Science, Engineering, and Research Identities in Graduate Engineering Education: A Latent Profile Analysis

Dr. Heather Perkins, Purdue University Northwest

Heather Perkins holds a Ph.D. in Psychology from North Carolina State University and a B.S. in Psychology from the University of Cincinnati. Before joining Purdue University Northwest as an Assistant Professor in 2025, she served as a Visiting Assistant Professor at Indiana University Bloomington and a Postdoctoral Researcher in the School of Engineering Education at Purdue University West Lafayette. Heather's work examines how identity shapes motivation, confidence, and academic trajectories, particularly in graduate and underrepresented student populations.

Mr. Matthew Bahnson, University of Cincinnati

Matthew Bahnson completed his Ph.D. in the Applied Social and Community Psychology program in at North Carolina State University. His previous training includes a B.A. in Psychology from the University of Northern Iowa and an M.A. in Social Sciences from the University of Chicago. Matthew's research focuses on sociocultural inequality in engineering graduate education with the intention of increasing diversity, equity, inclusion, and justice in STEM graduate education. He is completed a postdoctoral appointment in engineering education with the Engineering Cognitive Research Laboratory with Dr. Catherin Berdanier at Pennsylvania State University. He recently worked as a Research Scientist at Purdue University with the STRIDE research group directed by Dr. Allison Godwin at Cornell University. Matthew is currently an Assistant Professor of Psychology at the University of Cincinnati.

Dr. Marissa A Tsugawa, Truckee Meadows Community College

Marissa Tsugawa is an instructor at Truckee Meadows Community College who leverages mixed-methods research to explore neurodiversity and identity and motivation in engineering. They completed their Ph.D. in Engineering Education where they focused on motivation and identity for engineering graduate students.

Dr. Adam Kirn, University of Nevada, Reno

Adam Kirn is an Associate Professor of Engineering Education at the University of Nevada, Reno.

Science, Engineering, and Research Identities in Graduate Engineering Education: A Latent Profile Analysis

This empirical full research paper discusses a latent profile analysis of graduate students' engineering identities. Drawing from role identity theory [1], engineering identity captures how individuals come to view themselves as a particular "kind of person" within engineering [2] and that they belong in engineering. It is associated with important outcomes like career choice [3], [4], increased motivation [5], and persistence [6], [7], [8]. While research on undergraduate engineering identity has been extensive, the unique dynamics of identity formation among graduate engineering students are not as well understood [9]. Like undergraduate engineering identity, graduate engineering identity is theorized to encompass sub-constructs related to interest, performance/competence, and recognition [10], but it also expands into three domains of engineering, science, and research [10], [11], [12]. This is likely a reflection of the shift away from undergraduate coursework and internships [13], [14] and towards graduate dissertations and research assistantships [15], [16]. But exactly how these experiences and sub-domains interact is not known.

As graduate engineering identity has been recognized as a predictor of persistence [17], better understanding how the sub-domains of science, engineering, and research identity contribute to graduate engineering identity may also help predict attrition. Attrition is a serious issue in graduate engineering education, with more than half of doctoral students contemplate leaving without finishing [18]. While the stresses of graduate school are frequently cited as reasons for leaving, such as navigating advisor issues and relationships, dealing with the stress of coursework and qualifying exams, and managing research challenges and completing dissertations [19], [20], the new challenges in the domains of science and research may also contribute as students to learn anew what it means to 'be an engineer' [11]. Taken together, existing research establishes that engineering identity matters for persistence and that graduate students navigate multiple identity domains. However, we lack a person-centered understanding of whether graduate students experience aligned, misaligned, or uneven identity configurations, and their implications for attrition. Addressing this gap is necessary to clarify the structure and impact of graduate engineering identity and its impact on student outcomes.

The present study sought to examine how engineer, scientist, and researcher identities co-occur among graduate engineering students and relate to persistence using a two-step latent profile analysis [21]. First, latent profile analysis was used to identify latent groups or clusters of students characterized by similar patterns of identity across these domains. Next, auxiliary analyses were used to better understand how these latent profiles relate to other aspects of identity (e.g., discipline, race, international status, gender, and degree progress) and the relationships between the identity profiles and degree uncertainty. The research was guided by three overarching questions: RQ1: Do graduate students' engineering identities form latent profiles with distinct, complex interrelations across the SER domains, and what do these profiles communicate about heterogeneity in graduate engineering identity? RQ2: Is there an under/overrepresentation of students in specific latent profiles by discipline, race, international status, gender, or degree progress? RQ3: To what extent do the graduate engineering identity profiles explain degree uncertainty?

Method

Table 1: Demographics of final sample

Milestones Completed	Count	Percentage
No Milestones	324	29%
Comprehensive Exam (Written or Oral)	373	33%
Dissertation Proposal	286	25%
Dissertation Defense	51	5%
Not Answered	99	9%
Gender Identity		
Women	395	35%
Men	692	61%
Genderqueer, agender, or write-in	15	1%
Not Answered	28	2%
Race/Ethnicity		
Asian	315	28%
Black or African American	23	2%
Hispanic, Latino/Latina/Latinx, or Spanish	36	3%
Middle Eastern or North African	44	4%
Native Hawaiian or Other Pacific Islander	1	0%
White	581	51%
Multiple Options Selected	90	8%
Another race or ethnicity not listed above	10	1%
Not Answered	33	3%
International Status		
Domestic	718	63%
International	405	36%
Not Answered	10	1%

Participants. A national, stratified sample of U.S. graduate students were asked to participate in a study about their experiences, identities, and motivations during the 2017-2018 school year. Engineering departments were sampled by size and location to create a representative sample from approximately 120 institutions, with a final sample of 1754 students. Because Master's and doctoral programs differ in a number of ways, such as time to degree, financial costs, and research focus, only students enrolled in doctoral programs were used in the current analysis (*n* = 1133).

Measures. To measure graduate engineering identity, three scales refined for use with graduate engineering students were used to assess graduate engineering identity (for more information about the scale validation see [10]. Participants were asked to rate their agreement/disagreement on a 5-point Likert scale regarding their recognition, performance/competence, and interest as engineers, scientists, and researchers. Information about students' degree progress, race/ethnicity, international status, gender identity, and discipline were used in this analysis (refer to Table 1 for information about sample demographics). Full details regarding this aspect of data collection are included in Appendix A. Given the small numbers in some cells, students' race/ethnicities were grouped into three categories for analysis: White (n = 581), Asian (n = 315), Bi/Multiracial (n = 90), and Traditionally Underserved Students of Color (n = 114). Similarly, participants who identified as agender, genderqueer, and women were grouped based on their shared experiences with gender marginalization in engineering spaces (n = 413).

A write-in option was provided for students' majors, and 34 majors were identified. Similar but distinct majors were combined into larger, commonly reported disciplines as per [22]. The most frequently occurring majors were mechanical, electrical, material science, chemical, and civil engineering (full list of majors included in Appendix A). Lastly, two items were used to evaluate students' progress and intentions to persist. The first was drawn from interview data in a previous qualitative study in which students discussed the absence of clear goalposts and milestones in graduate engineering education ("I find it difficult to evaluate my degree progress" [10]). The second was rooted in standard measures of intentions to persist ("I intend to complete my graduate degree"). The second item was reverse-coded, and both were averaged to create the degree uncertainty variable. These items capture participants' perceptions of uncertainty and persistence, but are not direct measures of persistence outcomes. Intentions are central determinants of behavior, and they tend to be meaningfully associated with what people ultimately do although they are not perfect predictors [23], [24]. Accordingly, the results presented here should be interpreted as associations with persistence-related intentions and uncertainty, not as evidence of actual attrition.

Analysis

Less than 1% of the data for items used in this analysis were missing and Little's MCAR test was insignificant, so participants with missing data (n = 13) were dropped per recommendations from the literature [25], [26]. To screen for outliers, participant responses to the 9 identity scores (recognition, performance/competence, and interest across the science, engineering, and research domains) were checked using Mahalanobis' distance [26] and 27 participants were dropped.

