

Using Cognitive Interviewing to Test Youth Survey and Interview Items in Evaluation: A Case Example

Elisa LaPietra
Montclair State University

Jennifer Brown Urban
Montclair State University

Miriam R. Linver
Montclair State University

Background: Cognitive interviewing is a pretesting tool used by evaluators to increase item and response option validity. Cognitive interviewing techniques are used to assess the cognitive processes utilized by participants to respond to items. This approach is particularly appropriate for testing items with children and adolescents who have more limited cognitive capacities than adults, vary in their cognitive development, and have a unique perspective on their life experiences and context.

Purpose: This paper presents a case example of cognitive interviewing with youth as part of a national program evaluation, and aims to expand the use of cognitive interviewing as a pretesting tool for both quantitative and qualitative items in evaluation studies involving youth.

Setting: Youth participants were located in four regions of the United States: Northeast, Central, Southern, and Western. Interviewers were located at Montclair State University.

Intervention: Not applicable.

Research design: A cognitive interview measure was designed to include a subset of survey items, interview questions, and verbal probes, to evaluate if these items and questions would be understood as intended by both younger and older youth participants. An iterative design was used with cognitive interviewing testing rounds, analysis, and revisions.

Data Collection and Analysis: The cognitive interview was administered by phone to 10 male youth, five from the 10-13-year-old age range and five from the 15-17-year-old age range. Interviews were audio-recorded, transcribed, reviewed, and coded. Survey items and interview questions were revised based on feedback from the participants and consensus agreement among the evaluation team. Item revisions were included in further testing rounds with new participants.

Findings: As a result of using cognitive interviewing to pretest survey and interview items with youth, response errors were identified. Participants did not understand some of the items and response options as intended, indicating problems with validity. These findings support the use of cognitive interviewing for testing and modifying survey items adapted for use with youth, as well as qualitative interview items. Additionally, the perspective of the youth participants was valuable for informing decisions to modify items and helping the evaluators learn the participants' program culture and experiences. Based on the findings and limitations of the study, we give practice recommendations for future studies using cognitive interviewing with a youth sample.

Keywords: *cognitive interviewing; item validity; response error; verbal probes; pre-testing surveys; qualitative evaluation; interviewing children and adolescents; survey development*

Introduction

High-quality evaluation is critical for supporting evidence-based decision making (McDavid et al., 2013). Program evaluators may assess a program's effectiveness (McDavid et al., 2013), or what impact it has on its participants or the community (Tatian, 2016), and determine if the program needs to be modified, or if a new one is needed (Rossi et al., 2019). As evaluation often results in practical applications for programs (Powell, 2006), it is imperative that the data collected are valid and that accurate descriptions of a program's performance and effectiveness (Rossi et al., 2019) are what inform practice decisions (Powell, 2006). In order to improve data accuracy, it has become standard practice in evaluation to pretest surveys to ensure the items are understood by respondents as the evaluators intended (Newcomer & Triplett, 2015). Cognitive interviewing is a pretesting tool used by evaluators to increase item and response option validity (Ryan et al., 2012). Cognitive interviewing techniques are used to assess the cognitive processes utilized by participants to respond to items. This approach is particularly appropriate for testing items with children and adolescents who have more limited cognitive capacities than adults and vary in their cognitive development (Silva et al., 2019). In addition, children and adolescents have a unique perspective on their life experiences and context (Katz et al., 2017), making it essential to explore how they interpret items that have been developed or adapted for use with them in their context. This study presents a case example of cognitive interviewing with youth as part of a national program evaluation, and aims to expand the use of cognitive interviewing as a pretesting tool for both quantitative and qualitative items in evaluation studies involving youth. We further contribute practice recommendations gained from this cognitive interviewing study with children and adolescents.

