AC 2011-925: UTILIZATION OF A THINK-ALOUD PROTOCOL TO COGNITIVELY VALIDATE A SURVEY INSTRUMENT IDENTIFYING SOCIAL CAPITAL RESOURCES OF ENGINEERING UNDERGRADUATES

Julie Martin Trenor, Clemson University

Julie Martin Trenor, Ph.D. is an assistant professor of Engineering and Science Education with a joint appointment in the School of Materials Science and Engineering. Her research interests focus on social factors affecting the recruitment, retention, and career development of under-represented students in engineering. Dr. Trenor is a recent NSF CAREER award winner for her research entitled, "Influence of Social Capital on Under-Represented Engineering Students Academic and Career Decisions."

Matthew K. Miller, Clemson University

Matthew K. Miller, M.S., is a doctoral research assistant in the Engineering & Science Education department at Clemson University. He received his undergraduate and Master's degrees in Industrial Engineering from Clemson University.

Kyle G. Gipson, Clemson University

Kyle Gipson is a doctoral candidate in the School of Materials Science and Engineering at Clemson University. His research is focused on polymer nanocomposites for optical materials and increasing diversity within STEM disciplines through the creation of inclusive learning environments. Mr. Gipson received a B.S. in Physics from Wofford College and he also received his Master's of Science in Textile Technology from the Institute of Textile Technology.

Utilization of a Think-Aloud Protocol to Cognitively Validate a Survey Instrument Identifying Social Capital Resources of Engineering Undergraduates

Abstract

The use of verbal report (e.g. "think-aloud") techniques in developing a survey instrument can be critical to establishing an instrument's cognitive validity, which helps ensure that participants interpret and respond to survey items in the manner intended by the survey designer(s). The primary advantage of utilizing a verbal cognitive validation protocol is having evidence that survey items are interpreted by participants in the same way the researcher intended before the instrument is administered to a large sample. Think-aloud protocols have been used to accomplish different goals in a variety of fields, including engineering education where think-alouds are commonly used in problem solving research. While think-alouds have been used by engineering education researchers, the engineering education literature includes few resources for researchers regarding the use of these protocols with respect to large-scale survey development and refinement. In this paper, we present a protocol based on elements of think-alouds conducted inside and outside the engineering education domain. By presenting results and examples from our own experience suing this protocol, we aim to provide a cognitive validation model which may be useful to engineering education researchers designing their own survey instruments.

By following the model outlined in this paper, participants in our study verbalized several issues of concern when interacting with our web-based survey. These issues ranged from minor grammatical errors to serious cognitive mismatches which caused participants to interpret and/or respond to items differently than we intended. Participants were asked for suggestions to correct these issues, and changes were made to the survey based on this feedback. The survey was re-tested in two additional iterations of think-aloud sessions with new participants to ensure the revisions successfully remedied the issues encountered by previous participants. Finally, the refined survey was pilot tested and subsequently reviewed by an expert in the field before being administered at seven institutions. This paper includes evidence and specific examples of how the cognitive validation model resulted in a refined survey instrument, as well as recommendations for other engineering education researchers wishing to employ similar techniques in designing and validating survey instruments.

Introduction and Motivation

Much of the extant literature surrounding the establishment of reliability and validity for survey instruments largely focuses predominately statistical methods to establish such measures as construct or internal consistency within an instrument, which requires the use of rigorous statistical methods to compute coefficients such as Cronbach's alpha to verify that the instrument has achieved at least a minimum acceptable level of reliability and/or validity (e.g. Eris and colleagues[1]). Such statistical methods can establish a case for whether or not the instrument consistently and appropriately measures participant responses to items by measuring a variety of constructs, which include (but is not limited to) ensuring the items within the instrument have an appropriate coverage of the relevant content, are scored or evaluated consistently, and/or are responded to in a consistent manner by participants when interacting with a similar item[2]. While a survey instrument may provide technically consistent, reliable, and/or valid responses after undergoing rigorous statistical validation techniques, without considering **cognitive validity** it is possible that the survey could fall victim to participants responding to an item consistently but with an interpretation of the question that is different than what the survey designers had intended.

Ultimately, of course, the nature of a research instrument and goals of the research study determine what validation techniques are most appropriate. The engineering education literature, as consistent with the research of many other fields, contains many studies that document quantitative statistical analyses assessing an instrument's validity[3,4]; however, there is little guidance in the literature to help researchers establish whether or not the interpretation of an instrument is consistent with its intended design by identifying and understanding participants' thought processes that take place when responding to each item on the instrument. Other fields have used a variety of techniques falling under the scope of Verbal Report Methods (VRMs), in which subjects are asked to provide constant verbal feedback while performing a task[5,6]. VRMs have also been used to establish a case for the cognitive validity of various quantitative instruments, but there is little guidance in the literature for designing a complete cognitive validity evaluation of an instrument, particularly for instruments that measure self-report items.