To answer RQ1 we used latent profile analysis and the R package **mclust** [27]. For an exploratory approach to LPA, it's recommended to run nine models, each with an increasing number of latent classes [28], [29]. The fit statistics used in model evaluations vary across studies; for our data and our sample size, we gave greater weight to more conservative statistics (BIC, CAIC, ICL) to avoid the formation of spurious groups. As recommended in the literature [30], we primarily used fit statistics to eliminate models from consideration. The resulting models were graphed and analyzed visually for parsimony and theoretical fit until the best model was identified and chosen.

To answer RQ2 and RQ3 we used the LPA results in conjunction with chi-square and regression analysis. A group with a significant chi-square test and a standardized residual > |2| was considered meaningfully over- or under-represented in a particular profile. Hierarchical regression analysis was used to examine whether GEI significantly predicts degree uncertainty.

Students who had completed their dissertation and were essentially finished with their degree (n = 55) were dropped from the sample for this analysis. For all regression models, we checked VIF scores and examined diagnostics (non-normality of residuals, homoscedasticity, scale-location, and residuals vs. leverage plots) to confirm that regression assumptions were met.

Results

RQ1: Latent Profile Analysis. A more in-depth discussion of the analysis process is included in Appendix B. After a review of entropy scores, fit indices, and visual analysis of the 5- and 6-profile solutions, the 6-profile solution was chosen as having the best theoretical fit. Labels for the identified profiles were developed and are explained below, and Figure 1 shows the distribution of scores in each profile (the means and standard deviations are included in Appendix B).

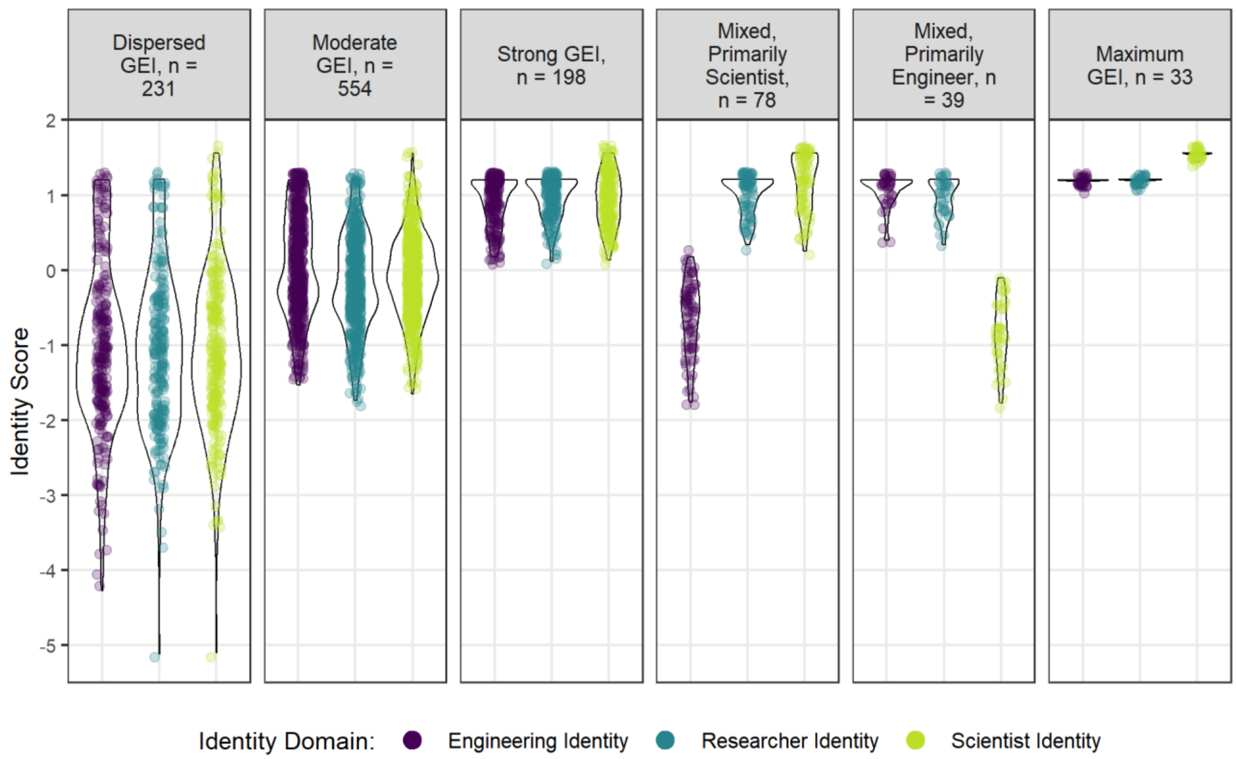

Figure 1. Violin scatterplots of identified latent profiles. Black outlines indicate the shape of the distribution. Identity scores are standardized, so a value of zero is average and the standard deviation is one. Scatterplots are jittered to aid in data visualization.

- Dispersed GEI (Profile 1, n = 231). Students in this profile had a wide degree of variability in their identity scores, with all participants reporting at least one identity score more than 1 SD below the mean.
- Moderate GEI (Profile 2, n = 554). Moderately strong GEI characterizes students in this profile. Their engineering, researcher, and scientist identity scores are generally within 1 SD of the mean and are generally congruent across domains.

- Strong GEI (Profile 3, n = 198). Strong GEI characterizes students in this profile. Their engineering, researcher, and scientist identity scores are above the mean and are generally congruent across domains.
- Mixed, Primarily Scientist (Profile 4, n = 78). Students in this profile have above-average researcher and scientist identity scores but below-average engineering identity scores.
- Mixed, Primarily Engineer (Profile 5, n = 39). Students in this profile have above-average engineering and researcher identity scores but below-average scientist identity scores.
- Maximum GEI (Profile 6, n = 33). Students in this profile have the strongest possible engineering, researcher, and scientist identity scores. They are all at the top of the scales and are congruent across domains.

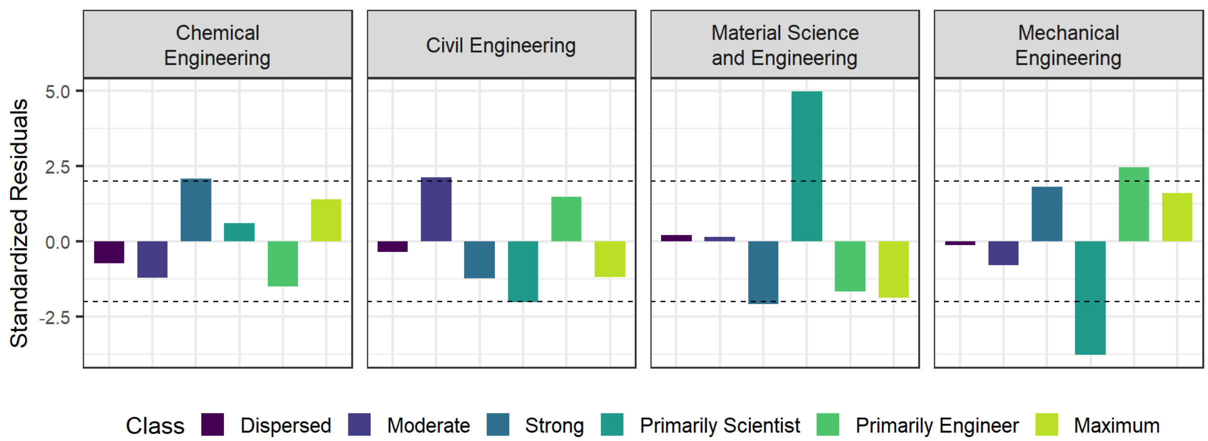

Figure 2. Plot of standardized residuals from chi-square test. Standardized residuals greater than |2| indicate meaningful over- or under-representation (depending on direction). For space, only majors that have meaningful under- or over-representation in one profile are plotted; full plots are included in Appendix B.

RQ2: Chi-Square Analyses. Five chi-square tests were used to test the distribution of a predictor (discipline, race, international status, gender, and degree progression) across the outcome (latent profile). The analyses for race ($p = .065$) and gender ($p = .348$) were not significant. We found significant differences in profile assignment by discipline, X^2 (df = 35) = 83.45, $p < .001$ (Figure 2).

