2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Paper ID #52652

Comparing Broad and Focused Interview Prompts for Biomedical Engineering Doctoral Students Discussing Well-being and Identity.

Kate Manion, Northwestern University
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.

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.

Comparing Broad and Focused Interview Prompts for Biomedical Engineering Doctoral Students Discussing Well-being and Identity

Introduction

This methods research full paper examines how different types of interview prompts influence the depth, diversity, and relevance of responses in qualitative interviews about well-being and identity among engineering doctoral students.

Qualitative researchers face a fundamental methodological question when designing interview protocols: Should questions probe specific subconstructs directly, or should they invite broader exploration that allows themes to emerge organically? Foundational qualitative methodology consistently argues that question format shapes data quality [1], yet empirical evidence directly comparing these approaches remains sparse. Studies comparing question types suggest format matters. Hansen and Świderska [2] demonstrated in representative surveys that question format affects not just response depth but content valence—closed questions yielded more negative attitudes while broad responses were "not only richer, but also more positive." In interview research, Ogden and Cornwell [3] analyzed 400 interview questions across ten health studies, finding that open questions predicted longer responses and more action-oriented accounts. However, this work focused on health research rather than educational contexts, did not examine specific student populations, and was not designed for high-stakes or longitudinal settings.

Within engineering education, qualitative methodology frameworks emphasize rigor and validation [4], [5]. Castillo-Montoya's Interview Protocol Refinement (IPR) Framework [6] provides guidance for developing effective protocols through alignment, conversation construction, feedback, and piloting. Yet these frameworks address protocol *development*—how to create good protocols—not protocol *comparison*—whether different questioning approaches yield systematically different data with the same participants. Bahnson *et al.* [7] demonstrated the value of pilot testing in refining interview protocols for engineering doctoral students, finding that an initial broad question ("tell me about your experiences as a graduate student") generated richer narratives than more specific identity questions, leading them to restructure their protocol. However, their work focused on improving a single protocol rather than comparing alternative approaches.

Studies examining complex constructs like well-being and identity in engineering doctoral students employ varied questioning strategies [8][9] without directly examining whether different approaches access these constructs differently. Within-subjects designs—where participants serve as their own controls—are well-established in quantitative research for reducing between-person variability [10], but no studies systematically compare qualitative question types with the same participants in a controlled crossover design.

This exploratory study addresses that gap through a within-subjects comparison of two questioning strategies. Broad prompts invite participants to explore how theoretical constructs or domains manifest in their current experiences, referencing frameworks introduced earlier while allowing participants to select which elements to discuss and how. For example: *"If you recall, one way to consider 'well-being' is through the PERMA-V framework (Positive emotions, Engagement... ...Vitality); how do you feel these concepts are manifesting in your current experiences?"*. Focused prompts target specific subconstructs, or relationships, experiences, or concerns that participants identified in previous interviews,

building on longitudinal knowledge of each participant's unique situation. For example: *"Last time you were telling me about your relationship with your P.I.; how has this been since our last interview?"* Both prompt types remain open-ended and maintain conversational flexibility; the distinction lies in whether the prompt allows participants to choose among theoretical constructs for discussion (broad) or pursues participant-identified priorities from prior interviews (focused). Importantly, both prompt types may reference theoretical frameworks introduced earlier in the study; the distinction is not between theory-informed and atheoretical prompting, but between prompts that allow participants to select which constructs or experiences to foreground versus those that target constructs or prior disclosures specified by the interviewer. Nine biomedical engineering doctoral students participated in multiple interviews during qualifying exam preparation—a high-stakes transitional period in doctoral education. Using a crossover design balanced participant characteristics across conditions, with each participant experiencing both protocol types in counterbalanced order.

This paper examines two research questions:
RQ1: Do focused and broad prompts differ in coding density?
RQ2: Do focused and broad prompts differ in code and thematic distribution and co-occurrence patterns?

Rather than assuming one protocol type is universally superior, this study examines how question format shapes qualitative data in practice. Findings provide initial empirical evidence to inform methodological decisions that researchers currently make based primarily on theoretical arguments, with specific implications for engineering education research on sensitive topics like well-being and identity.