We aim to describe a particular use of VRMs (specifically, think-aloud sessions) in this paper as an example of its potential utility for other researchers in engineering education by presenting an application of this model to our specific engineering education research project (details of which are given later in this paper). It is a primary goal of the current study to construct a model that engineering education researchers (and other relevant fields) can use to establish a case for the cognitive validity of their survey instruments, giving confidence that participants will have the intended interpretations of each items before the instrument is released on a larger scale for data collection.

Relevant Literature

Cognitive Validity

Work related to cognitive validation is in its relatively early stages[7] and is not typically considered when speaking of validity[8]. To date the concept of cognitive validity in education-related studies has been primarily used to analyze thought processes and higher-order thinking skills used in scientific and quantitative problem solving, but has also been considered in education research utilizing self-report items[8]. The process of cognitive validity and the resulting information obtained is described as "examining how respondents process thoughts, feelings, beliefs, or experiences as they respond to survey item"[8] (p. 140). Cognitive validation methods attempt to obtain evidence that the cognitive processes occurring when a participant takes the survey are consistent with what the survey designers had intended.

The cognitive validation of a web-based survey instrument can be related to a usability evaluation of a computer interface (as in both cases participants are completing a task on a computer), thus we can look to the usability literature in the psychology domain for guidance in sample size selection. Early research indicates that four or five participants will detect 80% of the usability problems in a user interface, with the most severe problems being detected by the first three participants, and each additional participant will be less likely to reveal new information[9]. However, Lewis determined that Virzi's conclusions regarding diminishing returns as additional participants are added actually depends on the likelihood of detecting problems[10] (so if problems are harder or less likely to be detected, more than five participants will be needed to attain the 80% problem detection). More recently, Faulkner demonstrated that in a repeated usability evaluation using groups of 5, as low as 35% of the problems were detected and to achieve a minimum detection of 80%, 10 participants should be selected[11].

Verbal Report Methods and Think-Aloud Processes

According to Stone and colleagues, "If you want to find out why participants are making their various choices while interacting with the interface, you will need to ask them about it"[12] (p. 478). Think-aloud processes, where participants are asked to vocalize internally-generated thoughts as they complete a given task[13], fall under a larger umbrella of cognitive techniques used to gain insight into human thought processes (e.g. verbal probing, interviewing, and focus groups), often referred to as Verbal Report Methods[5]. Think-alouds have been described as "the closest possible way to get to the cognitive processes of [participants]"[14] (p. 1). Thus, think-aloud techniques are a useful way to find out why these choices are being made, as the participant can take the survey with the researchers present, and the researchers can observe what is happening and ask questions as the participant interacts with the survey.

However, it is important to keep in mind that think-alouds are not the *only* way to gain insights into participants' thought processes and establish cognitive validity. One study in the health domain (measuring the effects of adult caregivers' social capital on the different aspects of well-being of the children for which they provide care) performed a cognitive validation of their instrument using cognitive interviews as the verbal report method[15]. Thus, it depends on the nature of the instrument, its application, and the goals of a particular cognitive validation process as to exactly what type of verbal report method is most appropriate.

There has been a recently growing trend to use qualitative techniques to improve the validity of instruments that are used to support primarily quantitative research goals[8] and there are some studies in the current literature that have utilized think-alouds to achieve this improved validity during instrument development[5,8,16,17]. While there is no singular protocol for conducting think-aloud sessions, current documentations of think-aloud studies and variations on protocols do provide some important points to consider, shown on the following page in Table 1. These points can keep the *design* of think-aloud sessions in check with their *purpose*: to ensure that participants are interpreting and responding to survey items in a way that is consistent with the intentions of the survey designers.