International status also differed significantly by profile assignment, X^2 (df = 5) = 19.38, $p = .002$. International students were significantly overrepresented in the Dispersed GEI and Primarily Engineer profiles and significantly underrepresented in the Primarily Scientist profile. Degree progress also differed by profile assignment, X^2 (df = 15) = 28.68, $p = .018$ (Figure 3). Review of the standardized residuals suggests there is over- and underrepresentation across all four levels of progress in three of the six profiles.

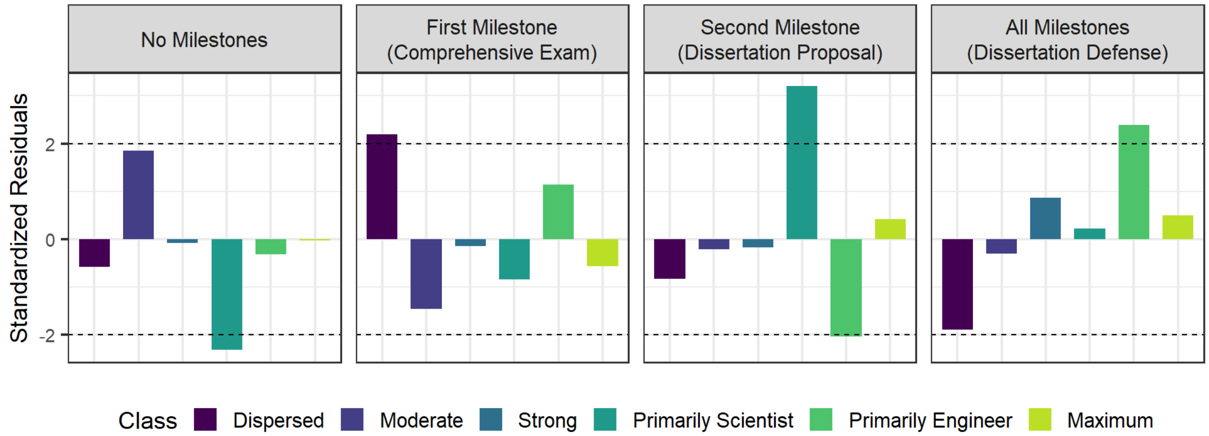

Figure 3. Plot of standardized residuals from chi-square test. Standardized residuals greater than |2| indicate meaningful over- or under-representation (depending on direction).

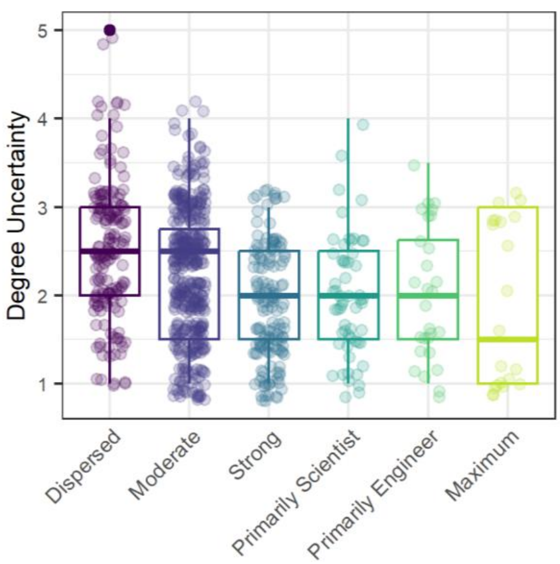

Figure 4. Degree uncertainty by assigned identity profile. Higher scores indicate greater uncertainty. Bold line indicates mean, boxes indicate quartiles, and solid dots indicate potential outliers.

RQ3: Regression Analyses. The first model confirmed the relationship between discipline, race, gender, degree progress, and degree uncertainty, F(13,809) = 3.33, $p < .001$, Adj. $R^2 = .04$ (Table 5). The second model examined the addition of profile assignment. The overall model was significant, F(18,804) = 7.62, $p < .001$, Adj. $R^2 = .13$, as was the increase in fit, F(5,804) = 17.89, $p < .001$. In both models, Bi/Multiracial students reported more uncertainty than White students; international students reported less uncertainty than domestic students; and students with one milestone completed reported more uncertainty than with none. Students in the Dispersed profile reported significantly more uncertainty than students in the Moderate profile (p < .001), while all other profiles except the Primarily Engineer profile reported less uncertainty (all ps < .025). Figure 5 shows degree uncertainty plotted by GEI profile.

Discussion

To briefly summarize the results discussed above: latent profile analysis of students' graduate engineering identities uncovered six profiles. Chi-square analyses found some uneven distributions by major, international status, and degree progress, and regression analyses found that race, international status, degree progress, and GEI profile were associated with degree uncertainty. In this section, we will review the results in more detail and connect them to the wider body of research, before providing recommendations for future research regarding engineering graduate students and engineering identity.

Table 5. Regression analysis results.

	B	SE	t	p	B	SE	t	p
Intercept *	2.14	0.08	25.79	<0.001	2.22	0.08	26.71	<0.001
Profile Assignment (0 = Moderate)								
Dispersed *	-	-	-	-	0.31	0.06	4.77	<0.001
Strong *	-	-	-	-	-0.37	0.07	-5.38	<0.001
Mixed, Primarily Scientist *	-	-	-	-	-0.35	0.10	-3.48	0.001
Mixed, Primarily Engineer	-	-	-	-	-0.20	0.14	-1.47	0.142
Maximum *	-	-	-	-	-0.35	0.16	-2.27	0.024
Major (0 = Mechanical)								
Bio	0.05	0.10	0.50	0.615	0.07	0.09	0.71	0.477
Chemical	0.09	0.09	0.98	0.328	0.11	0.09	1.24	0.216
Civil	0.01	0.10	0.14	0.888	-0.01	0.10	-0.11	0.911
CSE	0.14	0.10	1.33	0.185	0.12	0.10	1.25	0.210
Elec	0.03	0.09	0.29	0.772	0.01	0.09	0.06	0.951
MSE	0.16	0.09	1.74	0.083	0.14	0.09	1.50	0.135
Gender (0 = Men)								
Women & Gender Minorities	0.00	0.05	-0.06	0.949	-0.01	0.05	-0.27	0.784
Race (0 = White)								
Asian	0.04	0.08	0.53	0.594	-0.01	0.07	-0.18	0.856
Bi/Multiracial *	0.27	0.10	2.61	0.009	0.28	0.10	2.86	0.004
Underrepresented Minorities	0.06	0.09	0.64	0.520	0.04	0.09	0.40	0.689
International Status (0 = Domestic)								
International *	-0.25	0.07	-3.38	0.001	-0.25	0.07	-3.66	<0.001
Degree Progress (0 = No Milestones)								
Comprehensive Exam *	0.15	0.06	2.50	0.013	0.14	0.06	2.45	0.015
Dissertation Proposal	-0.05	0.07	-0.75	0.451	-0.04	0.06	-0.56	0.574

* $p < .05$

RQ1: Latent Profiles in Graduate Engineering Identity. Examining graduate engineering identity through the lens of latent profile analysis reveals some interesting patterns. Most immediately noticeable is the large group of students (48%) with GEI scores within one standard deviation of the mean (i.e., the Moderate GEI profile). This suggests that the normative experience for graduate engineering students is one of average identity across all three domains. We can compare students in this group to students in the Strong and Maximum GEI profiles and see that above-average identification across all three domains is comparatively rare. Students in the Strong GEI profile (n = 198; 17%) have above-average scores in all three identity domains, while students in the Maximum GEI profile (n = 33; 3%) have domain scores clustered around the ceiling. Outside of the Maximum GEI group, most students have some variability in their scores across the three domains (i.e., they are higher in some domains than others).

This variability is strongest for students in the Dispersed GEI profile (n = 231; 20%). Scores for students in this group range from 4 standard deviations below the mean to 1.5 above the mean. While the average means for this profile (Table 3) are below the average mean for the entire sample, the scores vary widely across the three domains. The high standard deviations and the absence of a uniformly 'Low GEI' group (to mirror the Strong and Maximum GEI profiles) suggests that there are few students who experience weak identification across all the domains, but a substantial number who experience weak identification in at least one domain.