Positionality Statement

The research team consisted of a faculty member at the study institution with experience of qualifying exams as an examiner, an engineering education post-doctoral researcher with separate doctoral experiences as an engineering doctoral student and as a STEM education doctoral student who navigated qualifying exams in both disciplines, and a biomedical engineering undergraduate student without experience of the qualifying examination 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, and genders was leveraged during coding discussions to interrogate assumptions about what constitutes well-being indicators in doctoral education contexts. The research team adopted a pragmatist stance focused on examining what differences exist in practice rather than making ontological claims about superior approaches. This methods research examines qualitative interviewing practices empirically while respecting the interpretive nature of interview data.

Methods

Participants

Participants were recruited from a pool of 2nd year biomedical engineering doctoral students at a mid-size private R1 university during the Spring term of one academic year as they were preparing for taking their qualifying exams (QEs). Demographic information was collected using a Qualtrics survey as part of a larger study.

Nine students each participated in multiple semi-structured interviews leading up to the QEs and an additional interview shortly after. Interview spacing varied across participants due to individual scheduling, with interviews typically occurring approximately 4–6 weeks, 2–3 weeks, and 1 week before QEs. 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 (bachelor's or greater) college 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. The study was approved by the institutional IRB (STU00223764); informed consent was given for both survey and interview.

Data

Interviews were conducted by the post-doctoral researcher over Zoom, each of which lasted 30–40 minutes and covered three major topics: 1) classwork, research efforts, and QE preparation strategies, 2) well-being, and 3) engineering and research science identity. Each participant's Interview #1 was conducted with a common semi-structured protocol which included brief explanations of the frameworks for discussing well-being (PERMA-V(N) [11] and engineering/research science identity in graduate students [12], [13].

Based on interviewer observations during Interview #1, participants were categorized by 1) stress level (low, medium, high) assessed through observed presentation (mood, affect, tone) and participant self-disclosure (e.g., *"I'm already pretty stressed"*), and 2) conversational style (forthcoming, intermediate, or reserved). These subjective classifications were used to generate approximately balanced groups in the crossover design, with the aim of mitigating potential confounds from interview timing relative to qualifying exams and longitudinal rapport development. Classifications were not used as outcome variables or analytic factors. The senior researcher was blinded to group assignments per IRB requirements. Participants were assigned to two groups with approximately equal representation of participants with each combination of characteristics. Table 1 shows the categorizations and assigned group for each participant.

Table 1. Participant characteristics used to determine cross-over study group assignment

Participant ID	Stress level (presented and self-reported)	Conversational style	English as first language	Assigned group
EDS 1	Low	Forthcoming	Yes	A
EDS 2	Medium	Forthcoming	Yes	A
EDS 3	Medium	Forthcoming	No	B
EDS 4	High	Reserved	Yes	A
EDS 5	Medium	Intermediate	No	A
EDS 6	High	Forthcoming	Yes	A
EDS 7	High	Intermediate	Yes	B
EDS 8	Medium	Reserved	No	B
EDS 9	Medium	Reserved	Yes	B

Interview #1 data were excluded from analysis because initial interviews employed only broad prompts and lacked the methodological variation necessary for prompt-type comparison. However, Interview #1 fulfilled essential preparatory functions: introducing theoretical frameworks to participants, eliciting participant-identified priorities and concerns that subsequently informed the content of personalized focused prompts in later interviews. Framework introduction established shared vocabulary for discussing abstract constructs, an approach that has precedent in engineering education research [14].

For Interviews #2 and 3, two modified protocols were used. For each section (topic) of an interview (e.g., well-being), participants were asked broad questions as well as focused questions that followed-up on specific comments made in proceeding interviews. For example, broad prompts took formats such as: *"If you recall, one way to consider 'well-being' is through the PERMA-V framework (Positive emotions, Engagement… …Vitality); how do you feel these concepts are manifesting in your current experiences"* whereas a focused portion of the interview might include a question such as: *"last time you were telling me about your relationship with your P.I.; how has this been since our last interview?"*.

A cross-over study design was used: for Interview #2, participants in Group A received broad prompts followed by focused prompts, whilst those in Group B received focused prompts first and broad prompts second. For Interview #3 the protocols for Groups A and B were reversed.