Table 1: Important Points to Consider when Conducting Think-Alouds[5,12,18]

- Tell the participant to voice any confusion or trouble they have when taking the survey.
- Make sure the participant is aware that the purpose of the study is to evaluate the *survey*, not the participant's performance. Treat the participant as more of a 'partner' in the study.
- If the participant seems to be struggling with a particular question, probe with a question to help fully understand the thought processes going through the participant's mind
    - What do you think this question is asking you?
    - How do you think you should answer this question?
    - Is this question confusing? *(Avoid asking if the participant is confused.)* If so, what would make this question less confusing?
    - What are you thinking about?
    - How did you arrive at that answer?
    - What does (a particular word/concept) mean to you?
- Researchers should consider the following questions if it is unclear whether or not a participant is having an issue:
    - Does the participant have an accurate internal representation of each question?
    - Does the participant have to re-read questions?
    - Does the participant seem to be giving a complete answer?
- Utilize *respective protocol*, where at the end of the survey (or at logical midpoints during a long survey) the participant will be asked to reflect upon the questions encountered and responses provided to determine if, after looking back, anything else seems confusing or if there is any additional information the participant thinks we should know but the instrument has not sufficiently drawn out of the participant.

Verbal Cognitive Validation in the Context of an Engineering Education Study

Purpose of the Research Study

This paper presents the utility of think-alouds as a VRM in the context of a particular engineering education study. The study is described here in order to lend context to the methods and results from the think-aloud protocol. It is the authors' intentions that using this specific research study as an example will aid other engineering education researchers in extending the potential efficacy of this process in other specific research contexts.

This work is part of a NSF-funded CAREER project (Grant Number EEC-0950710), which aims to significantly advance fundamental knowledge of social interactions that influence under-represented students' decisions to enter and persist in engineering. To accomplish this, we extend an established theoretical framework—social capital—to the field of engineering education. Simply put, social capital is "resources gained from relationships"[19]. Van Der Gaag described social capital as "an additional pool of resources embedded in the social networks of individuals, which can help to achieve individual goals in conjunction with or instead of personal resources."[20] In the context of engineering students' academic and career decisions, prior results suggest that students' decisions to select engineering as a college major and to persist in undergraduate engineering studies are influenced by the available resources in their social networks, as well as the activation of those resources[21-24]. Furthermore, under-represented students may utilize different mechanisms for developing and activating social capital. The project is a mixed-methods, multi-institution study. The first phase of the project involves developing and administering a specific kind of sociological survey instrument termed a "Name and Resource Generator" to approximately 1,500 engineering undergraduates at seven institutions in the U.S., from which descriptive statistics will be computed and group-level patterns in the survey data will be identified. Following the quantitative phase is a qualitative phase, where approximately 90 interviews will be conducted using grounded theory methodology.

The research study described herein aims to address the following research questions:

- How do social network indicators vary with gender, ethnicity and generational status in college for engineering undergraduates?
- What forms of social capital do students employ in making decisions about selecting and persisting in engineering at the undergraduate level, and how do these vary at the group level?

A think-aloud protocol was used in this study between survey instrument development and deployment. This step was included in the research process to give us confidence in the cognitive validity of the instrument, which was particularly useful because 1) this type of instrument is being newly developed and applied in the field of engineering education, 2) the large-scale nature of the survey makes it especially important to ensure as many issues as possible are identified and remedied before the survey is launched, and 3) the results obtained in the current quantitative phase of the project will help inform the following qualitative phase (interviews).

Description of Survey Instrument
It should be noted that in the present study, the nature of the Name and Resource Generator instrument does not require some of the statistical validation tests that are very common in research using surveys in educational research. The survey is not attitudinal in nature, and it does not measure constructs; rather, we are primarily measuring the number of ties socially relevant to a participant's education in engineering, the nature and strength of these ties, and the types of resources gained from these social ties and other types of people. While there are many situations where statistical validation assessments would be necessary in conjunction with cognitive validation exercises, this subject is outside the scope of the current research goals and protocol and is not discussed in further detail in this paper.

This quantitative phase of the study investigates the structure of participants' social networks and their available, accessed and activated social capital using a "critical incident" approach[25,26]. Two critical time periods, or decision nodes, in the academic and career decisions of participants are being explored: (1) when the participant made the initial choice to major in engineering, and (2) at the time of participation in the research study. This reflective approach also provides the ability to account for potential changes in social capital over time. The survey was adapted from two techniques commonly used by social scientists for social network mapping and social capital measurement to the specific context of engineering students' academic and career decisions[20]: the Name Generator and Resource Generator techniques. Data from the combined social network survey, termed the "Name and Resource Generator," or NRG, will be used to identify group-level patterns in participants' social networks and characterize those networks using descriptive statistics. This will be accomplished through quantifying various characteristics of individual participants' social networks (social network indicators) as well as their access and activation of embedded resources (social capital characteristics).

While the purpose of our particular instrument was to identify and quantify engineering undergraduates' access to and activation of social capital in making decisions about entering and persisting in engineering, our protocol and resulting changes made from think-aloud results in survey development process can serve as a model for other engineering education researchers.

