited to my wedding. My lab has always been very supportive, but the minute people find out oh, you're engaged. Oh, if you have kids are you going to stay home? You should stay home when you have to have kids. Instead of the focus being on hey, I succeeded as a scientist, it's oh, but now you're going to be a wife, or you're eventually going to be a mom. And those questions start. And I feel like I had to hide a part of myself because there's a lot of prejudice against females and making those assumptions.	
V2	PI Interactions, Career Planning, and Skill Development	Code V2 delves into PI relationships, career planning, and skill development. Themes include PI relationships, career aspirations, management skills, school influence, research navigation, and open communication.	Interviewee : And then, I think actually here in October, she's trying to send me for training for a week for some modeling. So that will definitely help my expertise as well as help me to meet other people in my area as well.	
V3	Technical Skills, Adaptation, and Emotional Reflections	Code V3 focuses on technical skills, adaptation, and emotional reflections. Themes include technical skills, adaptation to methods, mentorship, emotional reflections, career direction, and collaboration.	Interviewee: Yeah. I don't know if she knows that that's my plan. I wrote a mini review with the post-doc in my first lab, and we rewrote some of my boss' edits, and he got very, very upset, so I need to clarify what I'm allowed to do, which are her edits, so that doesn't cause a similar conflict. 16:30 And then we'll kind of restructure the figures as we both agree on and with-it's tricky where, again, she's young and	

Table 2. Interpretation of Year 4 11-Code Solution Based on LACOID Text Mining Results

V4	Identity, Institutional	
	Choices, and Reflective	
	Growth	

Code V4 centers on personal identity, institutional choices, and reflective growth. Themes include personal identity, institutional choices, yearly reflections, politics, collaboration, and future plans. like, ha ha, remember when you (inaudible), so we're kind of torn between knowing that she wants that and knowing that that would help her build her lab's reputation quickly and attract more funding quickly, which I absolutely appreciate, but she also knows that I hate grad school and that I want to be out, and also just planning her career, quantity may be just as valuable, if not more so, than quality, not in terms of a crappy paper, but that makes more sense.

is friends with a lot of young, but very, very successful PIs who have gotten a lot of early career high-impact papers, and so I think we're kind of torn between knowing that that's what she wants and her thinking that we can get our story into science, which, on the record, we can't, and I would love if I'm wrong and you can see my paper and be

Interviewee: The graduate students who were in the lab before me also were helpful just to get me an idea of what to expect and some of the different aspects of the department rules and everything. Then some of my friends in other labs are great. I don't know. It's nice to commiserate when you have a failed experiment and you're just like, Oh

V5 **Research Projects**, Code V5 explores research Timelines, and Project projects, timelines, and **Relationships** relationships. Themes include multiple projects, challenges, yearly reflections, project relationships, types and relevance, and funding considerations. V6 Academic Pathways, Code V6 emphasizes academic pathways, industry considerations, Industry Considerations, and and future goals. Themes include

Interviewee: So I've worked on some projects that were a little bit afield for my main focus as a researcher, so those took a little bit more effort 'cause I needed to spend more time reading. But really, it depended a lot. Some projects are harder to write, others are pretty darn easy to write.

Interviewee: Well, I do really like research and I like being in lab and I really like doing the hands-on stuff and asking a question and seeing if you can come up with an answer to the question.

	Future Goals	academic direction, faculty dynamics, postdoc plans, weekly meetings, industry considerations, and clear goals.	Being a post-doc seemed like an easy way to extend that lifestyle as well as kind of buy myself a little more time and hopefully, Well, time to both make the decision and also, get more training or think about what other things I would need to do if I want to pursue whatever I want to do as my long-term goal. If I want to do teaching at a teaching college, I will likely have to do more teaching before that, right, either at a university or a community college or something. Looking into ways to get that.
V7	Publication Processes, Writing Challenges, and Collaborative Authorship	Code V7 focuses on publication processes, writing challenges, and collaborative authorship. Themes include publication progress, writing insecurities, journal selection, collaborative writing, departmental requirements, encouragement, and challenges.	Interviewee : Yeah, I mean, if it's a paper that I'm the first author on that I'm like, I make all the decisions, I guess, really, in terms of who writes what. My advisor was very, very much left that up to me. Like, I wrote the paper, and then whatever was more or less done, he would check it out, make some suggestions and then,.
V8	Advisor Relationships, Communication Dynamics, and Personal Feelings	Code V8 explores advisor relationships, communication dynamics, and personal feelings. Themes include advisor dynamics, communication and expectations, interviewer's engagement, personal feelings, frustrations, and career preferences.	Interviewee: not mentoring right now? Okay. So I'm trying to think. I guess sometimes I don't know, maybe this is just like, an insecurity, maybe it's not true, but sometimes I think [name redacted], like with my main mentor, she can sometimes I wish she would just be a little bit more honest with me about something. Like, sometimes I feel like she's bothered by something. And maybe it has, it probably has nothing to do with me, but you know, sometimes I wonder, am I actually doing okay? I don't know. I feel like I don't get very much feedback. So
V9	Reflective Inquiry, Background Influence, and Interactions	Code V9 emphasizes reflective inquiries, background influence, and interactions. Themes include reflective questions, background influence, experiences,	Interviewer : Yeah. Okay. Are there other ways that you feel like your background shaped your experiences? So like, where you grew up, religion, gender? We've already talked about ethnicity, sexuality, social class Are there any other aspects of your background that shaped your graduate training experiences?