Also noteworthy are the mixed identity patterns in the Primarily Scientist (PS) and Primarily Engineer (PE) profiles. Although these profiles together make up a small percentage of students (7% and 3%, respectively) they highlight the unique ways in which identification across the three domains can emerge. The chi-square results discussed below suggests that these unique combinations of identity are associated with disciplinary culture and international status, and the regression results suggest that – despite comparatively low scores in the engineering or scientist domains – these students report relatively low degree uncertainty. However, due to the cross-sectional nature of our dataset, it remains unclear whether these differences are a product of disciplinary culture, or whether students choose different disciplines based on their entering GEI.

Taken together, these profiles suggest that graduate engineering identity is not defined solely by overall strength of identification, but by the patterning and alignment of science, engineering, and research identities within individuals. This shifts the focus from asking whether graduate students "have" engineering identity to examining how multiple professional identities are integrated or negotiated. Notably, later findings indicate that degree uncertainty is generally similar across several profiles, raising the possibility that different identity configurations may be adaptive in different disciplinary or institutional spaces. This underscores the need for discipline-specific and context-sensitive research to better understand the diversity of engineering identities present across subfields, rather than assuming a one-size-fits-all developmental trajectory.

RQ2: Demographics and GEI Profiles. Chi-square tests were used to examine the distribution of race, gender, discipline, and degree progress across the uncovered GEI profiles. Firstly, the distribution of students across the eight disciplines and six GEI profiles was not even. These findings suggest that involvement in different disciplines creates different 'balances' across the identity subdomains, which is consistent with other work that emphasizes differences in disciplinary culture [31], [32], [33], [34]. Since our data is cross-sectional, it's not possible to determine the direction of this effect. However, these results do suggest that the combination of identity domains is likely to be relevant to students' sense of belonging in their disciplines and programs.

Secondly, international status is also unevenly distributed across the GEI profiles. Previous research suggests that international students face language barriers, problems with finances, identity struggles, and high expectations when attending school in the U.S. [35], [36]. These stressors may lead international students to be less fixed in their roles as scientists, engineers, and researchers more than domestic students. Lastly, degree progress stage is also unevenly distributed across the GEI profiles. These findings suggest that students may have different identity needs as they progress through their programs. Facilitating engineering graduate students' scientist identities may be adaptive in earlier phases of their programs where they are 'learning the ropes' of independent research; however, the strong engineering identities of all milestone students suggests that engineering identity shouldn't be neglected.

RQ3: Demographics, GEI Profiles, and Degree Uncertainty. Hierarchical linear regression was used to examine the relationship between identity and degree uncertainty among graduate engineering students. There was an effect of race/ethnicity on degree uncertainty, with Bi/Multiracial students reporting significantly more uncertainty than White students, but no other differences. There is an extended body of literature that discusses racial discrimination and attrition in engineering, both at the undergraduate [37] and graduate [18], [38], [39] levels, but it's often unclear whether these samples include Bi/Multiracial students. Our findings may be an effect of monoracism, which includes traditional notions of racism aimed at specific groups, but also discrimination against individuals due to a lack of fit into discrete racial groupings [40]. However, this interpretation should be considered a hypothesis rather than a conclusion supported directly by the current data.

The absence of significant differences for underrepresented racial minority students should likewise be interpreted cautiously. It may reflect similarities in reported degree uncertainty within this sample, but alternative explanations – such as a filter effect or unmeasured factors – are also plausible. Importantly, because multiple racial and ethnic identities were aggregated into broader categories for analytic purposes, these findings may obscure meaningful within-group differences. This aggregation limits the specificity of demographic interpretations and underscores the need for future work with larger samples that can examine identity experiences with greater nuance.

There was also an effect of degree process. Students who had completed their comprehensive exam (first milestone) reported more uncertainty than students with no milestones. As the data used in this analysis is cross-sectional, we cannot make any firm statements about causality. However, these findings do suggest that students at different phases of their degree program experience different amounts of uncertainty and that the completion of more milestones may paradoxically increase uncertainty. Anticipating and providing timely support in the face of this expected uncertainty may be one way that faculty and administrators can help retain students. Similarly, international students also reported significantly less uncertainty than domestic students, which may be because the substantial commitment required to study abroad acts as a filter, with less-certain students choosing to pursue other opportunities.

These findings also suggest that students with more positive GEI experience less degree uncertainty, and are consistent with previous findings regarding engineering identity and persistence [6], [41], [42]. Students in the Dispersed GEI profile, which is characterized by high variability between the domains and low mean identity scores, were the only ones to report more uncertainty than the Moderate GEI group. Despite the differences in GEI across the profiles, such as uniformly high scores in the Maximum GEI profile or starkly higher scientist/researcher scores in the Primarily Scientist profile, the effects of the Strong GEI, PS, and Maximum GEI profile on degree uncertainty are similar ($-.34 < \beta s > -.36$). However, there is also no significant relationship between assignment to the Primarily Engineering profile and degree uncertainty. This suggests that strongly identifying as an engineer in graduate school does not translate to greater degree certainty, and that strongly identifying as a scientist, or identifying equally to multiple subdomains, is more beneficial. Contrary though it may seem, facilitating students' identification with the science and engineering domains may be more effective in preventing attrition than focusing on engineering identity alone.

Limitations. Several limitations should be considered when interpreting these findings. First, latent profile analysis (LPA) is a relatively novel, data-driven approach that lacks traditional measures of statistical significance or effect size, which complicates interpretation and generalization. The extent to which these profiles replicate in other samples of engineering graduate students – or engineering students more broadly – remains to be established, and the risk of statistical or interpretive overfitting warrants a more confirmatory follow-up. Additionally, students from diverse racial and ethnic backgrounds were combined into a single group of "Traditionally Underserved Students of Color." Although this approach acknowledged structural inequities while maintaining adequate cell sizes, it necessarily obscures within-group heterogeneity. Similarly, genderqueer and agender students were grouped with women for analytic purposes due to small sample sizes, which limits conclusions about the distinct experiences of gender-diverse students in engineering.

Acknowledgements. This work was supported by the National Science Foundation (EHR-1535453 and 1535254). The authors wish to thank the PRiDE research group and the North Carolina State research groups for their assistance in data collection and entry. Additionally, the authors thank the participants of this study for their openness in quantitative responses.

References

[1] J. E. Stets and R. T. Serpe, "Identity theory," in *Handbook of social psychology, 2nd ed*, in Handbooks of sociology and social research. , New York, NY, US: Springer Science + Business Media, 2013, pp. 31–60. doi: 10.1007/978-94-007-6772-0_2.

[2] J. P. Gee, "Identity as an Analytic Lens for Research in Education," *Rev. Res. Educ.*, vol. 25, pp. 99–125, 2000, doi: 10.2307/1167322.

[3] A. Godwin, G. Potvin, Z. Hazari, and R. Lock, "Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice," *J. Eng. Educ.*, vol. 105, no. 2, pp. 312–340, Apr. 2016, doi: 10.1002/jee.20118.

[4] Z. Hazari, G. Sonnert, P. M. Sadler, and M.-C. Shanahan, "Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study," *J. Res. Sci. Teach.*, p. n/a-n/a, 2010, doi: 10.1002/tea.20363.

[5] A. Godwin and A. Kirn, "Identity-based motivation: Connections between first-year students' engineering role identities and future-time perspectives," *J. Eng. Educ.*, vol. 109, no. 3, pp. 362–383, 2020, doi: 10.1002/jee.20324.

[6] B. Hughes, W. Schell, B. Tallman, R. Beigel, E. Annand, and M. Kwapisz, "Do I Think I'm an Engineer? Understanding the Impact of Engineering Identity on Retention," in *2019 ASEE Annual Conference & Exposition Proceedings*, Tampa, Florida: ASEE Conferences, Jun. 2019, p. 32674. doi: 10.18260/1-2--32674.

[7] H. M. Matusovich, R. A. Streveler, and R. L. Miller, "Why Do Students Choose Engineering? A Qualitative, Longitudinal Investigation of Students' Motivational Values," *J. Eng. Educ.*, vol. 99, no. 4, pp. 289–303, 2010, doi: 10.1002/j.2168-9830.2010.tb01064.x.