Methodology

Cognitive Validation Model

The model presented in Figure 1 is based on a variety of literature related to think-aloud protocol and cognitive validation in usability engineering, engineering education, and educational psychology research. This summarizes our process in using cognitive validity evaluation, and, we believe can be adopted or adapted by other engineering education researchers in their own work. In our study, we found the iterative approach, recommended when pre-testing questionnaires[5,17,27], to be particularly helpful in ensuring as many issues as possible—including potential issues—are identified and that the remedies are indeed effective at fixing the issues.

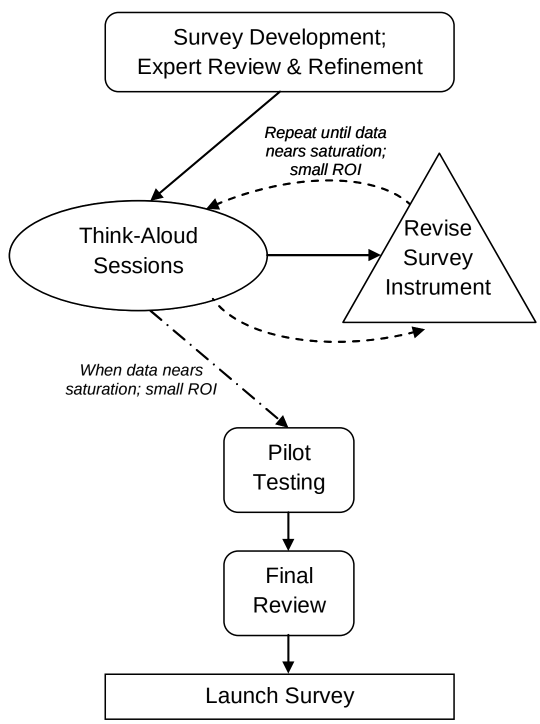

Figure 1: Survey Instrument Cognitive Validation Model

After initial screening, the instrument was tested in the first round of think-aloud sessions. The number of participants at this phase depends on details such as the nature of the survey, the complexity of the questions and information it is asking participants to provide, and the level of detail with which the internal review was conducted. In the present study, we followed Virzi's[9] recommendation to use four or five participants (which, as previously mentioned, could uncover up to 80% of the potential issues participants would face when interacting with the survey instrument), at which point the feedback received also became repetitive.

Then, the survey instrument was revised based on the feedback, observations, and suggestions that arose during the think-aloud sessions. After the initial revisions have been made, additional think-aloud sessions were conducted (with at least two or three new participants) to ensure that 1) the revisions have successfully remedied issues faced by previous think-aloud participants, and 2) no other serious issues were being detected within the instrument. If it was determined that criterion 1 and/or 2 are not satisfied, then another iteration of revisions and think-aloud sessions was conducted to ensure that as many potential issues with the instrument are identified and remedied. This iterative cycle of think-aloud testing and instrument refinement was repeated until the think-aloud feedback neared saturation—when feedback became repetitive, no new major issues were being identified by participants, and when there was a small return of investment (ROI) ratio between issues identified and resources spent on conducting the think aloud sessions.

When think-aloud feedback indicated no new (and few, if any, minor) issues, the survey was pilot tested (again, targeting a diverse demographic base consistent with the target strata identified in the NSF CAREER grant) to verify that that no major issues came up as the participants completed the survey, and also to verify that the questions were providing us with useful data. Following the pilot testing, an expert in the field reviewed the survey and the initial pilot data to identify any additional improvements to the instrument—particularly regarding questions that may not be necessary or do not provide useful data. After this review, the refined survey was launched, with additional confidence that participants' responses to the survey items will give the richer and more accurate data with which important research questions can be answered.

Recruitment

Participants for this study were recruited via e-mail to sign up for one-on-one think-aloud sessions. Recruitment letters were sent via e-mail to various undergraduate engineering courses and engineering programs at Clemson University and University of Houston. These programs were targeted with the goal of obtaining a stratified sample, specifically, achieving diversity in sex, ethnicity, and generational status.