Literature Review

High quality evaluation requires careful evaluation planning including the critical examination of measures and their

appropriateness for the target population (Urban & Trochim, 2009). Pilot testing is a standard evaluation practice for assessing survey items prior to actual use in an evaluation study. Cognitive interviewing is a pretesting technique used to determine if items are understood by the respondents as the evaluator intended and to explore sources of response error with the goal of improving validity (Buers et al., 2014; Karpen & Hagemeyer, 2016; Teal et al., 2015). Cognitive interviewing is used to understand how participants interpret survey items and response options and to obtain their recommendations for improving items, response options, and survey format (Beatty & Willis, 2007). While quantitative methods are used to establish the psychometric properties of measures, cognitive interviewing is used to understand the reasons for item non-response and other response errors such as problems with item content, survey format, recall errors, and response categories (Buers et al., 2014; Campanelli et al., 2015; Irwin et al., 2009). Participants' responses to verbal probes, or questions about the items, allow evaluators to determine if survey items function as intended (Beatty & Willis, 2007; Buers et al., 2014), and provide evidence for the confirmation, revision, deletion, or replacement of items (Pepper et al., 2018). Cognitive interviewing has been used in both large-scale evaluations (Irwin et al., 2009) and evaluation studies with limited resources (Ryan et al., 2012) and is an established evaluation tool in public health (Buers et al., 2014; Irwin et al., 2009; Sebastian et al., 2014), education (Hofmeyer et al., 2015; Pepper et al., 2018), and psychology (Castillo-Díaz & Padilla, 2013; McCoy, 2014). Cognitive interviewing has become a standard part of the survey development process (Karpen & Hagemeyer, 2017; Teal et al., 2014), and is particularly applicable when testing surveys with children and adolescents to obtain their unique feedback and perspective for improving items (Kenny et al., 2017).

Theoretical Model

Tourangeau's (1984) four-stage cognitive processing model provides a framework for the survey response process that supports the use of cognitive interviewing to identify sources of

response error. According to this theory, answering survey questions involves a complex cognitive process that includes four successive stages: comprehension, retrieval, judgment, and response (Tourangeau, 1984; Tourangeau et al., 2000). Comprehension is determined by how a participant organizes information and makes use of prior information and context to determine its relevance. Retrieval or recall is the participant's process of remembering relevant information, or the use of memory. A participant's judgments are made based on how this information is evaluated and integrated. Finally, the participant selects and reports a response. Within this theoretical framework, cognitive interviewing can be used at any of these four stages to explore how a participant, for instance, comprehends the intent of an item, how relevant information is retrieved or recalled, how an answer is decided or judged, and the participant's process of response selection (Tourangeau et al., 2000). Willis (2005) notes that a problematic response could occur because of an error in any one of these stages in the cognitive process, such as a participant not understanding an item, forgetting relevant information, overestimating a time frame, or being unable to find an appropriate answer choice.

While the cognitive processing framework is applicable to adult respondents, it is especially useful for conceptualizing specific issues related to the response process for children and adolescents as they have different cognitive capacities than adults (Silva et al., 2019). According to Beatty and Willis (2007), age and cognitive functioning are the most important factors that contribute to a respondent's ability to accurately interpret and respond to an item. The meaning of items can vary as a function of age and across developmental levels, with younger respondents having more limited cognitive processing and more difficulty with interpretation, vocabulary, reading, and abstract concepts (Koskey et al., 2010; Woolley et al., 2004). The extent to which survey questions can be read, comprehended, and answered by respondents of a certain age has been described in the literature as developmental validity (Silva et al., 2019; Woolley et al., 2004) or cognitive validity

(Karabenick et al., 2007). Tourangeau's (1984) model provides a framework for the cognitive processes used in the survey response process, including the developmental cognitive processing considerations specific to children and adolescents.

Cognitive Interviewing Techniques

To explore the cognitive processes used in answering survey items, evaluators use specific cognitive interviewing techniques. The techniques can be categorized into two approaches: think-aloud and verbal probing (Willis, 2005). For the think-aloud approach, the interviewer reads the items aloud and asks the participant to comment on the survey as they are completing it (Beatty & Willis, 2007). If the participant pauses, the interviewer may ask, "What are you thinking?" (Willis, 1999), or may direct the participant to "keep talking" or "keep thinking aloud" (Pepper et al., 2018). Some evaluators prefer this open-ended think-aloud technique as it minimally guides the participant (Willis, 2005).