[8] O. Pierrakos, T. K. Beam, J. Constantz, A. Johri, and R. Anderson, "On the development of a professional identity: engineering persisters vs engineering switchers," in *2009 39th IEEE Frontiers in Education Conference*, Oct. 2009, pp. 1–6. doi: 10.1109/FIE.2009.5350571.

[9] A. Elizondo-Noriega, N. Tiruvengadam, K. Cargile-Cook, J. Cross, and M. G. Beruvides, "Understanding Engineering Identity Formation Mechanisms in Graduate and Undergraduate Education: A State-of-the-Art Matrix Analysis," *IIE Annu. Conf. Proc.*, pp. 1–6, 2020.

[10] H. Perkins, M. Bahnson, C. Cass, M. A. Tsugawa-Nieves, and A. Kirn, "Development and testing of an instrument to understand engineering doctoral students' identities and motivations," in *2018 ASEE Annual Conference & Exposition*, 2018. Accessed: May 12, 2024. [Online]. Available: https://peer.asee.org/30319.pdf

[11] M. Bahnson *et al.*, "Inequity in graduate engineering identity: Disciplinary differences and opportunity structures," *J. Eng. Educ.*, vol. 110, no. 4, pp. 949–976, 2021, doi: 10.1002/jee.20427.

[12] H. Perkins, M. A. Tsugawa-Nieves, J. N. Chestnut, B. Miller, A. Kirn, and C. Cass, "The Role of Engineering Identity in Engineering Doctoral Students' Experiences," presented at the 2017 ASEE Annual Conference & Exposition, Jun. 2017. Accessed: May 12, 2024. [Online]. Available: https://peer.asee.org/the-role-of-engineering-identity-in-engineering-doctoral-students-experiences

[13] S. Chen, "Balancing knowing and not-knowing: an exploration of doctoral candidates' performance of researcher selves in the dissertation defence," *Assess. Eval. High. Educ.*, vol. 39, no. 3, pp. 364–379, Apr. 2014, doi: 10.1080/02602938.2013.834876.

[14] C. L. Colbeck, "Professional identity development theory and doctoral education," *New Dir. Teach. Learn.*, vol. 2008, no. 113, pp. 9–16, Mar. 2008, doi: 10.1002/tl.304.
[15] M. Bahnson, H. Perkins, M. Tsugawa, A. Kirn, and C. Cass, "Influence of Research Experience on Recognition and Identity Development in the Engineering Graduate Student Population," in *2018 IEEE Frontiers in Education Conference (FIE)*, Oct. 2018, pp. 1–4. doi: 10.1109/FIE.2018.8658764.
[16] M. Bahnson *et al.*, "Variance in Engineering Identity in Master's Degree-Seeking Engineering Students," in *2019 IEEE Frontiers in Education Conference (FIE)*, Oct. 2019, pp. 1–7. doi: 10.1109/FIE43999.2019.9028414.
[17] N. H. Choe and M. Borrego, "Prediction of Engineering Identity in Engineering Graduate Students," *IEEE Trans. Educ.*, vol. 62, no. 3, pp. 181–187, Aug. 2019, doi: 10.1109/TE.2019.2901777.
[18] M. Bahnson and C. G. P. Berdanier, "Current Trends in Attrition Considerations of Graduate Engineering Students in the United States," *Int. J. Eng. Educ.*, vol. 39, no. 1, Jan. 2023, Accessed: May 13, 2024. [Online]. Available: https://par.nsf.gov/biblio/10464171-current-trends-attrition-considerations-graduate-engineering-students-united-states
[19] M. Bahnson, G. Sallai, K. Jwa, and C. Berdanier, "Mitigating Ceiling Effects in a Longitudinal Study of Doctoral Engineering Student Stress and Persistence," *Int. J. Dr. Stud.*, vol. 18, pp. 199–227, 2023, doi: 10.28945/5118.
[20] C. G. P. Berdanier, C. Whitehair, A. Kirn, and D. Satterfield, "Analysis of social media forums to elicit narratives of graduate engineering student attrition," *J. Eng. Educ.*, vol. 109, no. 1, pp. 125–147, 2020, doi: 10.1002/jee.20299.
[21] Z. K. Collier and W. L. Leite, "A Comparison of Three-Step Approaches for Auxiliary Variables in Latent Class and Latent Profile Analysis," *Struct. Equ. Model. Multidiscip. J.*, vol. 24, no. 6, pp. 819–830, Nov. 2017, doi: 10.1080/10705511.2017.1365304.
[22] J. Roy, "Engineering by the Numbers," ASEE, 2018.
[23] C. J. Armitage and M. Conner, "Efficacy of the Theory of Planned Behaviour: a meta-analytic review," *Br. J. Soc. Psychol.*, vol. 40, no. Pt 4, pp. 471–499, Dec. 2001, doi: 10.1348/014466601164939.
[24] P. Sheeran and T. L. Webb, "The Intention–Behavior Gap," *Soc. Personal. Psychol. Compass*, vol. 10, no. 9, pp. 503–518, Sep. 2016, doi: 10.1111/spc3.12265.
[25] H. Perkins, M. Bahnson, M. Tsugawa-Nieves, A. Kirn, and C. Cass, "Making sense of missingness: Analyzing skipped items in engineering education research," 2019.
[26] B. G. Tabachnick and L. S. Fidell, *Using multivariate statistics*, 7th ed. in Using multivariate statistics. Boston, MA: Allyn & Bacon/Pearson Education, 2021, pp. xxvii, 980.
[27] L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery, "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models," *R J.*, vol. 8, no. 1, pp. 289–317, Aug. 2016.
[28] D. Oberski, "Mixture Models: Latent Profile and Latent Class Analysis," in *Modern Statistical Methods for HCI*, J. Robertson and M. Kaptein, Eds., in Human–Computer Interaction Series. , Cham: Springer International Publishing, 2016, pp. 275–287. doi: 10.1007/978-3-319-26633-6_12.
[29] D. Spurk, A. Hirschi, M. Wang, D. Valero, and S. Kauffeld, "Latent profile analysis: A review and 'how to' guide of its application within vocational behavior research," *J. Vocat. Behav.*, vol. 120, p. 103445, Aug. 2020, doi: 10.1016/j.jvb.2020.103445.

[30] J. J. Dziak, D. L. Coffman, S. T. Lanza, R. Li, and L. S. Jermiin, “Sensitivity and specificity of information criteria,” *Brief. Bioinform.*, vol. 21, no. 2, pp. 553–565, Mar. 2020, doi: 10.1093/bib/bbz016.

[31] N. E. Canney and A. R. Bielefeldt, “Differences in Engineering Students’ Views of Social Responsibility between Disciplines,” *J. Prof. Issues Eng. Educ. Pract.*, vol. 141, no. 4, p. 04015004, Oct. 2015, doi: 10.1061/(ASCE)EI.1943-5541.0000248.

[32] A. Danowitz and K. Beddoes, “Characterizing Mental Health and Wellness in Students Across Engineering Disciplines,” in *2018 CoNECD - The Collaborative Network for Engineering and Computing Diversity Conference Proceedings*, Crystal City, Virginia: ASEE Conferences, Apr. 2018, p. 29522. doi: 10.18260/1-2--29522.

[33] A. Kirn, B. Morkos, and L. Benson, “Work in progress: How differences in student motivation characterize differences between engineering disciplines,” in *2012 Frontiers in Education Conference Proceedings*, Seattle, WA, USA: IEEE, Oct. 2012, pp. 1–2. doi: 10.1109/FIE.2012.6462358.

[34] P. J. Palazolo, S. Ivey, and C. V. Camp, “Freshman Engineering Student Perceptions of Engineering Disciplines,” 2010.

[35] T. M. Bluestein, C. T. Amelink, and M. S. Artiles, “Campus Climate for Engineering Graduate Students: Examining Differences Between Domestic Minority, Domestic Majority, and International Students,” presented at the 2018 CoNECD - The Collaborative Network for Engineering and Computing Diversity Conference, Apr. 2018. Accessed: May 26, 2024. [Online]. Available: https://peer.asee.org/campus-climate-for-engineering-graduate-students-examining-differences-between-domestic-minority-domestic-majority-and-international-students

[36] C. Gautam, C. L. Lowery, C. Mays, and D. Durant, “Challenges for Global Learners: A Qualitative Study of the Concerns and Difficulties of International Students,” *J. Int. Stud.*, vol. 6, no. 2, pp. 501–526, 2016.

