

Conducting the Cognitive Interview: Sharing Experiences and Insight from Two Think Aloud Studies

Introduction

This theory paper focuses on a research methodology, using an autoethnographic approach to reflect on the use of cognitive interviewing (CI) as a method of increasing the quality and validity of questionnaires in pre-validation design and development stages. We first provide a brief review of cognitive interviewing, sometimes called “cognitive think-aloud interviewing” or “think-aloud interviewing,” before presenting a summary of two studies conducted by the authors that used CI. Differences between these two studies are discussed as comparative cases and advice is given to scholars considering the use of CI in their own research. While this paper is not an explicit guide to conducting CI, we do intend to provide advice and wisdom for researchers who are unfamiliar with CI as a method, grounded in our experience with the method. This paper is written with a particular focus on the use of CI in engineering education research (EER) but may be more broadly applicable to other social sciences domains.

To the knowledge of the authors, there is not a set of guidelines available for engineering education researchers interested in conducting CIs. Guidelines and public discussions of the advantages and drawbacks of methods and specific aspects of methodological use may be of particular importance for new researchers and for establishing standards of quality for survey measures within EER. While we acknowledge that the exercise of judgement by researchers is important and do not intend to offer a definitive guide for CI in EER, we hope that this work might increase the prevalence of conversations about research methods in conferences in our field, including guidance which supports new researchers, researchers new to CI as a method, and which supports the standards of quality and validity of research in EER.

Review of Cognitive Interviewing

Cognitive interviewing is a process by which one collects verbal information about how participants understand a topic or answer a question. CI can be used to determine whether items on a questionnaire are understood as intended, generate the desired information, and reflect participant experiences and understanding [1]. CI is most commonly used for the pretesting of surveys and can be used for educational research, but the method can also be used to understand cognition, such as by having participants think aloud while problem solving [1]. For the purposes of this paper, we will consider CI as a means of pretesting a survey for research, using CI for validation purposes while the survey is in drafting stages. CI can be used as an independent method, such as during think aloud studies of educational materials, however we are most concerned here with CI techniques which impact the design of written surveys for research. As a technique, CI has seen widespread use in this way across many fields including psychology, education, business, and political science [2]. In EER, CI has been used in this way to improve the design of measures of many topics, including professional skills development [3], social capital resources [4], and student responses to instructional strategies [5].

Cognitive interviewing requires participants to think aloud while completing a task. Drawing from reviews of the method, we here define *thinking aloud* as “requesting participants to openly reflect on their answers to survey questions and the processes by which they reach those answers, with limited interviewer interaction.” [1], [6], [7]. CI interviewers need not necessarily follow a uniform format; these researchers may choose to engage with participants via *concurrent probing*, where questions are asked during the interview in a process guided and usually pre-determined by the researcher and/or *retrospective probing*, where questions are asked in a debriefing session after the participant finishes taking the survey [7].

Critics of CI share concerns that task performance may be decreased by thinking aloud while also answering questions, leading to inauthentic results [8] [9] or may bias results by probing participants to think more intentionally than they may in a true survey setting [10]. While these criticisms are open to discussion for instructional or problem-solving settings, CI remains an important technique to explore individual cognition and processing [11] and to capture how individual aspects such as sociocultural contexts or language background might influence participant responses based on their lived experiences [12]. We consider CI to be particularly valuable for research studies which measure affective, attitudinal, motivational, cultural, and experiential research questions in engineering education environments.

At the time of the first of the two study design cases presented, the first author was a novice researcher who wished to ensure the maximization of reliability (and related, the minimization of variance between responses and interpretations) of the survey items he was drafting. In particular, this author was concerned with several questions, many of which involved achieving *content validity* [13], namely that the questions were related to the constructs they aimed to describe in the ways we intended. CI is particularly useful for providing evidence of content validity. Additionally, participant responses can reveal if questions are *double-barreled*: that there may be unintended second meanings or interpretations of the items. Finally, participant responses can reveal differences in how words are interpreted (e.g., is treating teaching like a duty always construed as negative?). In conducting CI during the draft stages of the two studies presented, the authors’ goals were to ensure that the items were *valid*, *easy to interpret* and *understand*, and that the survey items were *clear*. Further, we endeavored to be sure that the items were *relatable* and *relevant* to participants’ experiences.