Coding mechanics

Interview data were organized into *sections of meaning* and *distinct statements*. *Sections of meaning* were larger data portions that discussed a certain topic (e.g., mentoring received from PI). Embedded in these portions were *distinct statements*, which were specific clauses to which codes were applied. Initial coding was performed for other methodological work [11], [15] . Subsequently, transcripts were analyzed using hybrid inductive-deductive thematic coding [16]. The primary coder developed an initial codebook informed by PERMA-V(N) while remaining open to emergent themes, resulting in approximately 90 codes. Through collaborative iterative refinement involving regular calibration sessions between coders and the senior researcher, codes were systematically reviewed, combined, and pruned. This iterative process resulted in 75 final codes consolidated into 15 thematic families that formed the basis for subsequent analysis. For final analyses, the primary coder coded all transcripts which were then discussed to consensus as a team. Additionally, each *section of meaning* and *distinct statement* was categorized as originating in either an broad part of the transcript, or in a focused question part. All coding was performed using Atlas.ti.

During preliminary analysis, the team recognized that the crossover design—while successful in alternating protocol order—was not fully compatible with semi-structured interviews thus did not produce cleanly separable 'broad' versus 'focused' data within each interview. The broad prompts of both protocols elicited focused follow-up questions that emerged naturally from participant responses during semi-structured conversation. For example, even in sections beginning with broad prompts about well-being, the interviewer would probe specific participant statements with targeted follow-up questions. Consequently, many excerpts initially categorized as responses to broad prompts were actually responses to these emergent focused probes. To ensure accurate comparison, each *distinct statement* was reclassified based on the specific prompt type that directly elicited it, rather than the nominal protocol section in which it appeared. No recoding was required; the existing codes were retained while prompt-type classifications were corrected.

Analysis

To address RQ1, coding density was assessed through the mean number of codes applied per identified *distinct statement*; the presence of a statistically significant difference was tested with a two-tail t-test.

To address RQ2, code and thematic distributions were assessed by testing for differences in the presence or absence of each of the 75 codes and 15 thematic families between responses to focused and broad prompts. Fisher's exact tests were conducted for each thematic family. False discovery rate (FDR) correction was selected over the more conservative Bonferroni correction because this study explores patterns in a novel methodological domain (within-subjects protocol comparison) without *a priori* hypotheses about which specific themes would differ between prompt types. FDR correction balances Type I error control with statistical power in exploratory contexts where rejecting some true null hypotheses is acceptable given the goal of pattern discovery [17].

Code co-occurrence patterns were examined by identifying all possible pairwise co-occurrences among the 15 thematic families within *sections of meaning*. For each co-occurrence pair, Fisher's exact test was used to compare the frequency of that co-occurrence between focused and broad prompt conditions. False discovery rate correction was again applied.

Results

Initial analyses—where excerpts were categorized according to from which part of the protocol they emerged (broad condition or focused condition)—differed in number between protocol sections (focused: 19.9/interview, S.D.=7.1 statements; broad: 9.9/interview, S.D.=5.6 statements; $p<0.01$), however the two groups showed no statistically significant differences for response length per excerpt, thematic breadth, thematic coverage, or code co-occurrence patterns. Hereon, except where noted, all results are presented with re-categorized excerpts based on prompt type rather than protocol section as described above.

Unlike in the initial analyses, focused *prompts* did not generated a statistically different number of coded *distinct statements* than broad *prompts* (focused: mean=16.4/interview, S.D.=6.9 statements; broad: mean=13.4/interview, S.D.=5.3 statements; $p=0.15$). Additionally, the mean number of codes applied per excerpt were not statistically different between responses from each prompt type (focused: mean=4.37 codes, S.D.=1.26 codes; broad: mean=4.42 codes, S.D.=1.10 codes; $p=0.704$).

To address RQ2, calculations using Fisher's exact test with false discovery rate corrections indicated 17 / 75 codes and 6 / 15 themes were more prevalent in responses to one type of prompt that the other to a statistically significant level ($q<0.05$). Results of these calculations are presented in Figures 1 and 2 by showing each code (Figure 1) and theme (Figure 2) plotted on axes of relative prevalence (x-axis) versus the total code prevalence (y-axis). The relative prevalence is described by the odds ratio (see below); the odds ratio is transformed with the logarithm to preserve symmetric scaling in both positive and negative directions of the axis. i.e. equal prevalence of a code or theme in both types of prompt would produce an odds ratio of unity and be plotted on the y-axis.

$$\text{Odds Ratio} = \frac{C_F \Big/ (C_F + \overline{C_F})}{C_{OE} \Big/ (C_{OE} + \overline{C_{OE}})}$$

Where C (coded) is the number of excerpts with a specific code applied, $\bar{C}$ (not coded) is the number of excepts without the code applied, subscripts *F* and *OE* indicate focused and broad prompt types, respectively.

