ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022
Paper ID #38038

Using Natural Language Processing to Explore Undergraduate Students' Perspectives of Social Class, Gender, and Race

Umair Shakir (Graduate Research Assistant)

My academic background is a bachelor's and master's in civil engineering (University of Engineering Technology, Lahore, Pakistan), and Ph.D. (Engineering Education, VT, the USA, expected in Fall 2022). My ten years of professional experience range from NESPAK (5 years), to Dubai (1-years), and assistant professor (The University of Lahore-3 years). I am certified in Project Management Professional (PMP). During my Ph.D., I served as a graduate research assistant on NSF projects and a teaching assistant at the Department of Engineering Education, VT. I have analytical skills in python, and R studio. I am using the state of the art natural language processing techniques to analyze national scale qualitative data. My dissertation research is focused on Pell grants and accessibility to engineering colleges for underserved populations in the US. I aim to work on international donor education-related projects, particularly in Pakistan

Sarah Ovink

Sarah M. Ovink is an Associate Professor of Sociology at Virginia Tech. Her research focuses primarily on inequalities in higher education pathways by race/ethnicity, gender, and income, using mixed methods of inquiry. Her publications have appeared in a variety of journals, including Gender & Society, Social Currents, and Research in Higher Education. She is a co-editor of the volume, Intersectionality and Higher Education: Identity and Inequality on College Campuses (2019, Rutgers University Press). She is the author of Race, Class and Choice in Latino/a Higher Education: Pathways in the College-for-All Era (2017, Palgrave Macmillan). She is the recipient of a 2015 NSF CAREER award, investigating intersectional inequalities in STEM and non-STEM undergraduate pathways. From 2016–2020, she was a co-PI of the Life Sciences Mentoring Program, which matched incoming life sciences majors with near-peer mentors to provide mentoring training. Dr. Ovink received her Bachelor of Arts in Sociology at Kalamazoo College, and her Master of Arts and Ph.D. in Sociology at the University of California, Davis. She lives in Blacksburg, VA with her husband Eric, two kids, and two dogs.

Andrew Katz (Assistant Professor)

Using Natural Language Processing to Explore Undergraduate Students' Perspectives of Social Class, Gender, and Race

Motivation

Students' experiences in higher education settings are stratified across their racial and socioeconomic identities. This leads to existing institutionalized inequities across those groups being reproduced. The engineering education community has a collective consciousness that individualistic factors, for example, race, gender, sexual orientation, socioeconomic and disability status, should not influence who is most likely to matriculate, persist, or graduate in engineering. The subgroups within the engineering education community regularly implement and evaluate student support programs focused on diversity, equity, and inclusion. These programs aim to reduce academic, economic, or social barriers encountered by the minoritized student population on their paths into and through undergraduate engineering education. Empirical evidence about the effectiveness of those student support programs is divergent, which highlights complex, dynamic issues about the role of social class, gender, and race in the engineering education experiences of minority students. We as engineering education researchers must question how students' experiences in a higher education institution are contributing to the differences across the majority and minority student groups. In the present study, we aimed to explore the perspectives of undergraduate students about the roles social class, gender, and race play in shaping their educational experiences in a higher education institutional context.

Literature Review

Based on existing empirical evidence, we understand that students' educational aspirations and outcomes are mediated by their racial, gender, and socioeconomic traits. For example, Black and Latino undergraduates were significantly less likely to persist in STEM majors relative to the White and Asian [1], [2]. In a similar vein, low-income students are significantly less likely to graduate within six years relative to their high-income counterparts. In response to societal calls for social justice and equity, higher education institutions have adopted various financial, social, and academic policy tools to address the different needs of minority students. Financial aid programs are partially successful at mitigating fiscal barriers for STEM students in households with inadequate economic resources [3], [4]. Financial aid increases STEM credit accumulation since it obviates students' need to work while studying [5], [6]. Academic programs like STEM-related work-study employment, undergraduate teaching assistantships in STEM programs, and undergraduate research opportunities are the institutional strategies to positively impact the educational experiences of minority students [7]. Universities facilitate social clubs, peer study groups, and on-campus living communities for minority students to enable and enhance their social involvement and satisfaction [8].