Think-Aloud Protocol

After receiving the recruitment e-mail, potential participants were instructed to contact the first author to schedule a think-aloud session, which would last up to 90 minutes. The survey itself is designed to be completed in approximately 30 to 45 minutes; however, since participants in the think-aloud sessions would be talking about the questions and their answers much more than normal, we chose a 90-minute window to allow ample discussion time. Participants met with one or more research team members at an office equipped with a desk, computer, internet access, and chairs for the participant and researchers, or for participants at University of Houston, sessions were conducted via Skype. Participants were greeted and informed about the purpose of the NSF CAREER project, particularly emphasizing that our objective was to evaluate the survey instrument through the think-aloud testing, not the participants. Participants were provided with a bottle of water and were then instructed to begin taking the survey and reading out loud everything they saw on the screen. As participants took the survey, the researchers observed, took note of any issues the participant encountered or when/if body language indicated survey fatigue, and probed the participant with reminders to keep speaking or questions to elicit more information if the participant appeared confused or was not sure how to answer a question. After completing the survey and any final questions, the participant was thanked for their participation and given their advertised incentive (a $20 gift card to Amazon.com).

Participants
In total, 10 participants were recruited across the three think-aloud iterations that took place during this study. As shown in Table 2, participants included a variety of ethnicities (e.g. Caucasian, African American, Hispanic) and generational statuses (e.g. first-generation college students and non-first generation college). Thus, the demographics of the sample obtained for the think-aloud sessions is consistent with the targeted diverse strata of the larger NSF CAREER project (under-represented student groups in engineering). In addition to our primary strata, we were also able to achieve diversity in other characteristics, such as family background (e.g. biological parents still married, parents separated and remarried) and income level (e.g. low, lower middle, middle classes). Having a diverse group of participants evaluate the survey instrument helps ensure that the survey is viewed from as many different perspectives as possible.

Table 2: Participant Demographics Summary

Participants:	*n*	*Average Age (Standard Deviation)*	*Gender Ratio: Male/Female*	*Average Year in School (based on Years Attended, Class Standing)*	*Generational Status Ratio: First Generation College/non-FGC*
	10	21 (1.4)	30%/70%	2.9, 3.1	50%/50%
Ethnicity:	*White or Caucasian*	*Black or African American*	*Hispanic or Latino/a*	*Asian or Asian American*	*Other*
	30%	50%	10%	10%	0%
Parents' Income Level:	*Low*	*Lower Middle*	*Middle*	*Upper Middle*	*High*
	30%	20%	30%	20%	0%

Given Faulker's[11] findings about the number of participants necessary for verbal protocol analysis and the diminishing return on investment we later observed after the first nine participants, we were confident that 10 participants was a sufficient number to uncover the most severe issues—and a bulk of potential issues that participants may encounter— during this cognitive validation process.

Data Analysis
When reporting information about the various problems participants face when interacting with a survey, some literature has categorized these problems as mechanical or structural in nature (e.g. grammatical errors/typos) or cognitive in nature (e.g. participant cannot recall any relevant

information), with the acknowledgement that some problems have elements of both categories[5]. Typically, cognitive problems are considered to be more severe because these can lead to a significant mismatch between what the researchers are looking for and what the participant interpreted the question to be asking for. However, some mechanical or structural errors could become a cognitive problem if the error is so severe that it causes a participant to not fully understand a question or the provided answer choices.

During each revision decision point, issues were tracked in two lists. One contained more serious cognitive and structural (or both) issues that definitely needed to be fixed. The other list represented less serious or potential issues to be kept on a watch list. In the think-aloud sessions, we specifically looked for both types of issues as the participants took the survey, and in some cases we probed participants to determine what changes, if any, would be beneficial.

Results

During the process of following the cognitive validation model, three full iterations of revisions and retesting were conducted before the survey instrument received its final think-aloud test: five think-aloud sessions, followed by refinements to the instrument; three think-aloud sessions, followed by refinements to the instrument; one think-aloud session, followed by refinements to the instrument; and one last think-aloud session to confirm that all of the issues had been remedied and no new issues were being detected.

In the initial think-aloud sessions, the feedback revealed many issues within the survey instrument—both minor and major, and structural and cognitive. The predominant issue that arose was participants forgetting the time points for which they were answering questions (as a reminder, the survey prompted participants to answer questions 1) at the time they first considered engineering as a college major versus 2) at the present time), which was largely due to vaguely worded questions and/or question prompts not being noticed. It was also noted that several participants entered in an incorrect format answers to a question asking about how long they had known their influential people: instead of typing the number of years they knew someone (e.g. ‘7’) most would type the number plus years (e.g. ‘7 years’). While this was not a serious issue and did not affect participants’ actual answers, it is something that could potentially have caused us difficulty in analyzing the data, thus the answer format was changed to select a number of years from a drop-down list. When the first five participants had completed think-aloud sessions, this feedback was becoming repetitive and fewer issues were being noticed so we stopped scheduling new participants, refined the survey, and then continued think-aloud testing to 1) ensure that the previous issues were fixed and 2) to check for other issues.

