2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #52613

Reliability of PERMA-V(N): A Qualitative Coding Framework for Capturing Well-being Dimensions in Engineering Doctoral Students

Dr. Johan Tabora, Northwestern University

Johan Tabora is a Postdoctoral Scholar at Northwestern University, McCormick School of Engineering. He studies the intersection of race, power, and STEM education, and is interested in understanding how these interactions shape the experiences of underrepresented Asian American populations in STEM.

He completed his Ph.D. in science education at the Department of Curriculum and Instruction at the University of Illinois Chicago. He holds a Ed.M. in teacher education from the Harvard Graduate School of Education, a M.S. in teacher education from the Harvard Graduate School of Education, and a B.S. in aerospace engineering from the Florida Institute of Technology, Prior to doing a postdoc, Johan taught high school physics for 20 years in the Chicago Public Schools.

Kate Manion, Northwestern University
Dr. David P O'Neill, Northwestern University

David O'Neill is a Professor of Instruction and the Michael Jaharis Director of Experiential Learning for the Biomedical Engineering Department at Northwestern University. David read Engineering Science at University College, Oxford, receiving his M.Eng. and D.Phil. before undertaking a post-doc in the Department of Physiology, Anatomy and Genetics. During these years, he taught undergraduate tutorials for Keble, New, University, and Harris Manchester Colleges, was College Lecturer for New College and a Senior College Lecturer in Engineering Science for Keble College. He has interests in the formation of engineering identity, and increasing synopticism at a curricular level.

Reliability of PERMA-V(N): A Qualitative Coding Framework for Capturing Well-being Dimensions in Engineering Doctoral Students

Introduction

This methods full paper presents a reliability study for a modified well-being framework deployed to code interviews with engineering doctoral students (EDSs) during the peri-QE period—the weeks of heightened stress leading up to, and shortly after taking, qualifying examinations (QEs). The QE represents a critical milestone in engineering doctoral programs, serving to legitimize EDSs and admit them into doctoral candidacy [1]. However, research has shown that the purposes and formats of QEs remain ambiguous to students [1], [2], [3]. The combination of unclear expectations and high-stakes evaluation creates a particularly stressful experience for EDSs [4].

Well-being challenges among doctoral students exceed those of general graduate populations, with engineering students experiencing disproportionately high mental health difficulties—more than six times as likely to experience depression and anxiety as the general population [5], [6]. These well-being challenges directly constrain academic outcomes, with research demonstrating that well-being predicts research output, publication productivity, and self-efficacy [7], while poor mental health increases likelihood of research discontinuation[8]. Critically, well-being operates not merely as a mental health concern but as a foundational capacity that either facilitates or constrains the cognitive, social, and developmental processes essential for professional identity formation during the doctoral experience.

Current approaches to measuring doctoral student well-being have significant limitations. Traditional models focus primarily on mental health problems like depression and anxiety, missing the positive aspects of flourishing that are essential for professional growth [9]. Simple measures that look at just one dimension fail to capture how EDSs experience well-being, which changes dramatically from week to week—varying with assignment deadlines and shifting significantly across different phases of the semester [10], [11]. Academic milestones like QEs create predictable disruptions that require measurement tools sensitive to rapid change [12].

The PERMA-V framework offers a more comprehensive approach, measuring *Positive emotions*, *Engagement*, *Relationship*(s), *Meaning*, *Accomplishment*(s), and *Vitality* [13], [14]. Originally conceived as the PERMA framework by Seligman [9], this approach focuses on five pillars that contribute to an individual's "flourishing". Zhivotovskaya later expanded PERMA to include *Vitality*, creating the PERMA-V framework to more holistically capture a person's well-being [15]. Engineering education researchers have previously used the PERMA-V framework to explore aspects of well-being among EDSs [16].

However, our prior work revealed that the PERMA-V framework as typically used was unable to fully describe EDS experiences during the high-stress QE period, particularly those experiences that may be negative or emotionally neutral [17] . Critically, PERMA-V as a coding framework proved unable to distinguish between emotionally-neutral and negative aspects of the doctoral experience, suggesting the need for a framework capable of capturing how other elements of PERMA-V can be co-coded with positive emotions or opinions, with negative emotions or opinions, or with neither. This limitation is particularly problematic given the documented stress and negative emotional experiences associated with high stress settings such as QEs [18].

This paper presents PERMA-V(N), a qualitative coding framework that extends PERMA-V by adding a Negativity dimension for qualitative analysis of well-being in high-stress academic contexts and/or when positive and negative experiences co-occur. While developed and tested with EDSs during QEs, PERMA-V(N)'s constructs could apply to other high-stress contexts where systematic well-being assessment requires capturing both positive and negative dimensions simultaneously—including closing stages of doctoral work (dissertation writing and defense), medical licensing examinations, tenure reviews for faculty, or high-stakes professional certifications. This paper addresses two research questions to establish both the reliability and utility of PERMA-V(N):

RQ1: Can the framework achieve adequate inter-rater reliability when applied to EDS during QEs?