Despite the growth of some opportunities, underserved students continue to struggle long after their admission to those student support programs. These programs are generally rooted in an individual-centric perspective, which assumes that deficiencies in the life histories of minoritized students are responsible factors for inhibiting their college success. For example, student-faculty interactions typically have a positive impact on the academic performance of students, but Black students who interact more frequently with faculty are more likely to experience racial discrimination [9], [10]. This racial discrimination is negatively linked with students' retention, particularly in STEM fields. The on-campus social experiences of students from lower-social class households are different relative to their peers from higher-social class households, which contributes to differences in educational and career pathways of upper- and lower-social class students [11]. The students from the upper-social class consider their college enrollment period as an opportunity to utilize their preexisting capital to socially engage with others from similar backgrounds. The on-campus social prominence of upper-class students may consume the limited capital of lower-class students in imitating upper-class students, rather than using it for scholarly dedication and hard work.

Scholars of social justice and equity argue that structural characteristics of historical institutions and ethnic groups' relations are responsible for the prevailing norms of students' college success in engineering education culture[12]–[14]. Those institutionalized norms extend the privilege to the competitive and individualistic practices associated with the White majority students, but minority students conventionally do not reflect those practices. Therefore, scholars of social justice support that student support systems not only abandon the individualistic-deficiency perspective but also adopt the structural-centric perspective. For example, student support programs rooted in theoretical lenses of critical race and feminism will explicitly acknowledge the agency and habitus of minoritized students. Those programs will customize strategies and resources to augment the existing cultural and social capital of minoritized students. This enhanced capital will enable substantive improvements in the college success of those students.

In engineering education, [15], [16] contend that social class, race, and gender have been studied separately, and only a few published studies have focused on the students' embodied intersectionality marginalized ethnic and gender identities. For example, quantitative studies in engineering education, except large/national scale, generally aggregate African American, Hispanic and Native Americans female students due to the small numbers in each group, erroneously assuming that those minority female engineering students have similar academic and social experiences in higher education institutions [17]. [15] and his team published several studies investigating college-related outcomes disaggregated by gender and race using the MIDFIELD dataset of over 70,000 U.S. students. They argued that engineering education studies, which failed to disaggregate data intersectionally by gender and race, gave overgeneralized results since white males are overrepresented in engineering education. Recently, researchers have used intersectionality as a framework and examined how systems and experiences of social class, race, and gender impact students' experiences in engineering higher education settings [18], [19].

Theoretical Framework

Intersectionality as the framework is a lens to understand an individual's experiences by holistic convergence of their different identity dimensions (gender, race, ability, sexual orientation, and socioeconomic status) in an oppressive system [20]. Intersectionality theory rejects the additive approach which separately examines the experiences of individuals who encompass multiple marginalized identities. Researchers having an intersectionality perspective examine "what it means to live at the crux of structural inequality based on intersections" to address issues of power and privilege in society [13]. This study used intersectionality as the theoretical framework to answer the following research question: How do undergraduate students describe the role of social class, gender, and race in shaping students' college experiences in a higher education institution?

Methodology

Data Collection

This publication used data collected in a longitudinal study, six waves across the period of 2014-2021, of intersectional differences in college-to-career trajectories at an R1 Mid-Atlantic university in the US. During the first wave of the study between 2014-2016, the research team used the university email listserv, in-person presentations across academic departments and student organizations, and word of mouth to recruit participants. These efforts resulted in 113 first-wave interviews. Notably, this publication used only the data collected in the first-wave interviews.

In the first wave, participants included 41 engineering major (36.2%) and 72 non-engineering major (63.8%) undergraduate students. Of these, 54 (47.7%) identified as men, 58 (51.4%) as women, one as gender-fluid. Respondents included 43 White (38.1%), 29 Asian and Pacific Islander (25.6%), 19 Black (16.8%), 11 Latinx (9.7 %), 9 multiracial (7.9%), one Armenian, and one declined to answer. The participants also estimated their combined family's income range. Based on these qualitative responses, interviewees included 5 (< $25,000, 4.4%), 14 ($25,000 - $50,000, 12.3%), 81 ($50,000- 250,000, 71.6%), and 13 (>$250,000, 11.5%).