[37] B. Geisinger and D. Raman, “Why They Leave: Understanding Student Attrition from Engineering Majors,” *Int. J. Eng. Educ.*, vol. 29, pp. 914–925, Jan. 2013.

[38] E. O. McGee, P. K. Botchway, D. E. Naphan-Kingery, A. J. Brockman, S. Houston II, and D. T. White, “Racism camouflaged as impostorism and the impact on Black STEM doctoral students,” *Race Ethn. Educ.*, vol. 25, no. 4, pp. 487–507, Jun. 2022, doi: 10.1080/13613324.2021.1924137.

[39] E. O. McGee, D. M. Griffith, and S. L. Houston, “‘I Know I Have to Work Twice as Hard and Hope that Makes Me Good Enough’: Exploring the Stress and Strain of Black Doctoral Students in Engineering and Computing,” *Teach. Coll. Rec.*, vol. 121, no. 4, pp. 1–38, Apr. 2019, doi: 10.1177/016146811912100407.

[40] M. P. Johnston and K. L. Nadal, “Multiracial microaggressions: Exposing monoracism in everyday life and clinical practice,” in *Microaggressions and marginality: Manifestation, dynamics, and impact*, Hoboken, NJ, US: John Wiley & Sons, Inc., 2010, pp. 123–144.

[41] B. D. Jones, C. Ruff, and M. C. Paretti, “The impact of engineering identification and stereotypes on undergraduate women’s achievement and persistence in engineering,” *Soc. Psychol. Educ.*, vol. 16, no. 3, pp. 471–493, Sep. 2013, doi: 10.1007/s11218-013-9222-x.

[42] D. Verdín, “The power of interest: minoritized women’s interest in engineering fosters persistence beliefs beyond belongingness and engineering identity,” *Int. J. STEM Educ.*, vol. 8, no. 1, p. 33, May 2021, doi: 10.1186/s40594-021-00292-1.

Appendix A

Measures

To measure graduate engineering identity, three scales refined for use with graduate engineering students were used to assess graduate engineering identity (for more information about the scale validation see [10]). Participants were asked to rate their agreement/disagreement on a 5-point Likert scale regarding their recognition ("My parents see me as an engineer"), performance/competence ("I understand concepts I have studied in engineering"), and interest ("I enjoy learning engineering") as engineers, scientists, and researchers. This resulted in nine scores across three domains and three sub-constructs. The three subscales for each identity domain were combined into a single identity score for each domain. The scientist, engineering, and research subdomain items had excellent reliability ($\alpha = .91$, $.94$, and $.93$ respectively) and consisted of 14, 17, and 15 items each.

For this analysis, items that asked about students' degree progress, race/ethnicity, international status, gender identity, and discipline were used. Degree progress was measured by asking students to select all the degree milestones they had completed (comprehensive exam, dissertation proposal, and dissertation defense). For race/ethnicity, students were asked to indicate which racial and/or ethnic groups they identified with, and multiple selections were possible (refer to Table 1 for the full list of options). Write-in responses were recoded into existing categories where possible (refer to Table 1 for the full overview). Given the small numbers in some cells, students' race/ethnicities were grouped into three categories: White ($n = 846$), Asian ($n = 482$), Bi/Multiracial ($n = 123$), and Traditionally Underserved Students of Color ($n = 158$).

A similar measure was used for gender identity, also allowing multiple selections (again, Table 1 reports the full list of options). Of the 1754 participants who remained in the dataset after screening for missingness and outliers, 1586 participants identified as male or female (this includes 56 students who selected 'Cis' and 2 students who selected 'Trans' alongside one of the binary options), 13 identified as agender or genderqueer, 18 provided a written response, and 134 did not respond. Because there were so few participants identified as agender or genderqueer, they were combined with cis- and trans-women to create a variable for women and gender minorities ($n = 596$).

All students were also asked to write-in their country of origin, and 79 unique responses were identified (including those who did not respond, $n = 209$). 'USA' was the most common response ($n = 927$), followed by China ($n = 173$), India ($n = 143$), Iran ($n = 48$), and Bangladesh ($n = 23$). Approximately 33% of students ($n = 573$) identified as international.

As most degree programs require students to complete their comprehensive exam first, dissertation proposal second, and dissertation defense third, participants' degree progress was coded according to this highest milestone completed. Lastly, given the different demands of doctoral and master's programs, we limited our analysis to students who indicated they were pursuing a doctoral degree (n = 1173).

A write-in option was provided for students' majors, and 34 majors were identified, including an 'NA' category for students who did not respond. Similar but distinct majors were combined into larger, commonly reported disciplines as per Roy (2018). For instance, 'Biological Engineering' was added to a category with 'Biomedical Engineering' and coded as 'Bio Engineering (n = 133)' (refer to Table A1 for a full breakdown). Majors listed by 20 students or less that could not be combined into larger categories were recoded as 'Other Engineering' (n = 61).

Table A1: Graduate engineering majors

Reported Majors	**Count**	**Percentage**
Aerospace Engineering	50	3%
Agricultural & Biological Engineering	47	3%
Biological or Biomedical Engineering	133	8%
Chemical Engineering	190	11%
Civil Engineering	160	9%
Computer Engineering	33	2%
Computer Science	93	5%
Computer Science and Engineering	25	1%
Electrical & Computer Engineering	64	4%
Electrical Engineering	202	12%
Environmental Engineering	52	3%
Industrial Engineering	40	2%
Material Science and Engineering	185	11%
Mechanical Engineering	260	15%
Nuclear Engineering	59	3%
Other Engineering	61	3%
Not Answered	100	6%

Appendix B

Latent Profile Analysis. To answer RQ1 (do graduate students' engineering identities form latent profiles, and what do these profiles communicate about the heterogeneity in graduate engineering identity?) we used latent profile analysis. A form of cluster analysis, mixture modeling is used to identify hidden groups within a population through analysis of multivariate data. By identifying these latent groups, researchers can take a bottom-up approach to analysis and theory-building that prioritizes participants' reports, voices, and experiences instead of researchers' preconceptions and potential biases [42], [43]. However, researchers still make decisions about how to weigh fit indices which solutions to retain. LPA is also highly sample dependent, and while efforts are made to avoid overfitting solutions to sample-specific noise, the possibility is always a concern. The final selection of a profile solution is therefore guided by both statistical criteria and theoretical judgment, and these limitations should be kept in mind.

As discussed in the Analysis section, we used the R package **mclust** to identify latent profiles among participants. This tool uses model-based clustering to predict the parameters of the hidden groups and calculate the posterior probability (the likelihood that a case will fit into each of the groups). These values are iteratively refined through an expectation maximization (EM) algorithm, until either the parameters stop changing or the model fails to converge [25].

Various fit statistics are produced, with the AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and BLRT (Bootstrapped Likelihood Ratio Test) almost always reported. AIC and BIC scores are evaluated in relation to each other, with lower AIC scores and higher BIC scores indicating a better fit. Other fit statistics commonly reported include the CAIC (Consistent Akaike Information Criterion) and the ICL (Integrated Complete-data Likelihood). The CAIC is derived from the AIC score mentioned above, but includes a more stringent penalty for classes and thus produces more parsimonious solutions. Similarly, the ICL is offered as a complement to the BIC score, with additional penalties that recommend models with fewer latent classes [44].

The fit statistics used in model evaluations vary across studies; for our data and our sample size, we gave greater weight to more conservative statistics (BIC, CAIC, ICL) to avoid the formation of spurious groups (i.e., small groups produced when noise is misinterpreted as signal). These scores are evaluated relative to each other, with larger BIC and ICL and smaller CAIC scores indicating better-performing models. In addition to the listed fit statistics, entropy is usually considered as well, although it is a measure of the average accuracy in assigning cases to profile and not a fit statistic per se.