Verbal probing is also used to determine if items and response categories are being understood as intended. The interviewer asks follow-up questions after each item to better understand the participant's answers (Beatty & Willis 2007). Verbal probes can be scripted as part of the interview, used spontaneously in an unscripted manner (Willis, 2005), or used retrospectively after a survey is completed (Castillo-Diaz & Padilla, 2013). The type of probe selected depends on both the information needed and how the participant responds. For instance, Willis (2005) describes six categories of probes where the participant is asked: (1) comprehension/ interpretation questions about the specific meaning of a word, such as "What does 'race' mean to you?;" (2) paraphrasing questions to restate an item in their own words; (3) confidence judgment questions to assess how sure they are about their response; and (4) recall questions to determine how a participant knows how frequently or when an event occurred, such as "And how did you remember that you attended two times per week?" Other probes are 5) general, such as "What were you thinking about when you answered this question?" or 6) specific, such as "Why would someone

worry about getting the flu?” While evaluators have used a combination of both think-aloud and verbal probing techniques with adults (Domanska et al., 2018; Pepper et al., 2018; Vreeman et al., 2014), and youth (Koskey et al., 2010; Mulcahey et al., 2009; Silva et al., 2019), the emphasis is more frequently on verbal probing (Willis & Artino, 2013). Verbal probing is an especially useful technique with children and adolescents for exploring abstract concepts or context specific items that may be difficult for them to describe (Silva et al., 2019).

Sample Selection

Although the types of techniques used can vary, purposive sampling is typically used for selecting participants for cognitive interviewing studies. Participants are intentionally selected based on demographic or other characteristics specific to the study population (Willis, 1999), such as participants with medical conditions for health surveys (Buers et al., 2014; Irwin et al., 2009). A specific number of participants may be selected in order to increase variability, such as an equal distribution of men and women from various age ranges (Castillo-Diaz & Padilla, 2013). Selections can also be specific to racial or ethnic background (Reeve et al., 2011) or a particular life experience (McCoy, 2014; Norris et al., 2014). Due to the intensive nature of cognitive interviews (Teal et al., 2015), study samples can be effective with as few as 10 participants (Karpen & Hagemeyer, 2016; McCoy, 2014; Willis, 2005).

While cognitive interviewing is most commonly used with adults, they have also been completed with children as young as 8 years old (Irwin et al., 2009; Koskey et al., 2010; Mulcahey et al., 2009; Rebok et al., 2001), adolescents (Irwin & Stafford, 2016; Joffer et al., 2016; Koskey et al., 2010; Norris et al., 2014; Pepper et al., 2018; Silva et al., 2019), and young adults (Hinds et al., 2016; Karpen & Hagemeyer, 2016; Kramer & Schwartz, 2017). Although younger participants struggle with comprehension more than older children, they are able to understand items, response options, directions, and can identify problems with difficult language (Irwin et al., 2009; Rebok et

al., 2001). In order to include participants with varying cognitive capacity, evaluators typically select older and younger youth participants similar to their study population for the pretesting sample (Joffer et al., 2016; Kenny et al., 2017; Koskey et al., 2010; Kramer & Schwartz, 2017; Woolley et al., 2004).

Implementation

In addition to variability in the types of techniques used and samples selected for cognitive interviews, there is also variation in how cognitive interviews are implemented. For instance, adult cognitive interviews are often in-person, one-on-one meetings with an interviewer (Castillo-Diaz & Padilla, 2013; Reeve et al., 2011), but may also be conducted individually over the phone (Buers et al., 2014; Irwin et al., 2009). For adolescents, cognitive interviewing focus groups is an implementation approach that provides the advantage of peer interaction and support (Norris et al., 2014; Sebastian et al., 2014; Vreeman et al., 2014). Other studies with adolescents have used a one-on-one format to protect confidentiality and to create a more private environment in which adolescents may speak more openly (Silva, 2019). For children, creative approaches may be needed to engage them, such as the use of illustrations to represent concepts (Rebok et al., 2001).

Cognitive interviewing is typically implemented with an iterative approach: interviews are conducted in “rounds,” where the number of interviews completed in a round can vary, as well as the total number of rounds (Hinds et al., 2016; Irwin et al., 2009). Revisions are made to the survey after each round and retested with a new set of respondents, allowing for revisions to the survey during the cognitive interviewing process. There is currently no consensus as to how many rounds are considered adequate (Beatty & Willis, 2007; Hinds et al., 2016), but most cognitive interviewing studies include at least two rounds to test revisions from the first set of interviews (Willis, 2005).