The second author developed an interview protocol comprising six question categories: background; college decisions and academics; gender, race, income, disability status, and college; college life; goals; family; friendships and peer groups; and conclusion. The complete interviews ranged from 30 minutes to over two hours. The reader should note, that this publication used only the interview transcript portion related to gender, race, income, disability status, and college. This specific portion of the interview protocol is given in Appendix-A.

Data Analysis

A member of the research team manually extracted interview text relevant to gender, race, income, disability status, and college. These excerpts were then cleaned to remove filler language such as "Umm", "chew", "mhmm", etc. Here, a noteworthy step (for data analysis) is that interviewee's response is considered as a single contiguous block of text unless interrupted by the interviewer. The cleaned interviewees' responses are thematically analyzed using a human-in-the-loop approach in two sequential steps: (i) text clustering via natural language processing (NLP) techniques and (ii) manual thematic coding.

During step (i), we took blocks of raw text (as mentioned above) of interviewee responses and embedded these text blocks in a high dimensional vector space (768) with sentence transformers based on the MPNET architecture. The next step was to reduce these high-dimensional embeddings from 768 dimensions to a lower-dimensional space (<5). This was done to improve the performance of a subsequent clustering algorithm. Historically, clustering algorithms suffer in higher (768 or 1024) or intermediate dimensional space (in the range of 50-100) since every point (i.e., text embedding) is far from every other point [21]. Therefore, these embeddings undergo a combination of linear and nonlinear dimensionality reduction steps using principal component analysis (PCA) and uniform manifold approximation and project (UMAP) [21], respectively. This combination of PCA and UMAP techniques is efficient at maintaining as much possible variance in the original embeddings and allowing cluster algorithms to work efficiently.

We used PCA to reduce from the high embedding space into an intermediate embedding space since (a) PCA can result in losing valuable variance in the data if there is too sharp of a reduction in the process of dimension reduction, and (b) UMAP can reasonably function well in less than 100-dimensional space [21]. So, the two-prong goal here was to reduce original embeddings (768) to an intermediate dimensional space where as much as possible original variance (75-90%) is retained in the data while still enabling the UMAP step to work efficiently. Experimentally, we found that an intermediate embedding in the range of 65-80 dimensions appears to balance these two considerations best. Therefore, first, we used the PCA technique to reduce the original embedding space (768-dimensional space) into an intermediate embedding space (in the range of 65-80 dimensions). Second, we used the UMAP technique to reduce from intermediate embedding space to a lower-dimensional space (< 5). After those dimensionality reduction steps, we used an agglomerative clustering algorithm with ward linkage and a Euclidean distance metric to cluster the data (i.e., blocks of interviewee responses).

We will refer to the round of clustering based on blocks of the interviewee as round-one-clustering. While maintaining the numerical labels of round-one-clustering, the blocks of interviewee responses were split into sentences with the sentence parser from an open-source python NLP library, Spacy. The purpose of sentence splitting was to mitigate the issue that individual blocks of texts contained multiple sentences that might be thematically disconnected from each other. Conversely, the vast majority of single sentences might express only one topic.

After splitting the blocks while maintaining the labels of round-one-clustering, we took the sentences of interviewee responses and again embedded them in a high dimensional vector space (1024) with sentence transformers based on the BERT architecture. We replicate as described above the dimension reduction and clustering sequence. First, we used PCA to reduce those high dimensional embeddings (1024) to intermediate dimensional space (in the range of 60-85). Second, we used UMAP to reduce from intermediate embedding space to a lower-dimensional space (< 5). Lastly, we again used an agglomerative clustering algorithm with ward linkage and a Euclidean distance metric to cluster the data. We refer to these results as round-two-clustering. Here we want to highlight to the reader that round-two-clustering only works on the textual data already clustered together within boundaries of clusters developed in round-one-clustering.