Discussion of Results. A review of the entropy plot recommended a 5-profile solution, while the BLRT suggested that a model with 8 or fewer profiles would be best. A review of fit indices and entropy scores indicated that the 5- or 6-profile solutions would be the best fit for the data, as they balanced the lower BIC, CAIC, and ICL scores with higher entropy scores (refer to Table B1 for fit indices and other statistics).

Table B1: LPA Fit Indices & Statistics

	LL	AIC	CAIC	BIC	SABIC	ICL	BLRTS	BLRTS p	Entropy	Groups <5%
1 Profile	-4519.04	9056.07	9110.37	-9671.07	9072.78	-9671.07	-	-	1	0
2 Profile	-4285.58	8605.16	8707.71	-9248.77	8636.72	-9541	450.44	0.001	0.98	0
3 Profile	-4218.65	8487.3	8638.12	-9197.35	8533.71	-9535.1	79.55	0.001	0.84	0
4 Profile	-4128.84	8323.68	8522.75	-9047.04	8384.94	-9529.21	178.44	0.001	0.77	0
5 Profile	-4123.64	8295.29	8440.07	-8998.68	8339.84	-9456.41	76.49	0.001	0.85	1
6 Profile	-4099.49	8256.98	8431.93	-8974.44	8310.82	-9407.92	52.37	0.001	0.84	2
7 Profile	-4080.04	8228.07	8433.18	-8947.66	8291.19	-9448.65	54.91	0.001	0.79	2
8 Profile	-4045.21	8168.42	8403.69	-8926.07	8240.82	-9508.31	49.72	0.001	0.8	3
9 Profile	-4009.91	8181.82	8670.46	-8954.2	8332.18	-9648.95	-	-	0.76	2

The 5- and 6-profile solutions were chosen for further analysis to determine the best-fitting model. At this stage, the focus shifts from the comparison of fit indices and entropy scores to a visual analysis driven by theoretical understanding and the need for a parsimonious explanation of the data. Ultimately, the 6-profile solution was chosen. A hierarchical combination of the profiles (based on entropy criterion, according to the methodology described in [45]) suggests that when moving from the 5-profile to the 6-profile solutions, some participants are split from Profile 1 into a new profile (Profile 5). Profile 1 is the second-largest group in both solutions, with the lowest mean identity scores across all three domains. Profile 5 in the 6-profile solution is a small group (n = 39) characterized by a distinctive pattern of above-average engineering and researcher identity but below-average scientist identity. Given our study's focus on the combination of engineering, researcher, and scientist identities among graduate engineering students, we chose the less parsimonious but more descriptive model that offers additional insight into this small but theoretically interesting group of students. This also allows for comparison and contrast with students in Profile 4 (n = 78), who have above-average identity scores in the research and science domains but below-average scores in the engineering domain.

To better understand the patterns within the six profiles uncovered by the analysis, we examined the mean scores and standard deviations, density plots of the distributions, and scatterplots of individual scores across all identity domains for each uncovered profile (refer to Table B2). Based on the patterns of mean and variability, the profiles were labeled as discussed in the Results section of the paper.

Table B2: Means and standard deviations of identified profiles

Profile		**Means**			**Standard Deviations**	
	Scientist Identity	*Engineering Identity*	*Researcher Identity*	*Scientist Identity*	*Engineering Identity*	*Researcher Identity*
Dispersed GEI (n = 231)	-1.08	-1.06	-1.09	1.06	1.12	1.06
Moderate GEI (n = 554)	-0.08	0.06	-0.16	0.56	0.69	0.63
Strong GEI (n = 198)	0.94	0.90	0.95	0.36	0.31	0.27
Mixed, Primarily Scientist GEI (n = 78)	1.17	-0.64	0.99	0.37	0.49	0.26
Mixed, Primarily Engineer GEI (n = 39)	-0.84	1.06	0.98	0.46	0.21	0.25
Maximum GEI (n = 33)	1.54	1.19	1.20	0.04	0.03	0.03

Appendix C

Chi-Square Analyses. Five chi-square tests were used to test the distribution of a predictor (discipline, race, international status, gender, and degree progression) across the outcome (latent profile).

Discipline. For the discipline chi-square test, similar majors were combined to maintain adequate cell sizes:

- The ‘Other Engineering (n = 176)’ category included the miscellaneous disciplines described earlier as well as students in the Aerospace (n = 28), Environmental (n = 28), Industrial (n = 29), and Nuclear (n = 44) Engineering disciplines.
- Students in the Computer Engineering (n = 19), Computer Science (n = 55), and Computer Science and Engineering (n = 21) categories were combined into a single group (Computer Science and Engineering; n = 95).
- Students in the Agricultural and Biological Engineering (n = 27) were added to the Bio Engineering (n = 102) group (Bio Engineering; final n = 129).

This chi-square test was significant, X2 (df = 35) = 83.45, p < .001, although several cell sizes were small. Review of the standardized residuals suggests there is over- and underrepresentation among Chemical, Civil, Material Science, and Mechanical Engineering students in four of the six profiles (Figure C1).

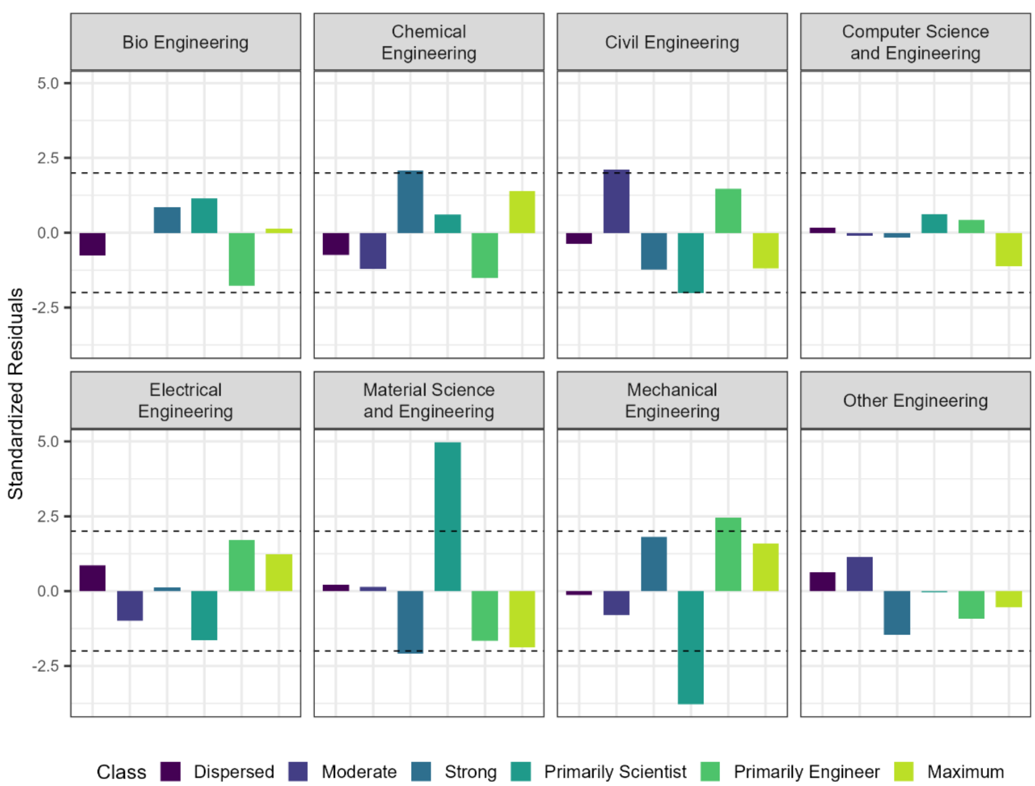

Figure C3. Plot of standardized residuals from chi-square test. Standardized residuals greater than |2| indicate meaningful over- or under-representation (depending on direction).

Chi-square tests were also used to examine the distribution of race and gender across the latent profiles. Race (p = .065; Figure C2) was not significant, although some of the standardized residuals were greater than two.