Analyzing the Data

Analysis of data from cognitive interviews is also often done using an iterative process

together with a variety of data reduction methods, such as qualitative written comments and various coding schemes (Castillo-Diaz & Padilla, 2013; Karpen & Hagemeyer, 2016; Pepper et al., 2018; Willis, 2005). In the qualitative written comment approach, experienced members of an evaluation team independently review and analyze interview transcripts. Each reviewer writes individual notes, including observations and comments regarding participants' responses to each item. Reviewers also record summary statements for each item including themes and problems. The reviewers then meet and compile their statements into a larger summary format, yielding one set of combined observations and identified problems for each item (Buers et al., 2014; Irwin et al., 2009; Teal et al., 2015; Willis, 2005). The summary data are analyzed using content analysis or other thematic coding systems (Teal et al., 2015; Willis, 2005). While not essential for identifying problems with a survey, coding schemes provide a structure for the categorization of identified errors or problems (Buers et al., 2014) and help organize qualitative written comments (Willis, 2005).

Survey Revision

After reviewing the results from data analysis, decisions for survey revisions are typically based on reviewer consensus, observation summaries, and meetings with the larger evaluation team (Willis, 1999). Additional feedback from an Advisory Board or other stakeholder groups may be included in decision-making about items (Teal et al., 2005). Willis (1999) noted the importance of considering the implications of cognitive interview findings, such as if the observation or interpretation of the item is specific to the individual or something that other participants may also erroneously interpret. Generally, the evaluators establish the criteria used to determine whether there is a problem with an item that requires modification (Boeiji & Willis, 2013).

Participants often identify a variety of problems with survey items that require revision. Examples of reported problems include multiple interpretations of questions,

response categories that were too specific, recall difficulty for specific time frames (Buers et al., 2014), and problems with question wording, ordering, and item format (Teal et al., 2015). The use of a technical term in an item, for example, can lead to comprehension problems (Buers et al., 2014). Additional problems can include redundant items, item vagueness, response category problems, reference period problems, and problems understanding the intent of a question (Reeve et al., 2011).

In addition to the above described problems, children and adolescents have more challenges comprehending items and response options than adults, especially if they include abstract concepts (Woolley et al., 2004), or do not relate to the youth's personal experience (Roberts, 2017). For example, Rebok et al., (2001) described 5-year-old children not understanding "healthy" as it was intended by the researchers, and 4- and 5-year-old children having difficulty with the concept of "neighborhood" (Katz et al., 2017). Similarly, Woolley, Bowen, and Bowen (2004) found that youth 7 to 11 years old struggled to understand the response option "a little like me" and the item "I am happy with myself," as they were too abstract for this age group. Even some middle school age students struggled with the concept of "improvement" (Karabenick et al., 2007). In a study including older and younger adolescents, the older, 17- to 18-year-old respondents, provided more complex interpretations and detailed reasoning related to their health than the younger, 12- to 13-year-old respondents (Joffer et al., 2016). Additionally, in a sample of 11- to 15-year-old youth, respondents had problems with item context, such as whether a year meant a calendar year or a school year, and with generationally-bound jargon, such as "peer pressure" and "hot temper." Some of these youth also struggled to understand the middle categories on response scales, including "neutral" and "moderately" (Silva et al., 2019). These findings related to comprehension, abstract thinking, and context require item and response selection revisions in order to improve their interpretation.

Item revisions may be as simple as changing font sizes (Irwin et al., 2009) or dividing a survey into sections to make it

easier to read (Teal et al., 2015). To help with item comprehension, definitions for particular words may be added directly to the survey text (Teal et al., 2015) or photos can be added for items that were not recognized by name (Hinds et al., 2016). Additional survey revisions could include changing or adding to response options, adding or eliminating items, adding probes, and rewording items (Buers et al., 2014). For youth, item and response choice interpretation can also be improved by incorporating their personal perspective and the language they use to describe concepts (Mulcahey, 2009). Simplifying terms, making response scales clearer, and avoiding jargon is particularly important with youth due to a more limited cognitive capacity compared to adults (de Leeuw & Borgers, 2004). For all participants, items that continue to be problematic after several rounds of testing and that lack consensus regarding revisions may be flawed (Beatty & Willis, 2007). These items can either be eliminated, or kept in the survey for field testing and further analysis (Irwin et al., 2009).

Item Selection

Testing and revising items to reduce sources of response error prior to field testing has made cognitive interviewing an important part of the survey development process (Irwin & Stafford, 2016; Teal et al., 2015). Evaluators may test all survey items or select a subset from the larger survey when deciding which items to use in cognitive interviewing. Subsets can be randomly selected items (Karpen & Hagemeyer, 2017), items that are potentially problematic as determined by expert review (Rothgeb et al., 2007), or an established subset of the survey, such as an item set domain (Irwin et al., 2009) or a measurement scale (Hinds et al., 2016). The focus of cognitive interviewing is generally on testing survey items and response scales (Pepper et al., 2018; Kramer & Schwartz, 2017), or the understanding of content or constructs (Joffer et al., 2016; Hofmeyer et al., 2015), and is not typically done with interview items.