Each cluster could consist of 20- 60 statements that discussed the same topic. Step (ii) of the data analysis procedure used in this publication aligned with Hatch's Inductive Analysis Model [22]. We used an iterative process of coding to ensure reliability in coding and conducting higher levels of coding. A member of the research team coded each cluster by reading the responses. When the coder identified a theme in that entire cluster then they assigned one label to each of these responses and moved on to the next cluster. We estimated this coding procedure was significantly quicker than traditional – on the order of a five-fold reduction in time [21]. This system also helped to improve consistency by analyzing across the entire collection of texts simultaneously and grouping similar items. This combination of affordances enabled the analysis of these interviews in a more manageable fashion. This first cycle of open coding resulted in a preliminary, descriptive codebook. During the second cycle of coding, we revisited NLP clustering results to develop salient domains by refining and grouping initial codes. We displayed codes associated with each domain on a physical whiteboard to reflect patterns in the raw data. We then reread the NLP clustering results to select direct quotes and passages from interviewee responses to support the prevalence of domains (and associated codes) and to define their essence and meaning.

We estimate that our created NLP system can facilitate and enhance a researcher (or a research team) analysis process in two major ways. First, it will improve the consistency by analyzing across the entire collection of texts simultaneously and grouping similar items. Second, it will significantly reduce time in handling and analysis of large volumes of text corpora that were previously unwieldy to handle. At this point, there are several notable limitations to emphasize.

Limitations

First, there is a limitation from the data input to the NLP system - we subjectively consider each time an interviewee responded to the interviewer as a single contiguous block of text related to a single topic unless interrupted by the interviewer. This assumption could be wrong and a single contiguous block of text of the interviewee's response may be related to different topics. The second round of clustering helped to mitigate the impacts of this limitation because it allowed for the identification of multiple topics in each block. Second, the NLP system used in this paper is in its early development stage. A user requires advanced knowledge and skills in both NLP and computer coding to manually adjust the codebase to fine-tune the parameters of NLP systems to produce optimal clustering results. Additionally, the process of adjusting parameters of NLP systems is an admittedly subjective one, balancing intra-cluster heterogeneity (or homogeneity) and the number of clusters. Nonetheless, we believe the potential of this approach to assist qualitative coding of a large collection of interviews offers new opportunities to the research community; therefore, we chose to present the work even in its nascent stages. Third, there were also additional limitations regarding the manual qualitative coding process of NLP clustering results. We considered clusters as noise when there was too much heterogeneity within a cluster. The clusters labeled as noise have been set aside during thematic analysis. With that in mind, we may have lost a nontrivial portion of interviewee responses in the present study. We believe we minimized the impacts of this limitation because the topics represented in some of these heterogeneous clusters appeared elsewhere in more homogeneous clusters and therefore were still represented in our final results.

Findings

We consolidated the findings of this study in Table 1 and supplemented these findings with illustrative quotations in the respective sections below.

Table 1: Summary of Domains and Findings

Domain	**Findings**
Social Class	• Students' experiences in college facilitated by family income • Student's academic progress in college inhibited by their limited financial support from their family • Social class does not matter in college
Gender	• Female students experienced their female gender as a social stigma • Engineering and/or STEM culture is dominated by the male gender • Women now have better opportunities than past
Race	• Systemic inequities exist between White and non-White races • Equal opportunities for all regardless of race, nowadays • White students tended to disfavor consideration of race in financial aid offers • Students socialize with same-race peers in college.

Domain: Social Class

We found four different themes in the domain of social class. First, students' experiences in college were facilitated by their family income. A White and continued-generation student expressed her feeling about the availability of parents' resources for college expenses as:

> "I realize that I'm lucky that I have parents who looked to the future and saved money for me to go to college."

In a similar vein, White male students from high-income families recognized agency given to them by being born in high-income families.

> "I feel not that it's not possible to go far without being born into but being born into a rich family is a, is overwhelming, puts you above almost anyone else"

> "Then, people who are financially well-off, just because that help, that gives you greater agency when it comes to what you want to do."

Second, high-income family students expressed concern about fellow students' academic progress due to limited financial support from their family and their employment during studying.

> "...paying for books and stuff and, and a laptop and everything, um, you know, if you don't have the books or the laptop, I don't think that you can do well at all."

> “I think, one of the best things about it, is not, is being able to fully focus on academics; I have friends who have to take up job, and then they have to balance having a job with studying”

Third, we found conversely that students considered college a level-playing field regardless of class status. An Asian/Pacific Islander, first-generation engineering student mentioned as:

> “If you're considered to be in a particular class especially in this day and age depending on income level that you have, I don't see a difference”

A fourth theme was that students did not talk about their family income to their peers, as mentioned by a student:

> “I feel like I don’t really talk about income level with other people, and I don’t know some people don’t feel comfortable talking about their income level so that’s not really a factor when I talk to other people or associate myself with everyone else.”