relationships and interactions, and reflective responses.

V10 Yearly Reflections, C Publication Timelines, re and Developmental ar Milestones T

Code V10 focuses on yearly reflections, publication timelines, and developmental milestones. Themes include yearly reflections, interviewer's questions, publication timelines, excitement, authorship, and development. **Interviewee**: The second year, what word would I pick? I guess the second year .. Okay, maybe I'll describe it, and then I'll come up with a word if I describe it. Second year I think was more, just kind of almost like getting my feet wet. Because, my project is just kind of a big project, and I wasn't building it on anything anyone else had done in the lab. It was a lot of just trying to figure out where we want to start, what direction we want to go to, and trying just a lot of things that maybe didn't pan out. It was a lot of learning for me. I don't know if I can say starting out, like hyphen that and have that be one word?

V11 Relationships, Professional Interactions, and Interviewer Inquiries Code V11 delves into relationships, professional interactions, and interviewer inquiries. Themes include interviewer's inquiries, relationships and interactions, transitioning topics, facilitation, and affirmations. Interviewee: Yeah. I would say that we have a good working relationship. It's more of a boss-and-employee relationship than a friend relationship, which is, again, one of the reasons I chose the lab. That's not to say that he doesn't care about me personally, because I got married a couple of months ago, and that was not an issue for me to take some time off to do that. He definitely cares about us as individuals. It's not anything like that, but, at the end of the day, on a day-to-day basis, it's like a business. All right, did you do this experiment? What were the results you got? What's the next experiment? It depends on the week, but there's some weeks where we won't see him at all. There's some weeks where we'll see him every day, but we do have a weekly lab meeting, so we see him usually at least once a week minimum, and like I said, you could see him every day, but at least once a week minimum.

Table 3. Inter	rpretation of	Year 7 4-Co	ode Solution	Based on L	ACOID Text	Mining Results
						0

Code #	Code Name	Code Description	Example Quotes		
V1	Reflections on Personal Growth and Challenges	Code V1 reflects bioscience Ph.D. students' experiences with professional relationships, personal choices, and decisions that shape their paths. Themes include PI interactions, cultural and gender dynamics, personal boundaries, academic path, growth reflections, and challenges.	Interviewee : I have done my own education and research on various topics of race, which has been also a nice outlet to have a regular meeting with a bunch of girls. Actually, they weren't people I knew. I knew one of them, and now I've made a few other friends through this. It was a good social break throughout a quite isolating year to have this book club, and it was also a nice outlet to talk to people about stuff in a way that wasn't-no one was gonna shame you for it, and to start learning on our own.		
V2	Navigating Career Paths and Professional Development	Code V2 delves into PI relationships, career planning, and skill development. Themes include PI relationships, career aspirations, management skills, school influence, research navigation, and open communication.	Interviewee : I would definitely say friends 'cause they're the people friends within the lab, and also outside the lab, but within the program. 'Cause those are the people that you can vent to, and also vent about people, or vent about science, your projects not working. That would probably be number one.		
V3	Work Environment, Independence, and Shifting Perspectives	Code V3 focuses on technical skills, adaptation, and emotional reflections. Themes include technical skills, adaptation to methods, mentorship, emotional reflections, career direction, and collaboration.	Interviewee : The rotation students, they're in the lab for six weeks, but the undergrads, the one who is still here, has been in the lab for three years now. The one who left recently, who just graduated, was in the lab about the same amount of time. Those relationships probably have developed more and probably in ways more subtle than I am even aware of.		
V4	Shifting Gears, Transitions, and Leadership Roles	Code V4 centers on personal identity, institutional choices, and reflective growth. Themes include	<i>Interviewee</i> : Also, they enjoy working with customers, so they've been hired for their social skills. They have good social skills. I enjoy having conversations with them, talking		

personal identity, institutional choices, yearly reflections, politics, collaboration, and future plans. with them, seeing them online when we have meetings. That part was even way better than my expectations. I didn't even think that-I didn't realize that until a couple weeks in when I was like...