Cognitive interviewing can be used to achieve the goals above, but it can also be used as an independent method of data collection or to support other analytical techniques. For example, qualitative analyses of CI data in EER, such as those in a recent paper by Li et al. [3] can avail researchers of important differences between groups (e.g., demographic groups) of participants with relevant experiences or identities and how these groups might trend differently in survey responses. Triangulating interview analyses with responses is another opportunity to analyze CI data. And as presented below, CIs of surveys which are designed using prior data can be useful tools for member checking in sequential mixed methods studies.

Research Questions

The research questions guiding this study represent the experiences of the authors while navigating the methods of a commonly used technique and is designed to be didactic in terms of research methods in EER:

RQ1: How can conducting cognitive interviews inform changes to survey questions?

RQ2: What lessons and practices can be shared from attempting different approaches to conducting cognitive interviews? Namely, how does the study design influence the outcomes of cognitive interviews?

RQ3: How does the number of interviewers simultaneously conducting a cognitive interview affect the quality of the cognitive interview method?

Researcher Positionality

While we do not present an analysis of data in this paper, the authors agree with recent calls in the field for engineering education scholars to include positionality statements in EER work, regardless of methodologies used (e.g., in [14], [15]). The first author is a graduate student in an Educational Psychology program whose research has primarily focused on EER. He led the design of the survey drafts and CI testing for both studies. At the time of the first study, he was more familiar with qualitative social sciences methods compared to quantitative methods, and consequently found CI to be an attractive validity-seeking method. The second author is a faculty member researching EER and is the primary investigator of the first study and a co-PI for the second study, who had not advised students conducting CIs at the time of the first study. The third author is a faculty member in an Educational Psychology program and the primary investigator of the second study. She is an experienced researcher who has used CI techniques on many projects and offered her own wisdom during the second study. The fourth author is an undergraduate engineering student who conducted many of the interviews in the first study together with the first author. The first and fourth authors were conducting CIs for the first time during the study and took extensive field notes to document their process. In the first study, the first and fourth authors were concerned about developing a survey as novices and lacking intuition about survey design, and they were encouraged by the second author to explore and document the methodologies they used. In both studies, the faculty team members met at least weekly with the student team members during all study stages to offer advice and provide accountability for progress. The first, second, and fourth authors have collaborated on previous reviews of methods within their own research experiences, driving them to consider comparisons between approaches to research methods [16]. The first and second authors made mid-career transitions to partake in EER scholarship at an institution without many formal EER resources. Thus, this research, and any work which provides resources for new or learning EER scholars, is of particular value to them. The authors acknowledge that their various stages of learning influenced their comparisons of methods, and that their desire to reflect on their learning is influenced by their own recognition of unequal access to professional training materials in EER methods across institutions – for example, many institutions do not have formal EER coursework or departments.

Methods

Our teams conducted two CI studies on similar exploratory survey drafts. Both studies sought to measure stressors for engineering students, one study for undergraduate students

and one for graduate students. Both studies were conducted at the same large, Midwestern university with a large college of engineering. The research design and draft measure of both studies were approved by the site university's Institutional Review Board before data collection began. A sample of $N = 13$ undergraduate, engineering degree-seeking students, sampled until saturation was reached, completed the first CI study and a sample of $N = 13$ doctoral engineering students, also sampled until saturation was reached, completed the second study. While both studies sought to recruit between 10 and 20 students for cognitive interviews, the equal number of participants between the studies is likely coincidental, relating to a nexus of factors including survey length, participant experience, and the timing of the studies.

Descriptions of the first sample and the development of the measure tested in Study 1 have been described previously [16], [17]. Descriptions of the development of the measure tested in the second study are also published separately [18].

The Cognitive Interview Process. The process we used for conducting cognitive interviews involved multiple steps, which we will articulate within this sub-section. The literature on the CI method provides helpful advice for scholars new to the CI process (e.g., [1], [2], [6], [7]). Before interviews were conducted, we drafted a set of survey items and organized them into related sections which we expected would group as latent factors once data from a large sample was collected. We then created and familiarized ourselves with a semi-structured cognitive interview protocol, which included examples of both *concurrent* and *retrospective probing* questions to ask participants. Example concurrent probing questions are: "What do you think this question is asking you?" and "Is there a different way you would respond to this question than with the [Likert-type] options presented?". An example retrospective probing question is: "Was this section relevant to your experiences as a doctoral student?". The two cognitive interviewing protocols used across the two studies were similar; a sample of introductory text and probing questions for a cognitive interview is provided in our Appendix, using our materials from Study 2.