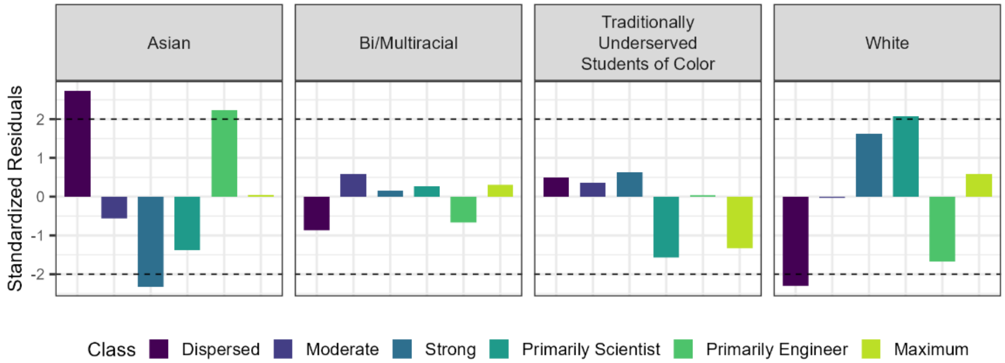

Figure C4. Plot of standardized residuals from chi-square test. Standardized residuals greater than |2| indicate meaningful over- or under-representation (depending on direction).

Considering these results, a follow-up chi-square test examining the relationship between race and major was also conducted, and was significant, X2 (df = 21) = 79.11, p < .001 (Figure C3). These results suggest that there is significant and meaningful racial under- and over-representation in different disciplines, and so differences in GEI according to discipline may be confounding the findings regarding race. A more in-depth, multivariate analysis would be needed to fully untangle these relationships.

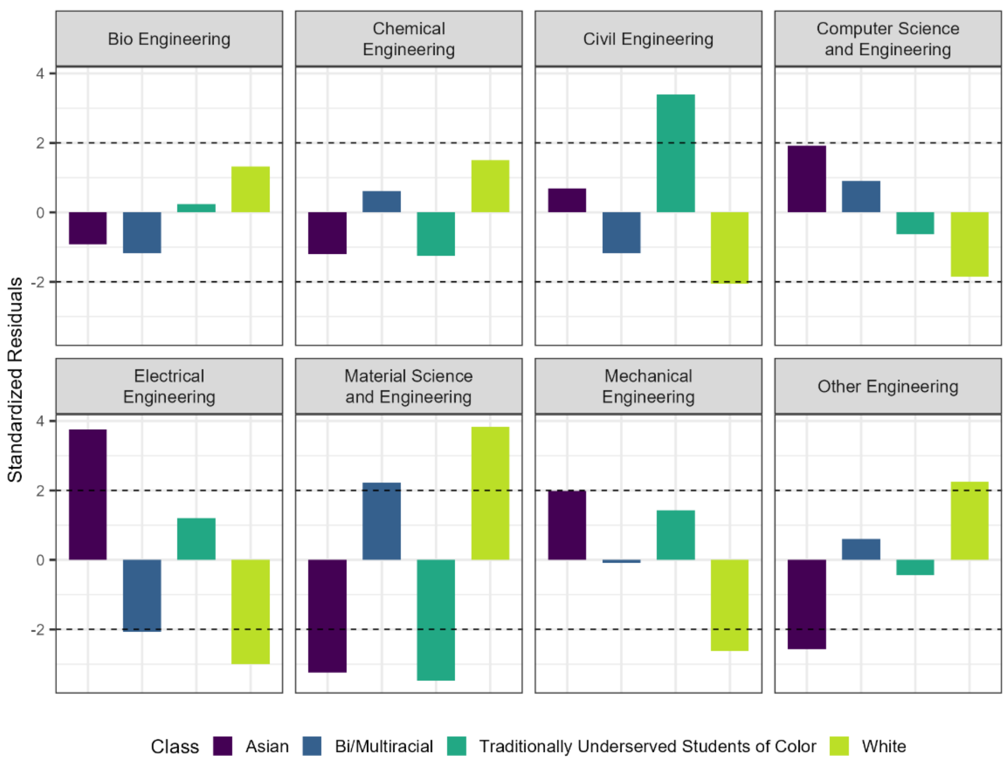

Figure C5. Plot of standardized residuals from chi-square test. Standardized residuals greater than |2| indicate meaningful over- or under-representation (depending on direction).

The chi-square test for gender was not significant (p = .348; Figure C4).

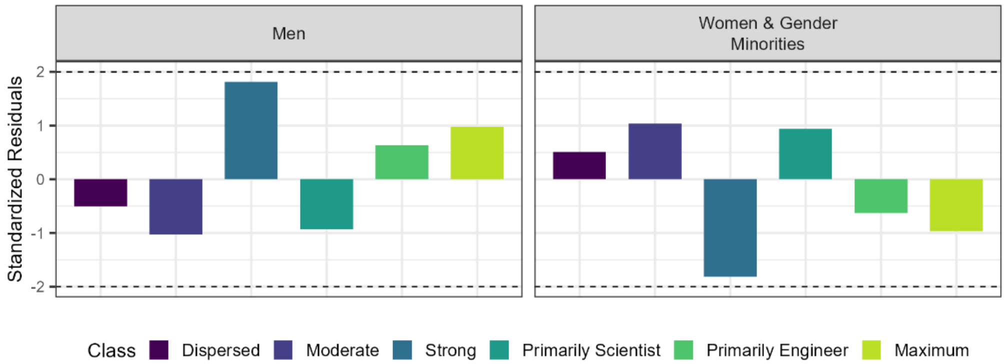

Figure C6. Plot of standardized residuals from chi-square test. Standardized residuals greater than |2| indicate meaningful over- or under-representation (depending on direction).

International status differed significantly by profile assignment, X^2 (df = 5) = 19.38, *p* = .002. International students were significantly overrepresented in the Disorganized GEI and primarily Engineer profiles, and significantly underrepresented in the Primarily Scientist profile (Figure C5).

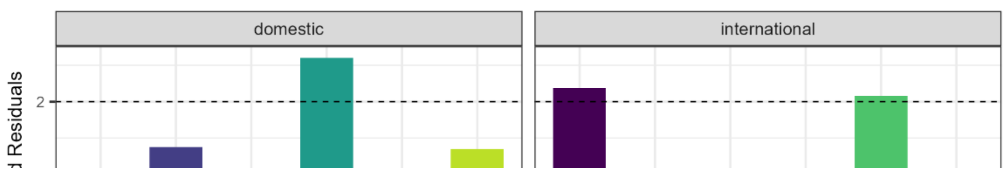

Figure C7. Plot of standardized residuals from chi square test. Standardized residuals greater than |2| indicate meaningful

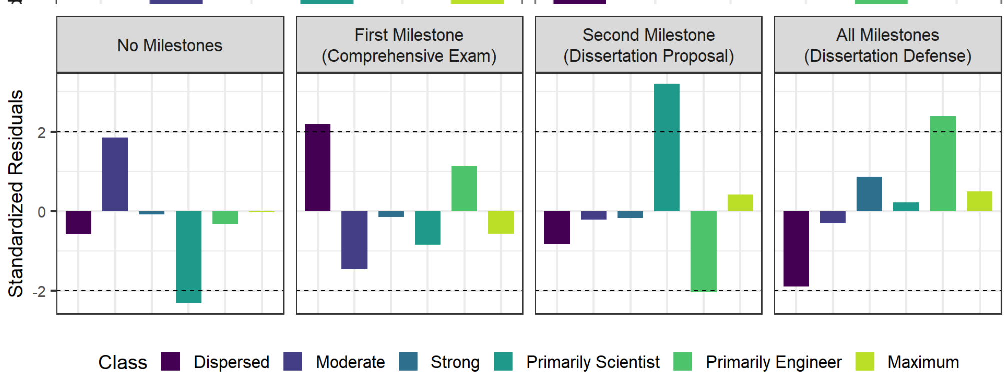

Figure C8. .Plot of standardized residuals from chi square test. Standardized residuals greater than |2| indicate meaningful over- or under-representation (depending on direction).

Degree progress also differed by profile assignment, X2 (df = 15) = 28.68, p = .018 (Figure C6). Review of the standardized residuals suggests there is over- and underrepresentation across all four levels of progress in three of the six profiles.