When deciding which items to test with youth, evaluators often include new items developed for the survey (Silva et al., 2019), modified items adapted for use with youth

(Rebok et al., 2001), and items established with other youth age groups (Koskey et al., 2010) or in contexts different from the study population (Bos et al., 2019). Even with established scales that have been used with other youth samples, the wording may not apply to the context of the study participants or may be outdated (Silva et al., 2019). In other words, established scale items may also need to be tested to determine if they will be interpreted as intended. After item selection and collaborative review with the evaluation team, cognitive interviews are typically completed prior to field testing a survey to improve the validity of the items (Sebastian et al., 2014; Teal et al., 2015).

BSA BEST Study Background

In the present study, cognitive interviewing was used as part of the survey and interview development process for the Boy Scouts of America Building Evidence in Scouting Together (BSA BEST) Study, a collaboration among the Boy Scouts of America (BSA), Montclair State University (MSU), and American Institutes for Research (AIR). The BSA BEST Study examined the BSA system to identify the existing trainings and other adult experiences that lead to the strongest youth character outcomes in Scouts. The primary goal of the initial phase of the BSA BEST Study was to develop and verify the theory of change for Scouts BSA (Urban et al., 2019), the BSA program for youth ages 11-17 years old.

The youth survey was developed through an extensive literature search for age-appropriate measures for each construct on the pathway model (theory of change). A total of 182 items from 32 existing scales were selected for the youth survey, in addition to demographic questions and items related to Scouting experiences. The evaluation team also developed a youth interview protocol. Constructs from the pathway model were selected to be included in the youth interview if: (1) researchers determined the constructs would be best addressed through interviews, and/or (2) constructs would contribute to triangulating interview questions and survey items to address youth character outcomes in Scouting (Urban et al., 2019).

The evaluation team was concerned that some of the selected measures designed for adolescents, young adults, or participants in non-Scouting settings, such as schools, could result in response errors when adapted for participants as young as 10 years old and for youth involved in Scouting. The youth cognitive interview protocol was developed to test a subset of survey and interview items adapted for the BSA BEST Study. Specifically, cognitive interviewing was used to determine if 29 survey items and their response options, as well as two interview questions, could: 1) be understood as intended by both younger and older Scouts from different regions of the country, 2) be modified according to the youth perspective in order to reduce response error, and 3) ultimately improve the validity of the survey items and interview questions.

Method

Sample

For this cognitive interviewing study, a convenience sample of 10 Scouts was recruited from four regions of the country as defined by BSA; two each from the Northeast, Central, and Western regions, and four from the Southern region as this region has the largest number of troops in the country. The Scouts were from independent troops in their region, except in the West where both Scouts were from the same troop. We purposively selected Scouts ranging in age from 10 to 13 years old and 15 to 17 years old to test how the items were understood by both younger and older Scouts. One younger and one older Scout were selected from the Northeast, Central, and Western regions. Two younger and two older Scouts were selected from the Southern region, resulting in five younger and five older Scouts in the sample. Seven of the Scouts identified as White/Caucasian and three identified as Asian/Pacific Islander. All participants identified as male.

In order to recruit Scouts for cognitive interviews, we first contacted BSA Councils and Scoutmasters and asked them to refer us to parents of Scouts. An introductory email was sent to parents, including a parental consent form and a youth assent form. When

completed consent/assent forms were received by the researchers, parents were contacted via email with a link to an appointment booking site specific to the study. After the parent scheduled an interview through the booking site, the cognitive interview items were sent via email to the parent and the Scout. A \$15 e-gift card to the National Scout Shop was offered as an incentive for participation, and emailed to the participant after the completed interview.