Domain: Gender

In the domain of gender, we also found variation in students’ descriptions about the role of gender in college experiences. First, we found female students experienced their female gender as a social stigma. A Black female student from a high-income family expressed her feeling as:

> “And men just look at women as objects and not as people”

An Asian/Pacific Islander female engineering student said:

> “it’s just they don’t care about if women even have a space in life, they’re just like we’re workers and we’re baby makers that’s it that’s literally our role”

The social pressure being women may give female students the motivation to pursue their goals. An engineering female student commented:

> “I mean, there are going to be people that underestimate you for being a female but…it’s the best feeling in the world to prove them wrong.”

Other students observed that the culture of engineering and/or STEM at the institution is dominated by males. A White-male engineering student mentioned:

> “... I’m in an engineering heavy thing, I don’t expect to see too many girls and when I do see a bunch in a class more than say 30 or 40%, it’s very surprising.”

A Black-male student majoring in liberal arts similarly observed:

> “...most groups here are um, pretty dominated by men, and are, let’s just talk about this fact that this is a school that is pretty focused on the STEM majors, and STEM majors are typically pretty high and men, so for those majors, men typically don’t try to be inclusive of women”

Nevertheless, there were opinions to the contrary, expressing how women were not excluded from engineering education, even though women are a minority in the engineering field. A Black-female engineering student mentioned:

> "I don't see where the men say oh women can't lead, or women can't be engineers"

Female students also acknowledged that women have better opportunities than in the past as a White-female stated:

> "I feel like, like women as a whole group have a better opportunity to get ahead than in the past, just based on, like, in the past, science was like a man's thing, and, like, women were mostly known, like, to do housework, a lot of, like, smaller, like, not as significant thing; now, it becomes more developed, there is feminism, equality for females and males, so they have a lot more opportunity."

A Black female student warranted better opportunities for females than past with cautionary tone:

> "Well, I don't wanna to say, "do not"; I think we have an equal chance, technically have an equal chance."

Domain: Race

Finally, as with the other two domains, we found variation in students' descriptions about the role of race in college experiences. A Black, first-generation, male student talked about existing inequities across White and Black races in the society:

> " Yeah, whites can get ahead more easily, and I mean blacks have made great strides obviously and it's easier for us than it used to be but still not nearly equal I don't think"

Conversely, another Black, first-generation, male student talked about equal opportunities for all regardless of race:

> "I feel like for the most part people are on the same playing field when it comes to college and getting ahead"

A White, male, engineering student supported those individualistic ideals of success in the current society:

> "I mean I don't really see a lot of discrimination or people doing worse just because they are from a minority group or not of the…I don't really see like one specific group of people doing worse it is kind of just random different people do better more it is just based on themselves not really one whole group."

Another important theme in our findings is that White students tended to disfavor consideration of race in financial aid offer as mentioned by a White, first-generation male engineering student:

> "Something or you don't get these five bonus points because you're white"

A Black, female engineering student talked about the disadvantage of being White for her friends as:

> “Oh, they’re white, of course they will just get in, but then you, I see plenty of my friends who are like we are like the middle class average white person, we get no financial aid.”

A Hispanic student mentioned advantages of affirmative actions related to minority races as:

> “But I still had the advantage of being Hispanic or whatever, and I’m sure that contributed in some way when you put it on an application because you know what’s it called, So I have an advantage even though I don’t really deserve and advantage because I came from a really stable family, with a stable income, so I guess that gives me an advantage more now than it did in the past.”

We also identified variations in the description of how students socialize with peers in college. On one end, students of the same race hang together naturally as illustrated by an Asian female student:

> “And, socially, whites are hanging out together, and Asians hang out together; within Asian students, Asian Americans hang out together, and international students have their own groups.”

On the opposite end, students mentioned they do not discriminate against students based on different races in a social setting as elucidated by a White female student:

> “I don’t think there’s really a big difference cause I interact with a lot of different people, and it doesn’t really matter their race.”