RQ2: What complexities of EDS well-being does PERMA-V(N) capture during QEs preparation?

Theoretical Framework

The development of PERMA-V(N) draws from contemporary well-being research recognizing that positive and negative affect represent distinct but co-occurring dimensions rather than mutually exclusive opposite poles of a single continuum [19], [20]. This theoretical foundation supports measuring both flourishing indicators (PERMA-V) and distress indicators (N) simultaneously, particularly during periods when students may experience research engagement alongside examination anxiety, or express negative psychological responses to, or assessments of, otherwise positive constructs such as engagement, relationships, meaning, accomplishment, or vitality. For example, students may describe high engagement with research while simultaneously expressing frustration about the engagement demands competing with examination preparation or report meaningful research relationships that create pressure and stress during evaluation periods.

PERMA-V(N) extends the established PERMA-V framework [9], [21], [22]. Second, *Negativity* (N) is introduced to capture negative psychological states and circumstantial assessments. These modifications recognize that psychological experiences manifest across emotional, evaluative, and circumstantial domains—particularly during high-stress periods when flourishing and distress co-occur.

This framework addresses three limitations in existing doctoral student assessment approaches: (1) deficit-focused measures that miss resilience and growth, (2) positive psychology measures that inadequately capture stress responses, and (3) single-dimensional tools that cannot represent the complexity of doctoral experiences during transitional periods. PERMA-V(N) positions well-being assessment as requiring both positive psychology constructs and *Negativity* recognition to inform comprehensive student support strategies.

Positionality Statement

The research team consisted of a faculty member at the study institution with experience of QEs as an examiner, an engineering education post-doctoral researcher with separate doctoral experiences as an EDS and as a STEM education doctoral student who navigated QEs in both disciplines, and a biomedical engineering undergraduate student without experience of the QE process. The undergraduate student brought a perspective which was unencumbered by doctoral-level academic stress but was knowledgeable about current concerns and stressors experienced by students in the shared mid-sized departmental student body. The team's familiarity with the institutional context provided valuable insight into local QE processes but required conscious attention to avoid assumptions about universal doctoral experiences. The diversity of career stages, QE experience levels, age, racial and gender identities was leveraged during coding discussions to interrogate assumptions about what constitutes well-being indicators in doctoral education contexts.

Methods

Participants

Participants were recruited from a pool of 2nd year biomedical engineering EDS at a mid-size private R1 university during the Spring term of one academic year. Students were recruited from this pool because they were preparing to take their (oral) QEs immediately after the term. Demographic information was collected using a Qualtrics survey that also collected data relating to measurements of Engineering Identity and Research Science Identity as part of a larger study exploring the experiences of 2nd year EDS during the QE process.

Nine EDS each participated in multiple semi-structured interviews leading up to their QEs (Spring, 2025). The participants comprised five female and four male students. Three participants identified as White, three identified as Latinx, two identified as Asian, and one identified as Asian and Latinx. Two participants identified as first-generation students. Six participants had engineering undergraduate degrees and three had science degrees. Specific combinations of participant characteristics are not given to reduce risk of identification. Sample size followed established guidelines for qualitative framework development (6–12 participants)[23], [24], prioritizing coding depth over participant breadth for initial reliability testing. The study was approved by the institutional IRB (STU00223764); informed consent was given as part of the survey and at the start of the interview.

Data

Interviews were conducted over Zoom, each of which lasted 30–40 minutes. Interview protocols were developed to be semi-structured and explore participant well-being, identities (engineering identity, research science identity, and other personal/demographic identities), and any change over time and interaction between these elements. Each interview covered topics such as classwork, research efforts, QE preparation strategies, relationships, and well-being. Prompts were intended to be neutral in nature to avoid predisposing participants to respond positively or negatively. Example interview prompts are given in Appendix A. Only portions of the interview that related to well-being (~50% of each interview) were used as data for the work described in this paper. Twelve interviews were required—six interviews were used for code development and coder calibration, and six were used for reliability assessment. Interview data were organized into *sections of meaning* and *distinct statements*. *Sections of meaning* were larger data portions that had a distinct topic (e.g. mentoring received from PI). Embedded in these portions were *distinct statements*, which were specific clauses to which codes were applied. Metrics for each of the interview portions used for reliability testing are given in Table 1. All coding was performed using ATLAS.ti [25].