Measures

To develop the cognitive interview measure, the evaluation team created a list of the survey items that met at least one of the following criteria: (a) modified by the evaluators; (b) not previously used with the full 10-17 age range; and (c) included wording or response options that were possibly confusing for Scouts as determined by the evaluators. We also included potentially confusing questions from the youth interview. A list of possible verbal probes was created for each item based on published guidelines (Irwin et al., 2009; Willis, 2005; see Table 1) and the anticipated problem with the item. Discussions with the larger evaluation team and feedback from BSA program professionals guided the selection of 29 survey items, their response options, and related probes; as well as two interview questions for the cognitive interviewing protocol. See Table 2 for examples of survey and interview items selected for the cognitive interview along with related verbal probes. In these examples, the probes for the survey item were used to evaluate the participant's general understanding of the question and response options, as well as the comprehension of specific words in the question. For the interview items, comprehension probes were used to explore the participants' understanding of the meaning of specific words. The cognitive interview also included instructions for the interviewer and an introductory script for the interviewer to read to the participants. The cognitive interview protocol and study procedures were IRB approved.

Table 1
Verbal Probe Examples from the Cognitive Interview (Classified by Willis [2005])

Type of Probe	Example
Paraphrasing	<i>In your own words, what does this statement mean?</i> <i>Using your own words, what do you think this series of questions is asking?</i>
General	<i>What did you think about when answering this question?</i> <i>How easy or hard is it to answer this? (Why?)</i> <i>What do you think about the answer choices?</i>
Specific	<i>Can you tell me more about why you answered _____ (i.e. strongly agree)?</i> <i>What are troop activities?</i> <i>Are there any other leadership roles available to you not listed here?</i>
Recall	<i>How did you determine how often you attend these meetings?</i> <i>What helps you remember when you held these positions?</i>
Comprehension/ Interpretation	<i>What do you think it means when people are similar to you in background?</i> <i>What is an example of a different culture?</i> <i>What do you think "Religious activities" includes?</i>
Spontaneous	<i>Can you tell me more about that?</i> <i>Can you give me an example?</i> <i>What do you mean by _____? (participant's word or phrase)</i>

Table 2
Cognitive Interview Survey and Interview Item Examples with Probes

Survey Item and Probes
<p>How often do you get to practice leadership skills at BSA meetings or events?</p> <p><input type="checkbox"/> never <input type="checkbox"/> at some meetings/events <input type="checkbox"/> at most meetings/events <input type="checkbox"/> at all meetings/events</p> <p>·<i>In your own words, what do you think this question is asking?</i></p> <p>·<i>What does it mean to you to "practice leadership skills?"</i></p> <p>·<i>Anything else? (ask for examples of leadership skills if not given)</i></p> <p>·<i>What did you think about that you answered...?</i></p> <p>·<i>Do the answer choices make sense to you?</i> <i>(If no) Can you tell me why they don't make sense?</i></p> <p>·<i>How could the answer choices be better?</i></p>
<p>-----</p> <p>Interview Item and Probes</p> <p>-----</p> <p>Values are important in Boy Scouts. What are your core values?</p> <p>·<i>In other words, what values do you live by?</i></p> <p>·<i>Can you give me some examples of core values (or values to live by)?</i></p> <p>·<i>Can you tell me more about (value they gave) and what that means to you?</i></p> <p>·<i>What do you think of the use of the word "core" in this question?</i></p>

Procedure

The cognitive interviews were administered by phone by two graduate-level experienced interviewers. Each interviewer received 10 hours of training specific to cognitive interviewing, including the purpose of cognitive interviewing and the use of verbal probes to gather information about items, response options, and survey format. As part of the training, the interviewers studied the items and probes from the protocol, the underlying constructs, the protocol script, and the cognitive interviewing procedures developed for the study. Interviewers practiced audio recording directly from the phone line using a connected recorder and headset. Finally, each interviewer participated in two audio recorded practice cognitive interviews with mock participants and received supervisory feedback.

The youth cognitive interviews were completed with 10 Scouts during June and July of 2018. At the beginning of the call, participants gave verbal assent. Interviewers then confirmed the participants had a copy of the interview items and response options in front of them. Next, the interviewer briefly explained the process by stating, "I'm going to be asking some questions about you and about Scouting. After you answer a question, I'm going to ask you to give me your opinion about it and what you think it means. There are no right or wrong answers. I just want to know what you think." Participants were instructed to follow along as the interviewer read the items and to state their answers out loud. Scripted verbal probes were asked following each of the participant responses. Spontaneous verbal probes were also used as needed for further clarification. The duration of interviews ranged from 26 to 46 minutes, with an average of 35 minutes. Upon completion of the interview, the audio file was uploaded to a secure shared drive and professionally transcribed by a secure contracted service. The transcripts were checked by trained research assistants who were part of the evaluation team.