In the next step, we determined the format of interviews, including the delivery mechanism (e.g., would the questions appear in a text document or in a sample survey environment?), the number of interviewers and roles of the interviewers, and the setting (e.g., in-person, Zoom) for the CIs. Participants were then recruited, given consent information, and scheduled to participate.

During the scheduled interview times, in an initial briefing, the participants were asked to record verbal consent, given information about the study including instructions on how to think aloud and respond, and given an opportunity to ask clarifying questions about the CI method. Then participants were shown the survey and responded aloud to how they would answer survey items. The interviewer(s) asked concurrent probing questions as participants answered these questions aloud, and at the end of each section of related items, the interviewers asked broader retrospective questions. These questions included assessing the overall clarity and design of the study, the relevance of the survey to participants' experiences, and the completeness (e.g., was any very relevant question missing?) of survey items. At the end of the interview, the interviewer(s) debriefed the participants by asking broader retrospective questions, thanking participants, and providing information

about compensation and future contact.

During the interview, the interviewer(s) collected detailed field notes. When there were two interviewers in Study 1, these notes were compared at the end of the interview in a short debriefing meeting. The interviewer(s) reflected about the interview, adding additional notes, and then made minor edits to the survey (e.g., grammatical clarifications) if making those edits would clearly improve the survey. Thus, the survey was iteratively adjusted so that new information could be produced in subsequent surveys. Finally, after the interviewer(s) felt that they were receiving similar feedback in repeated interviews, the studies were closed, and edits were made to form the final survey.

Table 1. The Cognitive Interview Process

Phase	Specific Steps	Notes on Steps
Before the interview	Draft survey items	If necessary, divide items into expected latent factors to best organize the survey
	Draft Cognitive Interview protocol	Draft introductory (briefing) text, consider how to address potential questions which may come up, familiarize yourself with the protocol
	Determine the format of interviews	Including the delivery mechanism, number and roles of interviewers, and setting
	Recruit participants to the study	Give study and compensation information, consent information, and schedule participant participation
During the interview	Initial briefing	Introduce participant to the study purpose, collect verbal consent, ensure participant understands to think aloud
	Provide survey and ask concurrent probing questions	As participants read the survey and answer the questions aloud, prompt participants for items which are confusing and to clarify how items are being interpreted
	Retrospective probing questions	At the end of sections of the survey or the entire survey, interviewer(s) ask questions about the clarity, relevance, and completeness of the survey
	Debriefing	Thank participants, ask for questions, give compensation information
After the interview	Record field notes	Take reflective notes and review notes taken during the interview, if applicable, compare across interviewers
	Edit the survey	Make changes to the survey iteratively based on interview findings and continue the CI process. If the researchers feel saturation is achieved, finalize the survey and stop collecting cognitive interviews

Study 1. The first study was a measure of stress culture for undergraduate engineering programs [16], [17]. Eighty-five items were developed using mixed methods research methods, including analysis of interviews with engineering undergraduate students and a

review of prior measures of culture and belonging for engineering students. The survey used a six-point Likert-type scale (Dis/agree, slightly and strongly dis/agree) with a 7th “no basis for judgement” item intended as a validity check to be used to remove items which did not often reflect participant experiences. The authors anticipated the survey would show 7-11 latent factors related to sources of stress in the later validation study, e.g., professors as a source of stress, classmates as a source of stress, etc. Items were both added and removed during the cognitive interview process and the final survey was 81 items long.

Participants were contacted in winter 2020-2021 via a newsletter email distributed to the site institution’s college of engineering. Participants registered their interest via a response to a brief survey form and scheduled their participation using Doodle.com. All participants were compensated with \$10 Amazon gift cards. Participants connected to interviews via Zoom and were audio recorded. Interviews lasted an average of 30 minutes each.