In the second round, we determined that most of the revisions to the instrument did correct the issues from which they originally stemmed, except for the fact that many participants still skipped over the yellow-highlighted prompts instructing them to think about a particular time point as they answered questions. However, some new issues did surface (e.g. some participants were not noticing the instructional sub-prompts—text below a question giving additional instruction, e.g. ‘mark all that apply’)—warranting another round of revisions (although most of these issues were less serious than in the first round).

During the third round, we were satisfied that the previous issues had been successfully remedied, but the participant also suggested removing a logo that appeared at the top of the page, which may be causing participants to skip over the yellow highlighted prompts. This change was made to the instrument, along with a couple minor structural corrections, resulting in the final revised survey instrument. The instrument underwent one additional think-aloud screening, at which point we were satisfied that most of the serious issues participants could face had been fixed.

A summary of the think-aloud sessions, including the number of structural and cognitive issues identified by each participant and an overview of the revisions made to the instrument at each iteration of the think-aloud/revision process, is presented in Table 3. Specific examples of issues identified and corresponding refinements to the survey instrument are shown on the following page in Table 4.

Table 3: Summary of think-aloud/revision iterations

<table>
<tr><th>Iteration #</th><th>Participant #</th><th>Number of Structural Errors Identified</th><th>Number of Cognitive Issues Identified</th></tr>
<tr><td rowspan="6">1</td><td>1</td><td>8</td><td>5</td></tr>
<tr><td>2</td><td>3</td><td>3</td></tr>
<tr><td>3</td><td>5</td><td>2</td></tr>
<tr><td>4</td><td>3</td><td>7</td></tr>
<tr><td>5</td><td>1</td><td>6</td></tr>
<tr><td colspan="3">Major issues fixed:
• 11 structural/typographical changes were made to question prompts and/or answer choices to improve comprehension; 5 questions were also moved
• Changed parent education question to offer four parent types: Mother, Father, Female Parent/Guardian, and Male Parent/Guardian
• Split up long lists of questions to improve flow and remind about time points
• Changed answer format for two questions (from text entry to drop-down list)</td></tr>
<tr><th>Iteration #</th><th>Participant #</th><th>Structural Errors Identified</th><th>Cognitive Issues Identified</th></tr>
<tr><td rowspan="4">2</td><td>6</td><td>3</td><td>3</td></tr>
<tr><td>7</td><td>1</td><td>2</td></tr>
<tr><td>8</td><td>4</td><td>1</td></tr>
<tr><td colspan="3">Major issues fixed:
• Increased font size on sub-prompts
• Changed parent types again: Mother, Father, and 2 Stepparents/Other Guardians
• Added reflective question at the end of each of the 3 major sections (instead of one at the end)
• Improved layout of Resource Generator question prompts
• 2 minor structural/typographical corrections</td></tr>
<tr><th>Iteration #</th><th>Participant #</th><th>Structural Errors Identified</th><th>Cognitive Issues Identified</th></tr>
<tr><td rowspan="3">3</td><td>9</td><td>4</td><td>3</td></tr>
<tr><td colspan="3">Major issues fixed:
• Remove NSF logo from all pages except the first
• Moved one question to improve flow
• 3 minor structural/typographical corrections</td></tr>
<tr><td>10</td><td>0</td><td>1</td></tr>
</table>

Table 4: Example issues and remedies during think-aloud/revision iterations

Iteration #	Participant #	Example issue identified by participant *[S=Structural, C=Cognitive, P=Potential]*	Prompt, before	Prompt, after
1	1	Unsure about difference b/w credit hours and years in school *[C]*	"What is your university class designation?"	"What is your university class designation, based on credit hours?"
	2	Potential for incorrect interpretations or assumptions *[PC]*	"[*name*] was knowledgeable about engineering…"	"[*name*] communicated knowledge about engineering…"
	3	Skipped highlighted prompt at top of page *[S]*	"As of today, please answer the following questions about your relationship with each of the people you listed."	Same *(effective remedy not yet identified)*
	4	Wording in answer choices inconsistent with wording in prompt, causing an interpretation different than intended *[S/C]*	"How influential do you believe this person was on your decision to major in engineering?" [answer choices ranged from 'not at all instrumental' to 'highly instrumental'	Replaced 'instrumental' with 'influential' in each of the answer choices
	5	Confused time point 3 times (began listing professors from college at the time point prior to majoring in engineering) *[C]*	*(Top of page)* "Before or during the time you were initially considering engineering as a college major, did you know anyone who…" *(13 questions later)* "talked to you about engineering career options?"	Same, but divided the long list of questions into pages with 4 or 5 questions each to keep the prompt at the top of the page within view
Iteration #	Participant #	Example issue identified by participant	Prompt, before	Prompt, after
2	6	Suggested modernizing answer choices pertaining to a question about communication *[PS]*	"What is the most frequently used form of communication with this person?" [answer choices: in person, phone, e-mail, text, social networks, handwritten letter, other]	Same, but replaced 'handwritten letters' with 'video conferencing, e.g. Skype'
	8	Suggested increasing font size of sub-prompts to increase visibility *[S]*	Sub-prompt text was one size smaller than the main prompt	Sub-prompt text is the same size as the main prompt
Iteration #	Participant #	Example issue identified by participant	Prompt, before	Prompt, after
3	9	Logo at top of page may cause participants to skip over yellow time point prompts *[PC]*	NSF logo separated the top of the page from the prompt	Logo was removed so the yellow prompt is at the top of the page