Data Analysis

The transcripts and audio recordings were independently reviewed by two experienced members of the evaluation team. Using the qualitative written comment approach (Willis 2005), reviewers independently recorded observations after each item and its corresponding response for each cognitive interview. The reviewers then met to discuss, reach consensus, and code their observations. These reviewer meetings occurred after the first, fourth, and tenth interviews.

Observations were coded as easy, moderate, or difficult. "Easy" codes indicated that a participant suggested a direct change, such as adding a response option of "None." Also included in this category were changes interpreted by reviewers as simple or straightforward, such as adding a probe for further clarification or changing a response set from a horizontal wrapping format to a vertical list to make it easier to read.

Observations coded as "moderate" indicated that a participant was unsure of the question wording or response options. Examples of observations coded as moderate were items that were understood by the participant in a more limited way than intended, items where participants had difficulty locating a response option that was there, and items where participants did not understand the response options as intended. Some of these observations were discussed with the larger evaluation team for additional feedback.

Observations coded as "difficult" indicated that the participant struggled significantly to answer a question. Examples included participants not understanding item wording, such as "What are your core values?"; participants indicating multiple interpretations of a particular word, such as "background"; response options that were too complex; and items that the evaluation team struggled with clarifying or simplifying for participants, such as an extensive list of activity categories. All observations coded as difficult necessitated discussion with the larger evaluation team.

A summary document was created after each of the three reviewer meetings that included revisions made (easy), suggested

revisions to review with the evaluation team with potentially brief discussion (moderate), and observations that would likely require extensive discussion with the evaluation team prior to any revision (difficult). If the team reached consensus, revisions were made to the cognitive interviewing protocol. Using an iterative process, revised items were included and tested in subsequent interviews, until consensus was reached that the item was being interpreted as intended, needed further modification, should remain in the protocol for additional testing, or be removed.

Results

As a result of cognitive interviewing, we were able to identify sources of response error, and make changes to the survey items and response options to make them clearer. Participants made three types of response errors (Tourangeau et al., 2000): (1) comprehension (how the intent of an item is understood), (2) judgment (how information is used to decide an answer), and (3) response

selection (how a response is selected). Slightly more than half of reviewer observations coded as difficult were related to comprehension problems, such as not understanding the intent of an item or an item having multiple meanings. Additional observations were coded as difficult due to problems with judgment, such as an explanation not supporting an answer choice, or problems with response selection, such as a participant answering with an activity that did not fit into the response options. Almost all of the reviewer observations coded as easy were related to problems with response option selection, such as a missing response option. Half of the observations coded as moderate were also due to problems with response option selection, such as participants having a clear answer, but not being able to find the appropriate response option. The other half were related to problems with comprehension and judgement. See Table 3 for a summary of the observation codes by type of response error. There were no findings related to the Scouts' region of the country.

Table 3
Observation Codes and Types of Response Errors

Observation Code	Type of Response Error			Total
	Comprehension	Judgment	Response Selection	
Easy	0	1	10	11
Moderate	2	2	4	8
Difficult	9	3	4	16
Total	11	6	18	35

Response errors led to one of three actions: (1) straightforward protocol modifications, (2) more complex protocol revisions, or (3) removing items from the protocol. A straightforward change included adding a "None" category to the response options. More complex revisions involved item wording and required discussion with the evaluation team. Several items received multiple revisions after reviewer meetings and discussions with the evaluation team. These revised items were included in the remaining rounds of cognitive

interviewing to test if they were understood as intended. Items that continued to be problematic for participants after 10 interviews were either eliminated or revised and retained for further testing in the pilot study. See Appendix A for a summary of all reviewer observations, revisions made to the survey and interview items, observation codes, and categories of response errors as a result of cognitive interviewing.

Youth Perspective

In our study, the youngest participants had more difficulty with item comprehension compared to older participants. For example, younger participants had difficulty answering the interview question, "What are your core values?" One 10-year-old Scout stated that he did not know what values were, and an 11-year-old Scout stated, "Like what you're best at, what you like doing most, and how enthusiastic you are about it." The older adolescents, however, were able to verbalize specific values such as equality, friendship, treating others the way you want to be treated, honesty, being respectful, trustworthy, and kind. In order to improve comprehension of this item, the probe was added, "What values do you live by?" This probe directed Scouts to their personal experiences related to values, enabling the younger Scouts to identify them in subsequent testing rounds.