Table 1: Characteristics of the six interview data excerpts used for reliability assessment

Interview ID	**Lines of text**	**Duration (min:sec)**	***# Sections of Meaning***	***# Distinct Statements***
EDS8_I1	108	18:40	13	37
EDS9_I1	151	17:10	10	37
EDS1_I1	155	19:00	12	30
EDS2_I2	82	15:08	10	37
EDS3_I2	225	21:47	10	25
EDS4_I2	150	16:44	11	27

Coding mechanics; calibration phase

For the calibration phase, data analysis followed a constant comparative approach [26] with three researchers using deductive coding based on the proposed PERMA-V(N) framework. Two researchers coded separately and then met with a third researcher who acted as the mediator. Data analysis started with each coder identifying *sections of meaning* and *distinct statements*, whilst simultaneously applying PERMA-V(N) codes: *Positivity* (P), *Engagement* (E), *Relationships* (R), *Meaning* (M), *Accomplishment* (A), *Vitality* (V), and *Negativity* (N).

Positivity encompassed positive emotions (joy, contentment), positive affect, and positive assessments of circumstances and events. *Negativity* included negative affect, negative emotions, and negative assessments of circumstances and events. These definitions acknowledge that psychological experiences manifest across emotional, evaluative, and circumstantial domains, particularly relevant during doctoral studies where students may discuss stressors in evaluative rather than purely emotional terms.

Discussions-to-consensus were had with the third team member to achieve calibrated efforts for identifying *sections of meaning* and *distinct statements*, and for the application of the appropriate code(s). Coders and the third member determined the parameters for each code.

Coding mechanics; reliability assessment phase

For the reliability assessment phase, one coder was responsible for identifying both *sections of meaning* and the *distinct statements*. Next, the file was split using the coding software and each of two coders applied codes separately considering each *distinct statement* in turn. Multiple codes, or none, could be applied to any one *distinct statement*.

In addition to applying codes to each *distinct statement*, each coder identified the *primary code* for each statement with any other applied codes determined to be *secondary codes*. The *primary code* was intended to capture the dominant aspect of PERMA-V(N) pertaining to each *distinct statement*. After code application, the separately coded transcript files were merged using the coding software such that codes from both coders were aligned with coder attribution preserved. Each coder then individually reviewed the merged file for differences in code applications and was able to revise their coding decisions if they had overlooked or misread a statement. Finally, the two coders met and discussed outstanding code disagreements. This discussion typically continued to consensus although reaching a final consensus was not required. Thus, for each code applied to each *distinct statement* there were four possible coding outcomes:

I. if both coders agreed (both applied the code or both chose to not apply the code), this was recorded as pre-review agreement and no discussion was had,

II. if one coder identified during the independent review stage that they had missed applying the code and agreed with the other coder's code application, this was recorded as post-review agreement and no discussion was had,

III. if coders disagreed on a specific code, each coder took turns to explain why they coded the way they did, explaining how their assumptions and thought processes drove their choice. If both coders agreed on the code (either application of the code, or agreement to not apply the code), this was recorded as post-discussion agreement,

IV. if, however, there was still disagreement after the discussion, this was recorded as post-discussion disagreement.

Cohen's Kappa (κ) calculation for reliability

Because the application of each code was a binary decision and multiple codes could be applied to a single *distinct statement*, inter-rater reliability was assessed using Cohen's κ [27]. Successive Cohen's κ values were calculated for three stages: pre-review agreement, post-review agreement (after individual revision), and post-discussion agreement for the two codes of focus—*Positivity* and *Negativity*—and separate calculations were performed for all instances of a code's application (primary and secondary combined) and for primary code applications alone. For other codes (*Engagement*, *Relationships*, *Meaning*, *Accomplishment*, and *Vitality*), calculations were performed for the post-review stage.

Results

Table 2 details the number of applications for each code from each coder for the six interviews analyzed in the reliability assessment phase of this study. Calculated values of Cohen's κ for each code are given in Table 3. To show the variation in Cohen's κ between analyzed interview transcripts, and how agreement increased through the review and discussion stages, individual Cohen's κ values are shown graphically in Figure 1 for *Positivity* and *Negativity*. All applications (*primary* and *secondary* combined) are presented in the left panel, and *primary* alone are presented in the right panel.

Table 2: Count of each code application by interview transcript and by coder. P and N code counts are reported at the post-review stage of analysis

Interview ID	Coder	Code application counts						
		P	**E**	**R**	**M**	**A**	**V**	**N**
EDS8_I1	1	4	15	10	0	3	5	16
	2	5	13	12	2	3	6	20
EDS9_I1	1	8	3	4	0	1	1	7
	2	9	3	4	0	4	1	7
EDS1_I1	1	8	15	5	0	6	0	13
	2	7	25	6	4	0	0	16
EDS2_I2	1	5	9	10	0	1	2	9
	2	2	14	7	2	5	3	7
EDS3_I2	1	4	3	8	1	3	0	13
	2	2	3	4	0	3	0	9
EDS4_I2	1	1	6	10	0	0	2	15
	2	0	11	8	1	3	2	11

Table 3: Cohen's κ measure of agreement for each code. Unless otherwise specified, agreement at the post-review stage of analysis is reported. Where used, parentheses indicate Cohen's κ calculations for the subset of applications as the *primary code.*

Code	Cohen's κ
P, pre-review	0.881 (0.966)
P, post-review	0.886 (0.980)
P, post discussion	1.000 (1.000)
E	0.536
R	0.806
M	0.937
A	0.797
V	0.988
N, pre-review	0.732 (0.902)
N, post-review	0.794 (0.948)
N, post discussion	1.000 (1.000)

At an individual interview/transcript level, Cohen's κ values varied at the pre-review ($0.3 \leq \kappa \leq 1.0$) and post-review ($0.38 \leq \kappa \leq 1.0$) stages, however complete agreement was achieved for every interview through discussion.