Following the think-aloud iterations, the survey was pilot tested with a group of 140 students, which verified that participants did not report any issues while completing the survey and that the resulting data overall provided good information. Then, an expert in the field reviewed the pilot data and the instrument, identifying 12 of 16 questions related to participants' communication patterns with the individuals they listed which did not provide much additional insight than we were able to gain with the remaining four questions. Thus, these questions were removed.

Implications for Engineering Education Research

This paper has outlined the process taken to establish cognitive validity during the development and refinement of a survey instrument. While the survey instrument is a new type of instrument designed for the engineering education community (and specifically measuring social capital), the present research has been able to glean much from relevant work in other fields and is thus being presented to the engineering education field as an application of the potential efficacy of utilizing a predominately qualitative methods to build confidence that a survey instrument is cognitively valid. This additional validity can result in identifying many revisions that will substantially improve the quality of the instrument and the data it produces.

It is important to remember that while the think-aloud sessions are the primary component of the cognitive validation model and these sessions can be very useful in determining the thought processes of actual participants as they interact with a survey instrument, this technique on its own is not sufficient for establishing a survey instrument's cognitive validity. There are other steps that need to be completed before and after the think aloud sessions to gain sufficient evidence that the instrument is cognitively validated, and it ultimately depends on the nature of the instrument being developed as to exactly what other steps are necessary. In the present example, before the think-aloud sessions began, a panel of experts in the field (the advisory board on the first author's NSF CAREER grant) took the survey in mock format and provided feedback regarding items that were confusing and corrected many grammatical issues before the first group of students looked at the survey. After the think-alouds were conducted, pilot testing and subsequent analysis helped verify that no major issues came up as the participants completed the survey and that the questions were providing us with useful data. Finally, the external expert review identified several questions that did not provide new information and could be eliminated to improve flow and reduce fatigue.

The cognitive validation model outlined in this paper was a useful guide to follow while developing a new type of survey instrument for use in the engineering education domain. Following this model resulted in the identification of several issues with the instrument covering a wide severity range. While some issues were fairly minor, these can distract participants while they are taking the survey and possibly weaken the perceived credibility of the instrument and the researchers who developed it. Other issues were more serious and would have resulted in the participant answering the question differently than intended. Being able to fix these issues—which could have led participants towards incorrect interpretations—*before* the survey is launched gives us confidence that the data obtained will be richer, more accurate, and more reliable. This will also improve the later qualitative phase of our mixed-methods project by informing it with information obtained through a more cognitively sound quantitative instrument. We feel that the elements of the cognitive validation model were relatively straightforward to implement and follow, and that other engineering education researchers would be able to realize similar benefits with little difficulty.

Acknowledgments

This paper is based on work funded by the National Science Foundation Early Career Development (CAREER) program (Grant No. EEC-0905710). We would also like to thank Dr. Kathy Zerda for her assistance coordinating the think-aloud sessions via Skype with the University of Houston.