The CI protocol included an initial briefing, recorded documentation of consent, explanation of the purpose of the study, and request for participants to *think aloud* while answering questions. To ease any potential cognitive burdens associated with replying, participants were not required to read questions aloud unless they preferred to. However, they were asked to justify their responses and to note any confusion about their interpretation of items. Instructions we gave to participants on how to think aloud were generally made to be vague so as to not bias participant thinking and to allow participants to engage with the items as authentically and naturally as possible. Items were divided into short sections grouped by anticipated factors. A protocol of *retrospective probing* debriefing questions was written for each section, including questions about how relevant the questions were to participants’ experiences, if any confusing items/item wording or grammatical errors were noted, and if any related topics were missing from the set of questions and thus not present in the survey.

With two exceptions due to scheduling issues, interviews were conducted by two interviewers in the first study. The first author conducted all interviews and was responsible for all email correspondences with participants for scheduling and consenting, and verbally introduced the interviews at the beginning of each interview section. The first author guided the participants through each section and asked *retrospective probes* after each section. The fourth author was present for nearly all of the interviews in the first study and took typed notes including participant feedback or confusion regarding items and trends in participant responses. Having access to the notes from the current and prior interviews, the fourth author asked *probing questions* concurrently to participation, allowing us to recall and ask about specific questions or words that had been misinterpreted by other participants.

Between interviews, iterative improvements of survey questions were made. Most frequently, grammatical adjustments to survey items were implemented between surveys. However, item addition or deletion or significant rewording of items was not done until meetings and discussions with the full project team. We decided this to save interview time on simple edits to the survey but to make sure that multiple participants experienced similar serious misinterpretations or lack of relevance to items we modified or removed, to better our understanding of how different participants understood the survey items.

Finally, during the interviews, we shared a text document containing the current draft of the

survey, where each section of questions was spaced apart to show where participation would be paused.

Study 2. The second study was a measure of sources of stress in doctoral engineering programs [19]. Sixty-five items were developed from interview data with doctoral students at the same institution in the previous academic year. Items were given on two response scales: an eight-point frequency scale ranging from never to daily and a six-point intensity scale ranging from no stress to extreme stress. An additional “does not apply/no basis for judgment” item was included as we did in Study 1. We anticipated 12 latent factors in the planned validation study, with approximately five to six questions per factor. Factors were based directly on themes from prior interview data. Leveraging our experiences with the first study, each item began with identical question stems “I feel stress when” to minimize the variance in participants’ interpretation of items. The same number of items were added to and subtracted from the survey during the cognitive interview process, resulting in 65 final items.

Participants were contacted in summer 2022 via an email distributed to doctoral students in several registered student organizations at the site institution’s college of engineering. $N = 7$ participants from the sample had provided interview data in a prior study used to design the survey. These participants were asked questions about the alignment of the interview themes with their prior interview participation as a means of member checking the interview results used to construct the survey. Participants registered their interest via an email response to the call for participation and scheduled their interviews via Doodle. All interviews were conducted on Zoom and participants were given \$30 Amazon gift cards for their participation for interviews that averaged 57 minutes each. These participants were given gift cards of a higher value due to two factors: (1) the additional duration of interviews for this study compared to the first and (2) the larger budget available for study participants within the second project.

The interview protocol used was nearly identical to the one used in Study 1; returning interview participants were asked additional questions about the relationship of items to their previous interviews as a means of member checking.

Unlike the interviews in Study 1, these interviews were conducted only by one investigator. Participants were enrolled into a space on the Canvas learning management system and responded to a draft of the survey in an environment similar to what participants in the full survey release would receive. The interviewer asked no *concurrent probing questions* during the first few interviews and later used field notes to develop prepared probing questions about specific problematic or questionable items but asked *retrospective questions* after each section. Similar to in Study 1, iterative improvements of the survey in the form of small edits were made between interviews.