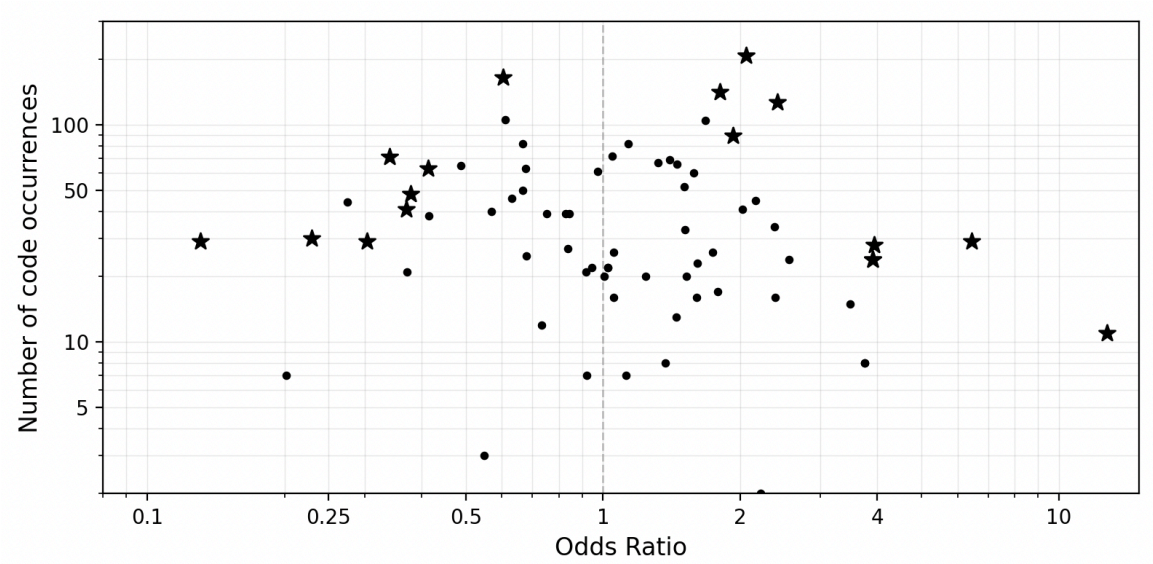

Figure 1. Relationship between individual code frequency and prompt type prevalence. The y-axis shows total occurrences of each code across all interviews. The x-axis shows the odds ratios comparing code prevalence in broad versus focused prompts, with positive values indicating greater prevalence in broad responses. Stars denote codes showing statistically significant differences after FDR correction ($q < 0.05$); circles represent non-significant differences.