Bibliography

1. Eris, O., Chen, H., Bailey, T., Engerman, K., Loshbaugh, H.G., Griffin, A., Lichtenstein, G., & Cole, A. (2005). Development of the Persistence in Engineering (PIE) Survey Instrument. *Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition.*
2. Salkind, N.J. (2006). Tests & Measurement for People Who (Think They) Hate Tests & Measurement. Thousand Oaks, CA: Sage Publications.
3. Van Duzer, E. & McMartin, F. (2000). Methods to improve the validity and sensitivity of a self/peer assessment instrument. *IEEE Transactions on Education*, *43*(2), 153-158.
4. Finelli, C.J., Szwalek, J.L., Carpenter, D.D., & Harding, T.S. (2005). A case study on research in engineering education: designing, testing, and administering the PACES-2 survey on academic integrity. *Proceedings of the 35th ASEE/IEEE Frontiers in Education Conference.*
5. Willis, G.B., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questions. *Applied Cognitive Psychology*, *5*, 251-267.
6. Payne, J.W. (1994). Thinking aloud: insights into information processing. *Psychological Science*, *5*(5), 241-248.
7. Thelk, A.D., & Hoole, E.R. (2006). What are you thinking? Postsecondary student think-alouds of scientific and quantitative reasoning items. *The Journal of General Education*, *55*, 17-39.
8. Karabenick, S.A., Woolley, M.E., Friedel, J.M., Ammon, B.V., Blazevski, J., Bonney, C.R., De Groot, E., Gilbert, M.C., Musu, L., Kempler, T.M., & Kelly, K.L. (2007). Cognitive processing of self-report items in educational research: do they think what we mean? *Educational Psychologist*, *42*(3), 139-151.
9. Virzi, R.A. (1992). Refining the test phase of usability evaluation: how many subjects is enough? *Human Factors*, *34*(4), 457-468.
10. Lewis, J.R. (1994). Sample sizes for usability studies: additional considerations. *Human Factors*, *36*(2), 368-378.
11. Faulkner, L. (2003). Beyond the five-user assumption: benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, *35*(3), 379-383.
12. Stone, D., Jarrett, C., Woodroffe, M., & Minocha, S. (2005). User Interface Design and Evaluation. San Francisco, CA: Morgan Kaufmann.
13. Ericsson, K.A. & Simon, H.A. (1993). Thinking-Aloud Processes. *Protocol Analysis: Verbal Reports as Data.* Cambridge, MA: MIT Press.
14. Katalin, E. (2000). “Please, keep talking”: the ‘think-aloud’ method in second language reading research. *Novelty*, *7*(3)
15. De Silva, M.J., Harpham, T., Tuan, T., Bartolini, R., Penny, M.E., & Huttly, S.R. (2006). Psychometric and cognitive validation of a social capital measurement tool in Peru and Vietnam. *Social Science & Medicine*, *62*(4), 941-953.
16. Beatty, P.C. & Willis, G.B. (2007). Research synthesis: the practice of cognitive interviewing. *Public Opinion Quarterly*, *71*(2), 287-311.
17. Bolton, R.N. (1993). Pretesting questionnaires: content analysis of respondents’ concurrent verbal protocols. *Marketing Science*, *12*(3), 280-303.
18. Anderson, C. (2004). How much interaction is too much? *Usability Interface*, *10*(3), 9-11.
19. Lin, N. (2001). *Social capital: A theory of social structure in action*: Cambridge University Press.
20. Van Der Gaag, M. P. J. (2005). Measurement of individual social capital, Behavioural and Social Sciences (Vol. PhD, pp. 266). Available from http://dissertations.ub.rug.nl/faculties/ppsw/2005/m.p.j.van.der.gaag/.

21. Trenor, J. M., Yu, S. L., Waight, C. L., Zerda, K. S., & Sha, T.-L. (2008). The relations of ethnicity to female engineering students' educational experiences and college and career plans in an ethnically diverse learning environment. *Journal of Engineering Education, 97*(4), 449-465.
22. Trenor, J. M., Yu, S. L., Waight, C. L., & Zerda, K. S. (2008). Influences for selecting engineering: Insights on access to Social Capital from two case studies. *Proceeding of the 38th Annual Frontiers in Education Conference, 2008*, Saratoga Springs, New York.
23. Fernandez, M. J., Trenor, J. M., Zerda, K. S., & Cortes, C. (2008). First generation college students in engineering: a qualitative investigation of barriers to academic plans. *Proceedings of the 38th Annual Frontiers in Education Conference*, Saratoga Springs, New York.
24. Trenor, J. M. (2009). A phenomenological inquiry of the major choice processes of an overlooked demographic: First generation college students in engineering. *Proceeding of the 2009 Research in Engineering Education Symposium.*
25. Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 33.
26. Woolsey, L. K. (1986). The Critical Incident Technique: An innovative qualitative method of research. *Canadian Journal of Counselling, 20*(4), 242-254.
27. Converse, J.M. & Presser, S. (1986). "Survey questions: handcrafting the standardized questionnaire." Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-063. Beverly Hills, CA: Sage Publications.