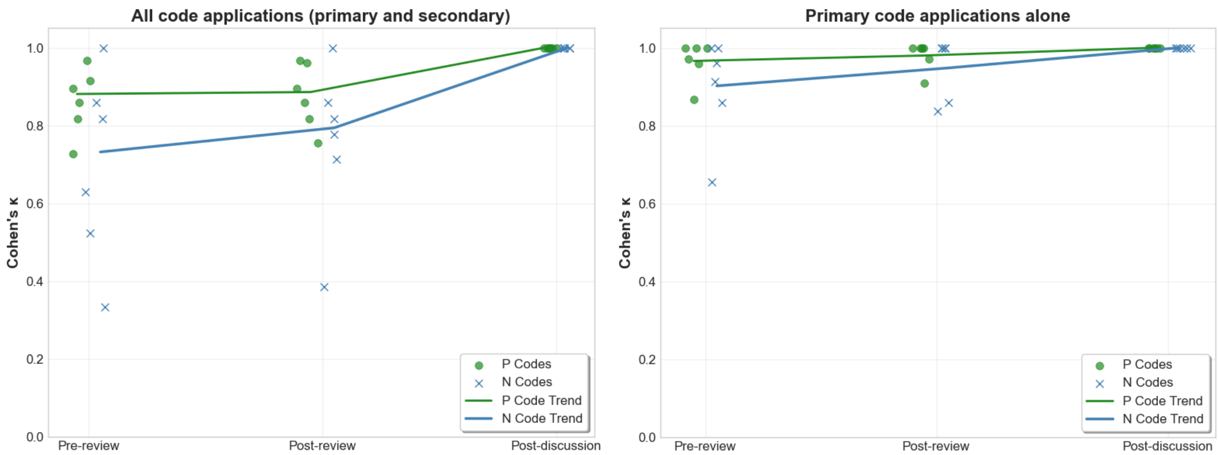

Figure 1. Cohen's κ for the codes *Positivity* and *Negativity*, calculated separately for each analyzed interview transcript. Trend lines show the pooled Cohen's κ for all transcripts combined.
$\kappa > 0.8$ is typically acknowledged as 'excellent agreement'; $0.6 < \kappa < 0.8$ is viewed as 'good agreement'

Discussion: Data Distribution

The distribution of codes in this study demonstrates patterns consistent with existing literature on doctoral student experiences during QEs. The substantially higher frequency of *Negativity* codes (n=188) compared to *Positivity* codes (n=104) aligns with documented research describing the stress experienced by engineering doctoral students during this period [18]. This nearly 2-to-1 ratio of negative to positive expressions provides empirical support for the characterization of QEs as high-stress academic milestones and supports the appropriateness of this participant sample for framework development.

Discussion: Inter-Rater Reliability Assessment

Multi-stage Agreement Convergence
The three-stage agreement process (pre-review, post-review, and post-discussion) demonstrated how PERMA-V(N) could be deployed with different levels of coder-collaboration—something that can be necessary when aiming for reliable coding of complex emotional data. The process revealed that a small but consistent fraction of initial disagreements resulted from oversight rather than fundamental interpretive differences, as evidenced by improved agreement following individual review phases. The structured discussion stage reliably resolved remaining disagreements to perfect agreement (κ = 1.0), suggesting that apparent coding conflicts often stemmed from different attention to contextual details rather than irreconcilable theoretical differences about framework application.

Overall Agreement for Negative Emotions
Cohen's kappa values for *Negative emotions* ranged from good to excellent across the reliability assessment interviews, with all disagreements resolved to perfect agreement (κ = 1.0) following coder discussion. While some individual interviews produced lower reliability metrics than desired, the overall pattern demonstrates the framework's applicability for capturing negative emotional experiences during high-stress academic periods.

Primary Code Reliability
Reliability improved substantially when *Negativity* was the *primary code* and *secondary coding* of *Negativity* were not considered (κ = 0.732 pre-review and 0.794 post-review became 0.902 and 0.948, respectively). Values of Cohen's κ greater than 0.8 are considered excellent agreement and $\kappa > 0.9$ is 'near perfect' agreement. This pattern demonstrates that the coders achieved higher consistency when *Negativity* represented the dominant theme of a statement rather than a co-occurring element.