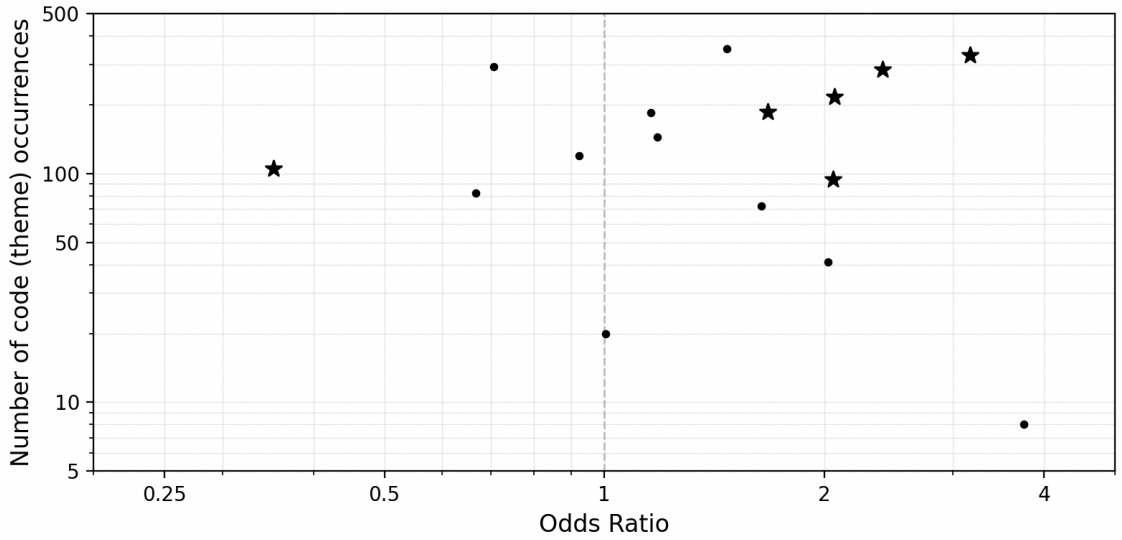

Figure 2. Relationship between thematic family frequency and prompt type prevalence. The y-axis shows total occurrences of each thematic family across all interviews. The x-axis shows the odds ratios comparing theme prevalence in broad versus focused prompts, with positive values indicating greater prevalence in broad responses. Stars denote thematic families showing statistically significant differences after FDR correction ($q < 0.05$); circles represent non-significant differences.

Figure 3 displays thematic co-occurrence patterns as network graphs for focused (left) and broad (right) prompt responses. Nodes represent the 15 thematic families; edge thickness indicates co-occurrence frequency within sections of meaning. Similar network structures across both conditions reflect that no co-occurrence pairs differed significantly (Fisher's exact tests with FDR correction, all $q > 0.05$). Readers should attend to whether the relative thickness of edges connecting each pair of thematic families differs substantially between the two panels. The broad similarity of the network across conditions visually supports the statistical finding of no significant co-occurrence differences.

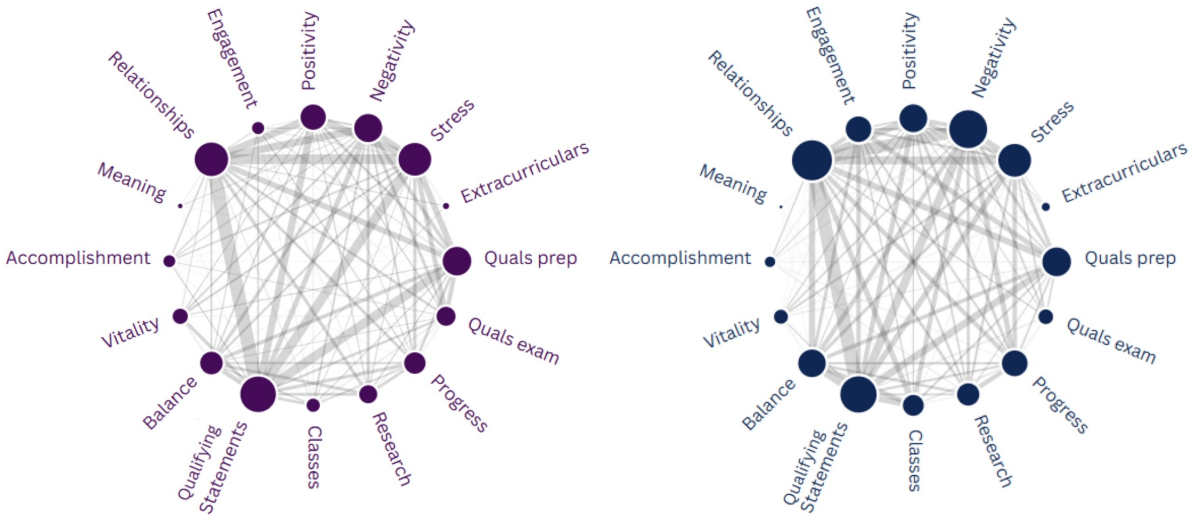

Figure 3. **Thematic family co-occurrence networks for focused prompts (left, purple) and broad prompts (right, blue).** Node size represents the frequency of each thematic family occurrence (larger nodes represent more frequent theme occurrence). Line thickness indicates the frequency of co-occurrence between thematic families within the same *section of meaning*.

Discussion

Initial analyses categorizing excerpts by nominal protocol section (focused versus broad) revealed no significant differences despite a large imbalance in the number of coded excerpts. However, reclassifying each statement by the actual prompt type that directly elicited it—rather than the nominal protocol section in which it appeared—both balanced the numbers of each type of response whilst also revealing six themes with statistically significant different prevalences. This reclassification emerged from recognition that semi-structured interviews inherently involve responsive follow-up questioning [18], [19]; sections beginning with broad prompts naturally elicited focused follow-up prompts by the interviewer. This recognition emerged through collaborative analysis sessions where team members with varying QE experiences interrogated what constituted broad versus focused responses, with the undergraduate researcher's outsider perspective proving particularly valuable in questioning our initial categorization assumptions. The reclassification represents adherence to actual conversational dynamics rather than nominal protocol design.