Discussion

The results demonstrated a wide variation in students’ perspectives of the role of social class, gender, and race in shaping their (and other students') college experiences. Such variation may challenge prevailing assumptions about the uniform effectiveness of student support programs across various racial, gender, and socioeconomic groups. [23] has given an example of Black lesbian engineering students whose reality cannot be adequately understood by discrete conversations on race, gender, and sexual orientation into three separate different groups. We should in designing both research and practice explicitly focus on the intersectionality of multiple marginalized identities.

The minoritized students in this study who embody intersectional marginalized identities face systematic oppression of classism, racism, and sexism. These inequitable experiences shape the decision-making of a minoritized student to enroll, persist or drop out from the engineering education ecosystem. Those students have to divert resources (cognitive, emotional, time) in navigating engineering spaces to avoid social pain caused by a misalignment between dominant engineering culture and the intersection of their socioeconomic, gender, and racial identities [[12]. Engineering culture has unique barriers for persons who embody intersectionality marginalized identities due to engineering being a predominantly White and male space. Yet, as we saw here in this study, students’ perceptions are more heterogeneous even within a particular group. Such heterogeneity underscores the importance of considering each student's lived experiences and the important nuances researchers and educators must consider.

Here we also want to expound on the implications of our created NLP system for the engineering education community. Because of the small numbers of historically marginalized groups in engineering, engineering education researchers have tended to obviate some elements of intersectional identities. In particular, those using quantitative methodologies may have considered it "methodologically necessary to aggregate all women together even when subjects' experiences differ by race or to aggregate all African American persons together even when their experiences differ by gender" [13]. Unfortunately, in doing so, quantitative research methodologies may lose the subtle nuances of students' experiences. On the other end of the methodological spectrum, qualitative research methodologies can capture the rich description of students' experiences, yet those methods can be resource-intensive and have issues related to scalability and transferability. These logistical challenges and intrinsic quality control issues in the qualitative research paradigm may be partially addressed by recent state-of-the-art developments in the NLP.

Conclusion and Future Work

This study publication was based on a research project that used a longitudinal qualitative research methodology. The research site was an R1, Mid-Atlantic university. For the first qualitative phase of the research project, in-depth semi-structured interviews were conducted with 113 undergraduate students at the university - 41 students were from engineering majors, and 72 were outside of engineering majors. Of the total 113, 5 % (6 out of 113), and 10% (12 out of 113) of the study sample reported their annual combined family income as less than 25 thousand dollars, or more than 250 thousand dollars, respectively. This study publication analyzed a portion of those transcribed interviews that were pertinent to the role of social class, gender, and race in shaping students' college experiences. With the help of an NLP, human-in-the-loop workflow, we took, embedded that text corpus, using a pre-trained transformer (a specific kind of neural network architecture trained to encode inputs and decode outputs), and perform a sequence of dimension reduction techniques capped with a final clustering step. The research team then utilized these groupings to perform a thematic analysis of this interview with a more nuanced understanding than only a human could do. Informed by an intersectionality lens, the results identified a wide variation (or complexity) in students' perspectives of the role of social class, gender, and race in shaping their (and or other students') college experiences.

In the future, we intend to enhance the scope of analysis using the same NLP-assisted approach used in this publication for the remaining portions of the interviews collected data in the research project. Those interviews contain questions and comments about e remaining portion of the project data related to students' high school to college transition, college experience, family and friends, and career goals. Further, we also intend to build a convenient application programming interface so that a researcher without a familiarity with computer coding could use our created NLP approach to analyze a large corpus of textual data in their work.

Acknowledgment
This material is based upon work supported by the National Science Foundation under Grant No. 1453213 and Grant No. 2107008

References

[1] L. T. O’Brien, A. Blodorn, G. Adams, D. M. Garcia, and E. Hammer, “Ethnic variation in gender-STEM stereotypes and STEM participation: an intersectional approach,” *Culture. Divers. Ethnic Minor. Psychol.*, vol. 21, no. 2, pp. 169–180, Apr. 2015, DOI: 10.1037/a0037944.