Findings

Here we summarize the changes made to surveys as a result of the two CI methods. In particular, we share several aspects of each study which were influenced by study design decisions including changes to the wording of items, the impact of the number of interviewers, the impact of the setting of the survey, and the advantage of designing CI studies to member check prior qualitative research. We describe changes which impacted the clarity of items, the identification of double-barreled questions, confusion or inconsistencies in

wording, grammatical conventions, and implications of the delivery mechanism. The changes described exemplify changes relevant to our first research question: *How can conducting cognitive interviews inform changes to survey questions?*

Clarity of Question Subjects. One consistent change required in both studies was to make the subjects of questions more clear. For example, in Study 1, one question stated, “My professors design courses to weed out weak students.” One participant said “*Calc II definitely felt like a weed-out course to me,*” suggesting that participants could misconstrue the scope of the survey to include topics outside of engineering. Thus, we decided to change “my professors” to “engineering professors” or “professors in my engineering department.” Although this misinterpreted scope occurred only in a few questions during interviews, we decided to change all questions to be explicit about experiences in engineering environments to ensure that other experiences were not captured by mistake [16], [17].

In Study 2, some questions were asked about research, and we used the word “labs” to refer to research groups, as the majority of engineering research groups at the site institution conduct bench science research. Participants whose work primarily focuses on coding, education, or theory wondered if these questions applied to them and we changed the word “labs” to “research groups” throughout the study.

Double-Barreled Questions. In some cases, participants exposed double-barreled questions. The first of these items found in Study 1 was: “Engineering professors and TAs expect students to compete in class.” One participant responded to this question with confusion, saying (paraphrased to protect participant privacy): “*I would answer this differently if you said engineering professors for one question and TAs for another one. I don’t think TAs do that, but I think some professors think that – that it’s helping us prepare for the real world.*” While the intent of the question was to ask about competition in courses as a stressor, some participants recognized that the attitudes of TAs and engineering professors towards students were different in their experience. As a result, we split questions which asked about multiple instructor types into two questions.

Another double-barreled question in our Study 1 draft was: “Engineering students do not prioritize their health and wellness.” Some participants suggested answers related to physical health, others mental health. Ultimately, we decided the information coming from this question was too inconsistent and thus asked a new question specific to mental health.

In Study 2, we had one unanticipated double-barreled question: “I feel stress when I feel unprepared for completing my coursework.” Participants answering that question considered topics such as meeting course prerequisites or keeping up with reading week-to-week, two very different constructs.

Resolving Other Confusing Question Elements. In Study 1, we noticed that grammatical conventions could also lead to confusion during cognitive interviews. Many questions began with the words “It is,” such as in the item: “It is normal for students to stay up all night doing work.” However, some non-native English speakers, familiar with the formal interrogative structure “Is it” to begin sentences, thought aloud and read the items as questions rather than statements, e.g., “Is it normal for students to stay up all night doingwork?” These questions can have different meanings; one asks participants to make a judgement while the other

identifies a norm. We removed this stem across all questions, e.g., this question became: “Engineering students commonly stay up all night working.”

In Study 2, aspects of the graduate experience such as work-life balance and milestones could cover a variety of topics; we provided parenthetical examples to some questions to clarify what sorts of phenomena were intended.

When completing a section on microaggressions in doctoral engineering spaces, one international participant in Study 2 said: “*My home country does not have a concept of microaggressions, I didn’t understand them when I came here and I don’t really get them still,*” leading to our decision to add a definition to microaggressions and an example in the survey.

Implications of Delivery Mechanisms. By sharing the questions on the Canvas LMS in Study 2 during the interviews, the same setting in which the survey would be delivered, we were able to iteratively make changes after interviews by adjusting the instructions and appearance of the survey in addition to making changes to the items, a difference between the changes resulting from the CIs in Study 1, where the questions were presented using a text file. During Study 2, we noticed that participants were likely to skip instructions at the top of the page but would be more likely to read instructions which came adjacent to a question, visibly changing the amount of text near the question. Participants in late stages of their PhDs also were unsure of how far back to think when considering the answers to questions about their experiences in their programs, causing us to add requests for them to think back only to recent experiences, and thereby increasing the validity of the measure.