Communication Style Effects on Coding Reliability
Inter-rater reliability varied across individual interviews for both *Positivity* and *Negativity* (Figure 1), despite all participants responding to nominally similar neutral prompts (Appendix A). Inspection of the subset of quotations where coders disagreed revealed patterns in participants' communication tendencies rather than question framing effects. Some students naturally used direct evaluative language (*"I feel overwhelmed", "I feel great about"*), while others embedded equivalent content in factual descriptions (*"The timeline is challenging", "My struggle meals are working"*). For most instances of coder disagreement, the undergraduate research assistant was the more likely source of a coding of *Positivity* or *Negativity*, suggesting that being a near-peer of participants aided transcript interpretation.

This variation represents a methodological consideration rather than framework limitation. The three-stage review process resolved all disagreements to perfect agreement (κ = 1.0) for both codes regardless of discourse style. However, researchers should anticipate that interviews with more embedded or subtle emotional or evaluative content may require additional discussion time to achieve reliability targets. This process additionally highlights the value of having near-peer perspectives on a research team.

Comparative Reliability: Negativity vs. Positivity
While *Negativity* achieved good to excellent reliability, Cohen's kappa values were

consistently lower than those for *Positivity* across all reliability assessment interviews. Several factors likely contribute to this pattern:

i. **Explicit Expression of Positive Experiences:** Interviewees demonstrated more explicit and direct language when describing *Positivity*, as exemplified by EDS2's clear statements: "*I feel great about my 'struggle meals'*". Such directness may reflect students' desire to specifically share 'bright spots' and to do so with direct language during periods of stress [28], [29], making positive codes more straightforward to identify and apply consistently.
ii. **Contextual Inference for Negativity:** Discussion of disagreements over coding of *Negativity* often required coders to read beyond the predetermined *distinct statements* in order to capture full meaning. For instance, EDS1's statement "*Pretty low, I'm not a good student*" required reviewing preceding context to understand that EDS1 was describing the low intensity of QE preparation and not describing intellectual capacity.
iii. **Coder Positionality:** Disagreements between coders did not demonstrate conclusive systematic bias patterns, though the undergraduate female coder showed slightly higher tendency to identify *Negativity* compared to the male postdoctoral coder. This non-significant difference may reflect documented benefits of near-peer perspectives, where individuals closer in developmental stage and academic experience demonstrate enhanced understanding of peer challenges and perspectives [30], and where female coders demonstrate enhanced sensitivity to emotional nuances, particularly for subtle or ambiguous emotional cues [31], [32].

Discussion: Utility of PERMA-V(N) Framework

Addressing Original PERMA-V Limitations

The original PERMA-V framework's inability to distinguish between emotionally-neutral and emotionally-negative aspects of doctoral experiences is demonstrated here through two engagement-coded statements that require different interpretations.

Consider the following statements from EDS1: "*Yes, I'm in class. And my last class is bioelectronics. And then so then you're [one is] prioritizing research, quals and classes....[I'm prioritizing research over quals] because quals are pass–fail, and my research is judged on quality.*", and *"I feel like there's like day-to-day stress of like I get all the stuff I need to get done today done and like, how much free time do I have?"*. Without the *Negativity* dimension that PERMA-V(N) adds, such statements would be coded solely as *Engagement*. With the *Negativity* dimension that PERMA-V(N) adds, the first statement remains coded solely as *Engagement*, but the second, which alludes to stress and lack of free time, can be coded with both *Engagement* (*primary code*) and *Negativity* (coded as a *secondary code*).

This expanded coding capability allows PERMA-V(N) to differentiate between neutral code applications and those that contribute to stress or distress, providing more nuanced understanding of how different wellbeing dimensions interact.

Approximately 20% of all *Positivity* code applications were adjudged as *primary*, whereas 56% of *Negativity* instances were viewed by the coders as being the *primary* code.

Co-occurrence of Positive and Negative Codes
The framework's ability to apply both P and N codes to single statements proved particularly valuable for capturing the complexity of doctoral experiences. EDS3's reflection, "*Well, I guess she's trying to leave space for me to think... which sometimes I feel like it's not like super helpful, because she's not like hand holding my way to the actual outcome. But I think it is good a good learning opportunity*" demonstrates simultaneous frustration (N) and appreciation (P) within a single experience. Similarly, EDS2's description of her current meal planning and cooking: "*I've definitely been having more 'struggle meals', but I feel great about my 'struggle meals'... I feel good about my struggle meals, my girl dinners*" captures both the negative acknowledgement of nutritional compromise and positive emotions about independence and self-care/coping strategies.

Expansion Beyond Discrete Emotions
The research team expanded the coding framework from *Negative emotions* to *Negativity* (and also *Positive emotions* to *Positivity*) to capture a broader spectrum of wellbeing indicators beyond explicitly stated emotional language. This decision aligns with contemporary affective science, which distinguishes between emotions (discrete, intense, short-lived responses) and affect (broader evaluative states that encompass emotions, moods, and attitudes) [20], [33]. *Negativity* represents a fundamental dimension of subjective experience that influences cognition, behavior, and overall wellbeing, often manifesting in graduate students' discourse through evaluative language, criticism, or expressions of dissatisfaction that may not explicitly reference emotional states [19].