The mean number of codes applied per distinct statement did not differ significantly between prompt types ($p=0.704$), indicating equivalent coding density across conditions. Coding density — the number of codes applied to a single statement — serves here as an index of response complexity: statements attracting more codes contain multiple co-occurring

constructs, reflecting responses that are more interwoven or multifaceted. Even if there was a statistically significant difference, higher density should not be interpreted as indicative of superior data; rather, comparable density across prompt types suggests that both elicitation strategies generate responses of equivalent analytical complexity, even where the topics emphasized differ. Whether greater thematic intricacy might be desirable depends on the research goals and analysis strategy of a given study.

Focused prompts generated shorter coded statements than broad prompts, yet coding density did not differ significantly. This supports anecdotal observations that broad prompts elicit extended elaboration on fewer topics, while focused prompts generate more discrete, codable units. Qualitative observations, both from the interviewer at the time, and the research team during data analysis, revealed that broad responses often involved participants "thinking-aloud"—processing and refining single ideas rather than introducing multiple distinct concepts. This distinction could have practical implications: focused prompts may be more 'efficient' when interview duration is constrained or comprehensive thematic coverage is a priority, while broad prompts allow deeper processing of participant-prioritized topics and, in some cases, more considered, developed or deeper transcript data. That both approaches generate comparably information-rich responses per excerpt challenges assumptions that one questioning approach universally produces "richer" data, instead suggesting that richness manifests differently.

False discovery rate (FDR) correction was employed rather than Bonferroni correction, reflecting this study's exploratory nature without *a priori* hypotheses [17]. That five of six significant findings under FDR ($q < 0.05$) remained significant under Bonferroni correction ($\alpha = 0.0033$) provides confidence in these results. Only Positivity did not reach significance under Bonferroni correction.

The five themes that were more prevalent in broad prompts (*Positivity*, *Qualifying descriptors*, *Quals Exam*, *Quals Prep*, *Stress*); one theme (*Engagement*) was more prevalent in focused prompts. This pattern suggests that when participants direct conversation, they gravitate toward discussing qualifying exam experiences and stress—topics understandably forefront in their thoughts.

The prevalence of the *Qualifying descriptors* in broad responses merits note. This code was used when interviewees responded with more descriptive language regarding exams, relationships, exam preparation and research activity, et cetera. Other times, the *Qualifying descriptors* code indicates hedging language and expressions of uncertainty about well-being. Their greater prevalence in broad responses suggests that conversational latitude elicits more nuanced, ambivalent assessments rather than definitive statements. This contrasts with advised caution [1] that overly specific probes can lead to overly reflective and intellectualized responses; the pattern here suggests broad prompts may elicit more reflective, qualified responses while focused prompts yield more definitive statements.

That both *Positivity* and *Stress* appeared more frequently in broad prompts (though only *Stress* remained significant after Bonferroni correction) suggests broad prompts allow expression of experiential complexity—both positive emotions and stressors—rather than systematically favoring either valence. This partially challenges finding that broad survey questions yielded "not only richer, but also more positive" responses [2]; in interview contexts, broad prompts elicited discussion of both positive and negative experiences rather

than favoring positive content. Both this and the previous point should be considered in the context that participants in this study were all transiting a high-stress experience.

Engagement—coded where interviewees described active involvement and absorption in research or QE preparation—was the sole theme more prevalent in focused prompt responses. Several factors may contribute. First, semi-structured interviewing involves natural follow-ups; when participants mentioned engagement during broad responses, interviewers likely probed further with focused questions, concentrating engagement discussion in the focused category. Second, direct questions about engagement may be less threatening than broad well-being invitations during stressful periods. When asked focused questions about specific engaging experiences, participants may provide straightforward affirmative responses, whereas broad well-being prompts during qualifying exam preparation may naturally draw attention to challenges. This aligns with research on power dynamics and participant agency [20]—participants may prioritize discussing stress when given latitude while responding readily to direct questions about positive engagement. Third, the constructs examined in this study differ in their temporal stability: stress is highly dynamic and situationally responsive, whereas engagement reflects more stable patterns of involvement in work. This state-trait distinction may partly explain the prompt-type difference observed---broad prompts, which allow participants to direct conversation, unsurprisingly surface the most contextually salient experience (acute pre-exam stress), while focused prompts may be necessary to elicit more dispositionally stable constructs that participants do not spontaneously foreground during high-stress periods. Fourth, focused prompts may elicit concise, quotable statements about engagement rather than extended narrative passages that constitute single sections of meaning. Future work should explore if pattens exist for which interviewees and during which interviews (timeline) *Engagement* and other statistically significant themes were most prevalent and whether prompt-type effects on temporally stable versus dynamic constructs replicate across other study contexts.