Discussion

CI as a Means of Member Checking. The stressors in Study 2 were described by participants during a longitudinal study. This study consisted of one year of repeatedly surveying and interviewing participants about stress, and then analysis of this year of data to find the most frequent and severe sources of stress and construct a survey about these stressors. Participants returning to the study to conduct a CI were asked to confirm that the themes and questions of the survey aligned with the topics discussed during their interviews. Participants emphatically agreed with the topics of the survey and the major themes of interview findings, with one saying, “*I feel like this survey was written for me specifically.*” Cognitive interviewing when methods are mixed in this way for the purposes of design is an effective technique to “double dip” by member checking the qualitative findings used to develop a larger scale survey, triangulating between these methods, and providing validity evidence both retrospectively to the prior interviews used to develop the survey and additionally to the measure being developed. Further, inviting past participants to conduct a CI is beneficial, for the investigator may have an established rapport with the participant making the interview easier and more natural to conduct.

Comparisons Between the Two Study Cases. While both CI studies focused on novel surveys of similar lengths and topics, the differences between the two studies in terms of the interview setting and the number of interviewers had implications on the findings.

Contributing to research question two, *What lessons and practices can be shared from attempting different approaches to conducting cognitive interviews? Namely, how does the study design influence the outcomes of cognitive interviews?*, in Study 2, the use of the same LMS space where the survey would eventually be distributed allowed us to assess the effectiveness of the instructions for the survey and observe the survey more authentically and thereby increase measure validity. While it was more challenging to prepare the draft survey for CI testing in an LMS space compared to sharing the interview questions in a text document, we believe that it is clearly advantageous to attempt to conduct CIs in settings as close to the actual interview setting (e.g., using Qualtrics or an LMS, using pen and paper, etc.) as possible.

Contributing to research question two as well as question three: *How does the number of interviewers simultaneously conducting a cognitive interview affect the quality of the cognitive interview method?*, the choice of the number of interviewers was a particularly salient factor in the execution of the studies. Most importantly, we found it easier to ask *concurrent probing questions* in the middle of the interviews with the help of two researchers. However, in our longer interviews with one researcher, we were able to ask more questions and complete more sections, leading to more productive interviews. From a discussion among the authors, Table 1 provides benefits and drawbacks to one or two interviewer CI formats.

Table 2. Pros and Cons for One and Two Interviewers for Cognitive Interviews

One Interviewer		Two Interviewers	
Pros:	Cons:	Pros:	Cons:
Scheduling with participants is easier	More cognitive note-taking burdens	Easier to record thoughts <i>in situ</i> and take notes	Scheduling with participants is more challenging
Consistent methodology: all interviews directed by one person	Splitting interviewer's focus across conducting the interview and thinking of questions	Enables delegation of tasks (managing the flow of the interview, asking probes)	Potential for more variance between interviews based on multiple interviewers
More comfort and more rapport for participants (no feelings of being "double teamed")	A single interviewer may miss an important insight while managing the interview	More researchers can potentially identify more potential problems with items	Participants may be less comfortable sharing
Interviews may be quicker or more efficient	Interviewer is susceptible to unexpected cancellations	Potential technological issues or conflicts of one interviewer do not necessitate cancellation	More probing questions may yield longer interviews

Table 2 reflects the authors' perspective that there are many advantages to both one and two interviewer CI studies. In addition to these considerations, other factors may impact the choice of one or two interviewers. If the potential interviewer(s) have relative power over the interviewees (e.g., if the interviewers are faculty while the interviewees are undergraduate students), it may be more advantageous to have only one interviewer, which may be less intimidating. Additionally, it may be more advantageous to only have one interviewer if the topic of the survey is uncomfortable or sensitive. However, if the two prospective interviewers may be of different genders or races, this may help participants of diverse backgrounds feel more comfortable. If the interviewers are relatively new to CI or even interviewing as a method, multiple interviewers may also help the interviewers to feel comfortable, leading to improvements in outcomes. Additionally, multiple interviewers may be useful if one interviewer is learning methods, as this interviewer can be directly mentored and given feedback. If the survey designers are interested in large amounts of item-level feedback, two interviewers, which may be conducive to more concurrent probing questions, may be more appropriate.

Authors' Reflection on the Benefits of the Method. We found that CI was an effective way to test our survey design, in particular when it came to language and cultural differences with international participants. Our project team does not include an international investigator, however, by including participants' experiences in a survey draft, we were able to find words or phrases which were confusing or idiomatic and adjust them.