This expansion addressed three coding challenges that emerged during preliminary analysis:

- **Challenge 1: Evaluative statements without emotional language.** Participants frequently expressed negative evaluations of their experiences using cognitive rather than emotional framing. For instance, EDS4's statement "*What's on my mind is why do they* [faculty] *feel like this is a good use of our time?*" clearly indicates *Negativity* toward institutional practices without employing emotional vocabulary. Under a strict Negative emotions framework, such expressions would be excluded despite their clear relevance to wellbeing assessment.
- **Challenge 2: Implicit affective content requiring inference.** Some statements contained ambiguous affective content that could create inter-rater reliability issues. EDS3's description of peer dynamics—"*There is also that comparison of each other's research achievements... that peer pressure is just what I'm referring to*"—exemplifies this challenge. The reference to "peer pressure" implies *Negativity*, but coding such statements under a narrow 'emotions only' framework would require coders to make potentially inconsistent inferences about unstated emotional responses.
- **Challenge 3: Compound statements mixing factual and affective content.** Graduate students often embedded evaluative content within seemingly neutral descriptions of their experiences. EDS4's response about QEs— "*Research is still continuing, unfortunately.* [I] *would like to just study literature reviews* [instead]"—demonstrates this pattern. The factual statement ("research is continuing") coupled with the evaluative modifier ("unfortunately") and implicit preference statement reveals *Negativity* toward the competing demands of research and exam preparation.

PERMA-V(N)'s expanded definition of *Negativity* (N) encompasses: (1) explicitly stated negative emotions, (2) evaluative statements indicating dissatisfaction or criticism, (3) expressions of preference for alternative circumstances, and (4) implicit negative evaluations embedded within descriptive content. This expansion increases the framework's sensitivity to subtle expressions of diminished wellbeing while maintaining clear coding criteria to preserve inter-rater reliability.

The enhanced *Negativity* dimension particularly improves capture of systemic and structural stressors—such as institutional practices, advisor relationships, and program requirements—that graduate students may discuss in evaluative rather than emotional terms, thereby providing a more comprehensive assessment of factors affecting doctoral student wellbeing.

Conclusion

This study establishes PERMA-V(N) as a reliable qualitative coding framework for capturing EDS well-being during QEs. Inter-rater reliability ranged from good to excellent across all dimensions (Cohen's $\kappa = 0.536$ to 0.988), with structured discussion resolving all disagreements to perfect agreement ($\kappa = 1.0$) for both *Positivity* and *Negativity*. This demonstrates that the framework provides sufficiently clear construct definitions to support reliable qualitative coding even when initial interpretations differ.

The addition of *Negativity* addresses a critical gap in the original PERMA-V framework. The substantially higher frequency of *Negativity* codes (n=188) compared to *Positivity* codes (n=104)—a nearly 2:1 ratio—demonstrates that PERMA-V alone would have failed to capture the majority of evaluative content in student interviews during this high-stress period. The framework's capacity to code co-occurring positive and negative experiences within single statements, alongside its expansion beyond discrete emotions to include evaluative content, provides nuanced understanding of well-being complexity during critical academic

PERMA-V(N) offers engineering education researchers and practitioners an improved tool for assessing EDS well-being during critical transitions, with specific applications for mentoring conversations, program evaluation, and targeted intervention development, moving beyond deficit-focused measures. Research using PERMA-V(N) can inform mentor training by highlighting how students experience co-occurring positive and negative states during high-stress periods, suggesting that check-in conversations should create space for both flourishing and distress rather than treating these as mutually exclusive.

Limitations and Future Work

The single-institution setting and focus on biomedical engineering during QEs limits generalizability claims. This controlled context was appropriate for establishing baseline inter-rater reliability, but framework transferability requires testing across institutional cultures, engineering disciplines, and academic milestones. While the sample size (n=9) was adequate for reliability assessment, future work should examine whether PERMA-V(N) maintains comparable performance with different doctoral populations and across varied high-stress academic contexts beyond qualifying examinations.

Acknowledgements
This material is based upon work supported by the National Science Foundation under Grant 2306239. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[1] S. Basir and E. Burkholder, “Investigating faculty perspectives on written qualifying exams in physics,” *Phys. Rev. Phys. Educ. Res.*, vol. 20, no. 1, p. 010139, 2024, doi: 10.1103/physrevphyseducres.20.010139.

[2] M. Hooper, J. Tawney, E. Palmer, J. Ragan, M. Hooper, and Y. Gonzalez, “Student-led program to improve equity in Ph.D. oral qualifying exams,” *2023 ASEE Annu. Conf. Expo. Proc.*, 2023, doi: 10.18260/1-2--44337.