[2] V. Lundy-Wagner *et al.*, “Gaining Access or Losing Ground? Socioeconomically Disadvantaged Students in Undergraduate Engineering, 1994–2003,” *J. High. Educ.*, vol. 85, no. 3, pp. 339–369, 2014.

[3] S. Goldrick-Rab, *Paying the Price: College Costs, Financial Aid, and the Betrayal of the American Dream*. Chicago, IL, UNITED STATES: University of Chicago Press, 2016. Accessed: Mar. 25, 2020. [Online]. Available: http://ebookcentral.proquest.com/lib/vt/detail.action?docID=4519377

[4] G. T. Henry, R. Rubenstein, and D. T. Bugler, “Is HOPE Enough? Impacts of Receiving and Losing Merit-Based Financial Aid,” *Educ. Policy*, vol. 18, no. 5, pp. 686–709, Nov. 2004, doi: 10.1177/0895904804269098.

[5] B. L. Castleman, B. T. Long, and Z. A. Mabel, *Financial Barriers to STEM Study in College: Causal Effect Estimates of Need-Based Grants on the Pursuit and Completion of Courses and Degrees in STEM Fields*. Society for Research on Educational Effectiveness, 2014. Accessed: Apr. 13, 2021. [Online]. Available: http://eric.ed.gov/?id=ED562774

[6] B. L. Castleman, B. T. Long, and Z. Mabel, “Can Financial Aid Help to Address the Growing Need for STEM Education? The Effects of Need-Based Grants on the Completion of Science, Technology, Engineering and Math Courses and Degrees.,” *J. Policy Anal. Manage.*, vol. 37, no. 1, pp. 136–166, 2018.

[7] M. J. Chang, J. Sharkness, S. Hurtado, and C. B. Newman, “What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups,” *J. Res. Sci. Teach.*, vol. 51, no. 5, pp. 555–580, 2014, doi: 10.1002/tea.21146.

[8] M. Ong, J. M. Smith, and L. T. Ko, “Counterspaces for women of color in STEM higher education: Marginal and central spaces for persistence and success,” *J. Res. Sci. Teach.*, vol. 55, no. 2, pp. 206–245, 2018, doi: 10.1002/tea.21417.

[9] K. A. Griffin, “Institutional barriers, strategies, and benefits to increasing the representation of women and men of color in the professoriate: looking beyond the pipeline,” *High. Educ. Handb. Theory Res. Vol. 35*, pp. 1–73, 2019.

[10] K. A. Griffin, “Achieving Diversity at the Intersection of STEM Culture and Campus Climate,” 2019.

[11] E. A. Armstrong and L. T. Hamilton, *Paying for the Party: How College Maintains Inequality*. Harvard University Press, 2013. Accessed: Nov. 18, 2021. [Online]. Available: http://www.jstor.org.ezproxy.lib.vt.edu/stable/j.ctt2jbr6k

[12] A. L. Pawley, “Learning from small numbers: Studying ruling relations that gender and race the structure of U.S. engineering education,” *J. Eng. Educ.*, vol. 108, no. 1, pp. 13–31, 2019, doi: 10.1002/jee.20247.

[13] A. E. Slaton and A. L. Pawley, “The Power and Politics of STEM Research Design: Saving the ‘Small N,’” Jun. 2015, p. 26.1564.1-26.1564.13. Accessed: Jul. 30, 2021. [Online]. Available: https://peer.asee.org/the-power-and-politics-of-stem-research-design-saving-the-small-n

[14] C. Baillie, A. L. Pawley, and D. Riley, Eds., *Engineering and Social Justice: In the University and Beyond*. West Lafayette, IN 47907 USA: Purdue University Press, 2008. doi: 10.2307/j.ctt6wq5pf.

[15] S. M. Lord, M. M. Camacho, R. A. Layton, R. A. Long, M. W. Ohland, and M. H. Wasburn, “WHO’S PERSISTING IN ENGINEERING? A COMPARATIVE ANALYSIS OF FEMALE AND MALE ASIAN, BLACK, HISPANIC, NATIVE AMERICAN, AND WHITE STUDENTS,” *J. Women Minor. Sci. Eng.*, vol. 15, no. 2, 2009, doi: 10.1615/JWomenMinorScienEng.v15.i2.40.