No thematic co-occurrence pairs differed in prevalence significantly between prompt types. This suggests that while prompt type may have some influence on which themes are discussed, it does not alter how themes cluster together in participants' responses. When participants discuss *Stress*, regardless of prompt type, they are equally likely to also discuss related themes. This provides methodological reassurance: prompt type steers conversation toward certain themes but appears, from these data, not to fundamentally restructure conceptual relationships between well-being constructs as participants experience and articulate them.

Taken together, these findings suggest that prompt type functions as a conversational steering mechanism rather than a determinant of data quality. Broad and focused prompts do not produce ‘richer’ or ‘poorer’ data; they produce differently distributed data of equivalent informational density. This challenges common qualitative methods assumptions that question format primarily affects response quality, suggesting instead that question format affects topical emphasis while maintaining comparable analytical richness.

Limitations

The sample comprised nine biomedical engineering doctoral students at a single institution. Findings may not transfer to other disciplines, institutional contexts, or doctoral populations. Only well-being sections of transcripts from Interviews 2 and 3 were analyzed; patterns may differ for identity constructs or other theoretical domains. The reclassification by actual

prompt type, while methodologically justified, represents a *post hoc* decision requiring cautious interpretation.

Some degree of temporal confounding likely remains in the data as a result of stress intensification closer to the QE. However, the study design—where participants received both broad and focused prompts within each interview, with alternating sequence order—was chosen to minimize temporal effects. Comparisons between interviews #2 and #3, and between which prompt type came first, were purposefully excluded from analysis because of the likelihood of confounding effects.

This analysis relied on a codebook developed specifically for this dataset through hybrid inductive-deductive coding informed by PERMA-V(N) frameworks. The codes and thematic families that emerged—and consequently, the differences observed between prompt types—reflect both the theoretical framework and the emergent patterns in these specific interviews. A codebook developed with different theoretical commitments or at different levels of granularity might reveal different prompt-type effects.

Conclusion

Neither focused nor broad prompts were shown with these data to produce "better" responses. Prompt type selection should align with research goals and constraints. Focused prompts generate more coded statements per unit time and may be more efficient when duration is constrained. Alternatively, broad prompts likely allow participants to direct conversation toward salient topics.

For doctoral student well-being research during high-stress periods, a hybrid approach may be optimal: begin with broad prompts to build rapport and allow participant-driven disclosure about stress and challenges, then employ focused prompts to ensure systematic coverage of positive well-being constructs like engagement that participants may not spontaneously emphasize. This approach leverages both questioning styles while attending to power dynamics and emotional complexity in interviewing stressed populations [20], [21].

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. Kvale and S. Brinkmann, *InterViews: Learning the Craft of Qualitative Research Interviewing*. SAGE Publications. [Online]. Available: https://books.google.com/books?id=bZGvwsP1BRwC

[2] K. Hansen and A. Świderska, "Integrating open- and closed-ended questions on attitudes towards outgroups with different methods of text analysis," *Behav. Res. Methods*, vol. 56, no. 5, pp. 4802–4822, 2024, doi: 10.3758/s13428-023-02218-x.

[3] J. Ogden and D. Cornwell, "The role of topic, interviewee and question in predicting rich interview data in the field of health research," *Sociol. Heal. Illn.*, vol. 32, no. 7, pp. 1059–1071, 2010, doi: 10.1111/j.1467-9566.2010.01272.x.

[4] J. Walther *et al.*, "Qualitative Research Quality: A Collaborative Inquiry Across Multiple Methodological Perspectives," *J Eng Educ*, vol. 106, no. 3, pp. 398–430, 2017, doi: 10.1002/jee.20170.