[3] S. Guloy, G. Hum, and D. K. O’Neill, “Voices at the gate: Faculty members’ and students’ differing perspectives on the purposes of the PhD comprehensive examination,” *Assess. Eval. High. Educ.*, vol. 45, no. 2, pp. 279–291, 2020, doi: 10.1080/02602938.2019.1637819.

[4] J. E. McLaughlin, K. Morbitzer, M. Meilhac, N. Poupart, R. L. Layton, and M. B. Jarstfer, “Standards needed? An exploration of qualifying exams from a literature review and website analysis of university-wide policies,” *Stud. Grad. Postdr. Educ.*, vol. 15, no. 1, pp. 19–33, 2024, doi: 10.1108/sgpe-11-2022-0073.

[5] T. M. Evans, L. Bira, J. B. Gastelum, L. T. Weiss, and N. L. Vanderford, “Evidence for a mental health crisis in graduate education,” *Nat. Biotechnol.*, vol. 36, no. 3, pp. 282–284, 2018, doi: 10.1038/nbt.4089.

[6] K. Levecque, F. Anseel, A. D. Beuckelaer, J. V. der Heyden, and L. Gisle, “Work organization and mental health problems in PhD students,” *Res. Polic.*, vol. 46, no. 4, pp. 868–879, 2017, doi: 10.1016/j.respol.2017.02.008.

[7] F. Zhang, K. Litson, and D. F. Feldon, “Social predictors of doctoral student mental health and well-being,” *PLoS ONE*, vol. 17, no. 9, p. e0274273, 2022, doi: 10.1371/journal.pone.0274273.

[8] C. M. Hazell, L. Chapman, S. F. Valeix, P. Roberts, J. E. Niven, and C. Berry, “Understanding the mental health of doctoral researchers: a mixed methods systematic review with meta-analysis and meta-synthesis,” *Syst. Rev.*, vol. 9, no. 1, p. 197, 2020, doi: 10.1186/s13643-020-01443-1.

[9] M. E. P. Seligman, “Flourish: A visionary new understanding of happiness and well-being.,” *Flourish: A visionary new understanding of happiness and well-being.*, pp. xii, 349–xii, 349.

[10] J. Stubb, K. Pyhältö, and K. Lonka, “Balancing between inspiration and exhaustion: PhD students’ experienced socio-psychological well-being,” *Stud. Contin. Educ.*, vol. 33, no. 1, pp. 33–50, 2011, doi: 10.1080/0158037x.2010.515572.

[11] K. J. Corner, “Exploring the reliability and validity of research instruments to examine secondary school principals’ authentic leadership behavior and psychological capital,” The University of Arizona, 2015.

[12] L. A. Clark and D. Watson, "Constructing Validity: Basic Issues in Objective Scale Development," *Psychol. Assess.*, vol. 7, no. 3, pp. 309–319, 1995, doi: 10.1037/1040-3590.7.3.309.

[13] J. Butler and M. L. Kern, "The PERMA-Profiler: A brief multidimensional measure of flourishing," *Int. J. Wellbeing*, vol. 6, no. 3, pp. 1–48, 2016, doi: 10.5502/ijw.v6i3.526.

[14] M. L. Kern, L. E. Waters, A. Adler, and M. A. White, "A multidimensional approach to measuring well-being in students: Application of the PERMA framework," *J. Posit. Psychol.*, vol. 10, no. 3, pp. 262–271, 2015, doi: 10.1080/17439760.2014.936962.

[15] A. Eacker, "Well-being definition and measures in medical education.," in *Medical Professionalism Best Practices: Addressing Burnout and Resilience in Our Profession*, R. L. Byyny, R. Byyny, S. Christensen, and J. D. Fish, Eds., in JAMA. , 2020, pp. 91–103. doi: 10.1001/jama.1971.03180370073017.

[16] K. Shanachilubwa, G. Sallai, and C. G. P. Berdanier, "Investigating the tension between persistence and well-being in engineering doctoral programs," *J. Eng. Educ.*, vol. 112, no. 3, pp. 587–612, 2023, doi: 10.1002/jee.20526.

[17] J. Tabora and D. P. O'Neill, "Work in Progress: Exploring Qualifying Exam Experiences in Engineering Doctoral Students using Well-Being Constructs," in *2025 ASEE Annual Conference & Exposition*, in 2025 ASEE Annual Conference & Exposition Proceedings. Jun. 2025. doi: 10.18260/1-2--44263.

[18] J. Mirabelli, J. Cromley, K. Jensen, D. Robbennolt, and A. Hart, "Work in Progress: Exploring the Landscape of Stressors Experienced by Doctoral Engineering Students," *2023 ASEE Annu. Conf. Expo. Proc.*, 2023, doi: 10.18260/1-2--44263.