Another satisfying element of CI is the degree of collaboration between participant and researcher. By requesting feedback on the research product, the participant becomes an active participant in the research process, translating their experiences to design. For studies of experience, attitude, and culture, we find that CI is an excellent means of increasing participant voices and utilizing participants' expertise in their own lived experiences.

CI is particularly useful in sequential mixed methods designs, where interviews can be used to inform the design of surveys. While these interviews are less open-ended or phenomenological than typical styles of qualitative research, they allow researchers to understand additional context related to the topics they are investigating. This helps not only to provide validity evidence to the survey topic through the CI process itself, but also to triangulate data sources to provide further evidence of understanding the topic being studied.

Finally, CI can help researchers to build an intuition for what factors might be predictors of interview response trends. For example, in Study 2, we noticed that the descriptions of stress for classes and milestones were described much more severely by PhD students who were relatively earlier in their programs. This correctly predicted what we later found in the survey: that year inprogram correlated positively with those specific stress subscales.

Conclusions

Cognitive interviewing is an effective way of providing evidence of the content validity of survey questions during the survey drafting process. Beyond uncovering confusing, poorly worded, or erroneous items, CI can allow researchers to identify unanticipated double-barreled items, misinterpretations due to culture and identity which would otherwise be difficult to predict and build intuitions for their survey and its trends. Designing CI to occur

in the setting (e.g., Qualtrics) where future participants will take the survey, rather than with a draft of the survey can also improve design outcomes. Further, depending on the type of probing questions, topic, and participants desired, one or two interviewers may be preferred by the investigator.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Numbers 1943541 and 2034800. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank the survey participants for their insights and contributions to our research. The authors thank Study 1 advisory board member Dr. Alison Godwin, who provided an example draft of a CI protocol which we closely followed in developing the protocol for our initial Study 1 cognitive interviews.

References

- [1] P. C. Beatty and G. B. Willis, "Research synthesis: The practice of cognitive interviewing," *Public Opinion Quarterly*, vol. 71, no. 2, pp. 287-311, 2007.
- [2] R. Tourangeau, L. J. Rips, and K. Rasinski, "The psychology of survey response," *Cambridge University Press*, Cambridge, UK, 2000.
- [3] T. Li, E. Holloway, K. Douglas, J. Martin, and V. Bill, "WIP: Think-aloud interviews for assessment of engineering students' opportunities to practice professional skills," *Paper presented at 2022 ASEE Annual Conference & Exposition*, Minneapolis, MN, 2002.
- [4] J. M. Trenor, M. K. Miller, and K. G. Gipson, "Utilization of a think-aloud protocol to cognitively validate a survey instrument identifying social capital resources of engineering undergraduates," *Paper presented at 2011 ASEE Annual Conference & Exposition*, Vancouver, BC, CA, 2011.
- [5] M. DeMonbrun, C. J. Finelli, M. Prince, M. Borrego, P. Shekhar, C. Henderson, and C. Waters, "Creating an instrument to measure student response to instructional practices," *Journal of Engineering Education*, vol. 106, pp. 273-298, 2017.
- [6] G. B. Willis, S. Schecter, and K. Whitaker, "A comparison of cognitive interviewing, expert review, and behavior coding: What do they tell us?" in *American Statistical Association, Proceedings of the Section on Survey Research Methods*, Indianapolis, IN, 2000.
- [7] G. B. Willis, "Cognitive interviewing," *SAGE Publications, Inc.*, 2005. <https://dx.doi.org/10.4135/9781412983655>
- [8] M. C. Fox, K. A. Ericsson, and R. Best, "Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods," *Psychological Bulletin*, vol. 137, no. 2, pp. 316-344, 2011.