[5] M. Borrego, E. P. Douglas, and C. T. Amelink, "Quantitative, Qualitative, and Mixed Research Methods in Engineering Education," *J Eng Educ*, vol. 98, no. 1, pp. 53–66, 2009, doi: 10.1002/j.2168-9830.2009.tb01005.x.

[6] M. Castillo-Montoya, "Preparing for Interview Research: The Interview Protocol Refinement Framework," *Qual. Rep.*, 2016, doi: 10.46743/2160-3715/2016.2337.

[7] M. Bahnson, M. Wyer, and C. Cass, "WIP: Pilot Testing an Interview Protocol to Explore Perceived Bias and Engineering Identity in Graduate Engineering Students," in *2019 ASEE Southeastern Section Conference*, 2019. doi: 10.1109/fie.2018.8658669.

[8] A. Kirn, A. Godwin, C. Cass, M. Ross, and J. Huff, "Mindful Methodology: A transparent dialogue on Adapting Interpretative Phenomenological Analysis for Engineering Education Research," *2017 ASEE Annu. Conf. Expo. Proc.*, 2017, doi: 10.18260/1-2--28669.

[9] 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.

[10] G. Charness, U. Gneezy, and M. A. Kuhn, "Experimental methods: Between-subject and within-subject design," *J. Econ. Behav. Organ.*, vol. 81, no. 1, pp. 1–8, 2012, doi: 10.1016/j.jebo.2011.08.009.

[11] J. Tabora, K. Manion, and D. P. O'Neill, "Reliability of PERMA-V(N): A Qualitative Coding Framework for Capturing Well-being Dimensions in Engineering Doctoral Students," in *(under review) 2026 ASEE Annual Conference & Exposition*, 2026.

[12] H. L. Perkins, M. Bahnson, M. A. Tsugawa-Nieves, A. Kirn, and C. Cass, "Development and Testing of an Instrument to Understand Engineering Doctoral Students' Identities and Motivations," in *2018 ASEE Annual Conference & Exposition*, in 2018 ASEE Annual Conference & Exposition Proceedings. 2018. doi: 10.18260/1-2--30319.

[13] E. P. L. Cisneros and D. P. O'Neill, "Comparing Implicit and Explicit measurements of Engineering and Research Science Identities in Engineering Doctoral Students," in *2025 ASEE Annual Conference & Exposition*, in 2025 ASEE Annual Conference & Exposition Proceedings. Jun. 2025. doi: 10.18260/1-2--29006.

[14] J. A. Henderson, B. L. McGowan, J. Wawire, L. S. S. Benjamin, K. L. Schaefer, and J. D. Alarcón, "Photovoice: Visualizing the engineering identity experiences of sophomore students," *J. Eng. Educ.*, vol. 112, no. 4, pp. 1145–1166, 2023, doi: 10.1002/jee.20555.

[15] J. Tabora, K. Manion, and D. P. O'Neill, "Using Composite Vignettes to Examine Identity and Well-being in Biomedical Engineering Doctoral Students During the Qualifying Examinations Process," in *2026 ASEE Annual Conference & Exposition*, Jun. 2026.

[16] J. Fereday and E. Muir-Cochrane, "Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development," *Int. J. Qual. Methods*, vol. 5, no. 1, pp. 80–92, 2006, doi: 10.1177/160940690600500107.

[17] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc.: Ser. B (Methodol.)*, vol. 57, no. 1, pp. 289–300, 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.

[18] B. DiCicco-Bloom and B. F. Crabtree, "The qualitative research interview," *Méd. Educ.*, vol. 40, no. 4, pp. 314–321, 2006, doi: 10.1111/j.1365-2929.2006.02418.x.

[19] M. DeJonckheere and L. M. Vaughn, "Semistructured interviewing in primary care research: a balance of relationship and rigour," *Fam. Med. Community Heal.*, vol. 7, no. 2, p. e000057, 2019, doi: 10.1136/fmch-2018-000057.

[20] O. Karnieli-Miller, R. Strier, and L. Pessach, "Power Relations in Qualitative Research," *Qual. Heal. Res.*, vol. 19, no. 2, pp. 279–289, 2009, doi: 10.1177/1049732308329306.

[21] E. J. Alessi and S. Kahn, "Toward a trauma-informed qualitative research approach: Guidelines for ensuring the safety and promoting the resilience of research participants," *Qual. Res. Psychol.*, vol. 20, no. 1, pp. 121–154, 2023, doi: 10.1080/14780887.2022.2107967.