[16] M. W. Ohland *et al.*, “Race, gender, and measures of success in engineering education,” *J. Eng. Educ.*, vol. 100, no. 2, pp. 225–252, 2011, doi: 10.1002/j.2168-9830.2011.tb00012.x.

[17] J. A. Whittaker and B. L. Montgomery, “Cultivating institutional transformation and sustainable STEM diversity in higher education through integrative faculty development,” *Innov. High. Educ.*, vol. 39, no. 4, pp. 263–275, 2014.

[18] M. Ong, N. Jaumot-Pascual, and L. T. Ko, “Research literature on women of color in undergraduate engineering education: A systematic thematic synthesis,” *J. Eng. Educ.*, vol. 109, no. 3, pp. 581–615, 2020, doi: 10.1002/jee.20345.

[19] S. Secules, N. W. Sochacka, and J. Walther, “New Directions from Theory: Implications for Diversity Support from the Theories of Intersectionality and Liberatory Pedagogy,” presented at the 2018 CoNECD - The Collaborative Network for Engineering and Computing Diversity Conference, Apr. 2018. Accessed: Jul. 30, 2021. [Online]. Available: https://peer.asee.org/new-directions-from-theory-implications-for-diversity-support-from-the-theories-of-intersectionality-and-liberatory-pedagogy

[20] K. Crenshaw, “Mapping the margins: Intersectionality, identity politics, and violence against women of color,” *Stan Rev*, vol. 43, p. 1241, 1990.

[21] A. Katz, M. Norris, A. M. Alsharif, M. D. Klopfer, D. B. Knight, and J. R. Grohs, “Using Natural Language Processing to Facilitate Student Feedback Analysis,” presented at the 2021 ASEE Virtual Annual Conference Content Access, Jul. 2021. Accessed: May 11, 2022. [Online]. Available: https://peer.asee.org/using-natural-language-processing-to-facilitate-student-feedback-analysis

[22] J. Amos. Hatch, *Doing qualitative research in education settings*, 1 online resource (xi, 299 pages) vols. Albany: State University of New York Press, 2002. [Online]. Available: http://site.ebrary.com/id/10587283

[23] L. Bowleg, "When Black+ lesbian+ woman≠ Black lesbian woman: The methodological challenges of qualitative and quantitative intersectionality research," Sex Roles, vol. 59, no. 5, pp. 312–325, 2008.

Appendix A

Excerpt of Interview Protocol Related to Social Class, Gender, and Race

Background	- Age
	- Race/ethnicity (when you are asked)
	- Race/ethnicity (at home or among friends)
	- Born in the United States?
	- IF NO: Keeping in mind that your responses are confidential and you can opt out of responding, what is your current citizenship status?
	    - Probe: are you currently working towards gaining citizenship in the US?
	- Please tell me a bit about your life growing up: your siblings, your parents, and the places you have lived.
	- What was the highest grade/degree your parents completed?
	- IF COLLEGE: what college?
	    o Probe: [if R says parents did not attend] Did they participate in any post-high school training of any kind?
	- Occupation(s) of parent(s)
	IF ANY SIBLINGS: What are your siblings doing now?
	- Do your parent(s)/guardian(s) own or rent their residence?
	- Can you estimate your parents' income level?
	    o Probe: [Show card] Here are a number of income ranges. Would you say your parents' combined income is (less than 25k, between 25-50k, between 50-100k, 100-250k, more than 250k?)

Social Class, Gender, and Race	Are there any people or groups that have a better opportunity to get ahead today than in the past? -IF NO: So there are no people or groups that have better opportunities today as compared with in the past? -IF YES: Please tell me more about that. What makes you think so? How does knowing that make you feel (good, bad, it's unfair, etc.)? Do you think that people of all races have an equal chance to do well at [this college]? Why? Why not? Do you think that people of all genders have an equal chance to do well at [this college]? Why? Why not? Do you think that people with disabilities have an equal chance to do well at this college? Why? Why not? Do you think this college is accessible for people with disabilities? Do you think people of all income groups can do well at [this college]? Why? Why not? What is it like to be [insert respondent's stated race/ethnicity] at this school? What is it like to be [stated gender] at [this college]? What is it like to be [stated income level] at [this college]? What is it like to be [a person with a disability] at this college?