- [9] J. Russo, E. Johnson, and D. Stephens, "The validity of verbal protocols," *Memory and Cognition*, vol. 17, pp. 759-769, 1989.
- [10] J. W. Schooler, S. Ohlsson, and K. Brooks, "Thoughts beyond words: When language overshadows insight," *Journal of Experimental Psychology: General*, vol. 122, pp. 166-183, 1993.
- [11] G. B. Willis and A. R. Artino Jr., "What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys," *Journal of Graduate Medical Education*, vol. 5, no. 3, pp. 353-6, 2013.
- [12] E. R. Gerber, "The view from anthropology: Ethnography and the cognitive interview," in *Cognition and Survey Research*, M. Sirken, D. Herrmann, S. Schechter, N. Schwarz, J. Tanur, and R. Tourangeau, Eds. New York, NY: Wiley; 1999. pp. 217–234.
- [13] S. G. Sireci, "The construct of content validity," *Social Indicators Research*, vol. 45, no. 1/3, pp. 83–117, 1998.
- [14] A. Godwin, "Sitting in the tensions: Challenging whiteness in quantitative research," *Studies in Engineering Education*, vol. 1, ed. 1, pp. 78–82, 2020.
- [15] C. Hampton, D. Reeping, and D. S. Ozkan, "Positionality statements in engineering education research: A look at the hand that guides the methodological tools", *Studies in Engineering Education*, vol. 1, ed. 2, pp. 126–141.
- [16] J. Mirabelli, K. Jensen, S. Vohra and E. Johnson, "Exploring the exploratory factor analysis: Comparisons and insights from applying five procedures to determining EFA item retention," *Paper presented at 2022 ASEE Annual Conference & Exposition*, Minneapolis, MN, 2022.
- [17] K. Jensen, E. Johnson, J. Mirabelli and S. Vohra, "CAREER: Characterizing undergraduate engineering students' experiences with mental health in engineering culture," *Paper presented at 2022 ASEE Annual Conference & Exposition*, Minneapolis, MN, 2022.
- [18] K. Jensen, S. R. Vohra, J. F. Mirabelli, A. J. Kunze, I. Miller, and T. E. Romanchek, "CAREER: Supporting undergraduate mental health by building a culture of wellness in engineering," *Paper presented at 2021 ASEE Annual Virtual Conference*, 2021.
- [19] J. F. Mirabelli, J. Cromley, K. Jensen, D. Robbennolt, and A. Hart, "Work in progress: Exploring the landscape of stressors experienced by doctoral engineering students," *Paper presented at 2023 ASEE Annual Conference & Exposition*, Baltimore, MD, 2023.

Appendix: Cognitive Interview Think Aloud Protocol (Study 2)

Opening Interview Briefing:

Our research team is developing a survey questionnaire that measures doctoral student experiences with stress in engineering programs. The goal of the survey is to measure how the experience of different stressors may correlate to students' intention to persist in engineering programs. In this phase of the research, we are seeking your input on the relevance and clarity of our draft of survey items.

We have a preliminary survey that we would like you to give us feedback on. It is important for our research that we design survey questions that can be understood by all respondents. We want to check two things in this interview:

1) That the way we have designed the survey and asked questions that translate to your experiences

And 2) That there are no errors on the survey that will impair your understanding of the questions (e.g., word choice, typos, design, etc.)

You will be asked to "Think Aloud" as you take the survey. We want to understand what you are thinking as you go through the questions. Anything is relevant. Even the smallest comment you have can mean there's an improvement to be made in our design. Please say anything as it comes to mind. In addition to saying out loud your interpretations of the questions, we also ask that you point out any typos, grammatical errors, confusing wording or anything that throws you off for any reason. I will ask minimal questions at the end of every survey section and may prompt for additional information as you go through the survey.

We'd like you to complete the survey section by section. There are a total of 10 sections, and we'd like you to stop once you've reached the end of each section. If we reach the end of a section and it is near the end of the survey time, we will stop there – it is okay if we do not complete the entire survey.

Concurrent Probing Questions:

What are you thinking about now?

What do you think this question is asking you?

Who/What do you think this question is about?

How do you think you should answer this question?

How do you think your peers might answer this question?

Is this question confusing? (Avoid asking if the participant is confused).

If so, what would make this question less confusing?

How did you arrive at that answer?

What does (a particular word/concept) mean to you?

Do the answer options in this question make sense to you?

Is there a different way you would respond to this question than with the options presented?

Retrospective Probing Questions:

Was there anything missing for this section?

Was there anything worded strangely, or difficult to interpret in this section?

Was this section relevant to your experiences as a doctoral student?

Are there any additional experiences that we should have asked about? [If so:] How do you suggest we ask about that?

Did you feel this survey served the purpose described at the beginning of this interview?

Do you have any additional feedback/questions?

Is there anything else we should ask about?