[19] D. Watson and A. Tellegen, "Toward a Consensual Structure of Mood," *Psychol. Bull.*, vol. 98, no. 2, pp. 219–235, 1985, doi: 10.1037/0033-2909.98.2.219.

[20] J. A. Russell, "Core Affect and the Psychological Construction of Emotion," *Psychol. Rev.*, vol. 110, no. 1, pp. 145–172, 2003, doi: 10.1037/0033-295x.110.1.145.

[21] B. L. Fredrickson, "The Role of Positive Emotions in Positive Psychology," *Am. Psychol.*, vol. 56, no. 3, pp. 218–226, 2001, doi: 10.1037/0003-066x.56.3.218.

[22] B. L. Fredrickson and C. Branigan, "Positive emotions broaden the scope of attention and thought-action repertoires," *Cogn. Emot.*, vol. 19, no. 3, pp. 313–332, 2005, doi: 10.1080/02699930441000238.

[23] J. J. Francis *et al.*, "What is an adequate sample size? Operationalising data saturation for theory-based interview studies," *Psychol. Heal.*, vol. 25, no. 10, pp. 1229–1245, 2010, doi: 10.1080/08870440903194015.

[24] G. Guest, A. Bunce, and L. Johnson, "How Many Interviews Are Enough?," *Field Methods*, vol. 18, no. 1, pp. 59–82, 2006, doi: 10.1177/1525822x05279903.

[25] Scientific Software Development~GmbH, *Atlas.ti*. 2024.

[26] B. G. Glaser, “The Constant Comparative Method of Qualitative Analysis,” *Soc Probl*, vol. 12, no. 4, pp. 436–445, 1965, doi: 10.2307/798843.

[27] J. Cohen, “A Coefficient of Agreement for Nominal Scales,” *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960, doi: 10.1177/001316446002000104.

[28] F. Fang and X. Tang, “The Relationship Between Chinese English Major Students’ Learning Anxiety and Enjoyment in an English Language Classroom: A Positive Psychology Perspective,” *Front. Psychol.*, vol. 12, p. 705244, 2021, doi: 10.3389/fpsyg.2021.705244.

[29] C. Li, J.-M. Dewaele, and G. Jiang, “The complex relationship between classroom emotions and EFL achievement in China,” *Appl. Linguistics Rev.*, vol. 11, no. 3, pp. 485–510, 2020, doi: 10.1515/applirev-2018-0043.

[30] S. U. Abeywardana, S. Velasco, N. Hall, J. Dillon, and C.-A. Chun, “Near-peer Mentoring in an Undergraduate Research Training Program at a Large Master’s Comprehensive Institution: The Case of CSULB BUILD.,” *UI J.*, vol. 11, no. 1, 2020.

[31] A. H. Fischer, M. E. Kret, and J. Broekens, “Gender differences in emotion perception and self-reported emotional intelligence: A test of the emotion sensitivity hypothesis,” *PLoS ONE*, vol. 13, no. 1, p. e0190712, 2018, doi: 10.1371/journal.pone.0190712.

[32] T. M. Chaplin and A. Aldao, “Gender Differences in Emotion Expression in Children: A Meta-Analytic Review,” *Psychol. Bull.*, vol. 139, no. 4, pp. 735–765, 2013, doi: 10.1037/a0030737.

[33] K. R. Scherer, “What are emotions? And how can they be measured?,” *Soc. Sci. Inf.*, vol. 44, no. 4, pp. 695–729, 2005, doi: 10.1177/0539018405058216.

Appendix: Example prompts used in the semi-structured interviews

Discussion Topic: Current activities

1. Are you taking any classes? What's the workload like?
2. Tell me about your current level of research activity. What kinds of activities are you engaged in?
3. Tell me about how you are preparing for the qualifying exams. What strategies are you employing? How do you balance quals prep with classes and/or research.

Discussion Topic: Well-being

1. Tell me about your current feelings and physical health. Have there been any changes over time / since our last interview?
2. Tell me about your current stress levels. Are there particular stressors (people? Events? Expectations?)
3. Tell me about your well-being. This could be physical, emotional, psychological.
4. Are your well-being, stress, anxiety levels mostly stable or more variable?
5. Looking more broadly, do you have any thoughts on how your well-being and/or stress compare to your peers?

Discussion Topic: Identity

1. Tell me about how your identities might be shaping your preparations for the qualifying exams? [identities might include: identity as an engineer, as a research scientist; or other personal identities such as race, gender, first-generation, domestic/international, etc.]
2. You did the Qualtrics survey which measures strength of *Identifying as an Engineering, and as a Research-Scientist.* There are three main components to each identity: being *recognized* as an engineer/RS, being *interested* in the subject, and in your *performance or competence*.
    a. Do you have any reflections on how you identify with these two Professional Identities and with any of these 'sub-constructs' in particular (recognition, interest, performance)?
    b. Do you sense any change in the strengths of these identities in recent months?
3. Do you feel your EI and RSI identity complement each other, is there any tension between them?