There was also variation among participant responses in defining certain terms, such as "background" and "cultural diversity." For example, from the survey item, "I prefer the company of people who are very similar to me in background and expressions," background was described by several Scouts as one or more of the following: culture, ethnicity, race, nationality, or religion. A younger Scout described background as how well a person socializes. An older Scout related background to economic status and urban versus rural living environments. Similarly, from the survey item, "Cultural diversity within a group makes the group more interesting," cultural diversity was described in various ways, such as different religions, languages, cultures, places, countries, and "from all sorts of walks of life," indicating a lack of agreement on its meaning.

Additional findings included participants recognizing response options as too limited and missing "None" or "Don't Know" categories, as well as response options that were too broad or presented in a complex format. One response scale had response selection errors related to judgment, where the participants' answer choices did not reflect their explanations. For instance, when asked about the number of hours spent on an activity, one participant verbalized "one hour

and 30 minutes," but then selected "30 minutes to less than an hour" from the response options. In this same response scale, other participants did not recognize the response option "more than 2 hours" to include their verbalized answers, "three hours or more" or "up to five hours." This response scale received several revisions (Appendix A) and was retested in subsequent rounds of cognitive interviewing.

Youth Context

An unexpected finding was the themes that emerged from participants' responses to the survey items and the interview questions related to the culture of Scouts BSA and their Scouting experiences. All ten Scouts participated in Cub Scouts from a young age, primarily beginning in 1st or 2nd grade. They were able to describe opportunities to practice leadership skills at Scouting meetings and activities, and all of them had been camping with their troop and attended BSA summer camp. Almost all Scouts strongly agreed with the item, "I feel that my Scoutmaster provides me with choices and opportunities for participating in Troop activities." One Scout, however, disagreed, stating that the choices and opportunities come from the Scouts themselves. Many of the Scouts used this item to talk more about their Scoutmaster. For example, an older Scout responded, "A hundred percent, I felt that definitely applied to my troop, that our Scoutmaster definitely supports us in that sense because he's always out there. He's always helping to plan the meetings and he's always getting the younger Scouts, the Scouts who aren't involved, to participate." Other Scouts described their Scoutmaster as inclusive, teaching, helping, guiding, or leading.

In addition, from the interview item, "Do you feel that your troop values your background, attributes, and experiences for example, your race, gender, culture, or orientation?" several Scouts responded that their troop respects each other's backgrounds and one noted that he feels they respect his religion and culture. Another Scout stated that the troop tries to learn about different cultures, stating "Well, we have sometimes intermittently we'll have international nights

where everybody brings in their own ethnic dishes, and they'll make an ethnic presentation based upon where they're from. So our troop kind of uses it as a learning opportunity to learn about different cultures." Similarly, another Scout noted, "So our troop accepts a lot of people. I don't know. I feel like a lot of our meetings are about different religions, and different ethnic groups. And also special needs. So, we make meetings for those, so we show I guess, our respect for everything." Also in response to this question, while several Scouts acknowledged feeling personally valued or respected by the troop, they shifted the focus to their contribution to the group as being more important. For instance, one older Scout noted the following:

I think my troop has a lot of respect for my values, for my background, attributes, and experiences, but I don't think they necessarily value it as something significant individually. I think it's more of the troop sees my experiences as a contribution to a more general effort, if that makes sense.

Finally, several Scouts felt that being valued by the troop was directly related to being included, everyone being able to participate, and having a role in the troop, more so than individual differences. The themes that emerged from cognitive interviewing with these youth participants as a result of verbal probing both survey and interview items, gave the evaluation team a deeper understanding of Scouting and their experiences. This context knowledge was important as it informed our decisions for modifying items, response selections, and interview questions in order to align them with the Scouts' perspective.

Discussion

As a result of using cognitive interviewing to pretest survey and interview items with youth in Scouts BSA, we identified problems related to comprehension, judgment, and response selection. These response errors were revealed by younger Scouts when they were asked to define abstract concepts, by older and younger Scouts when they were asked to define ambiguous terms, and when participants were

selecting responses or were asked for their opinions about response options. The various challenges the Scouts had with the items and interview questions reflected problems with validity, or the participants not understanding the items as we intended them to. These findings support the use of cognitive interviewing for testing items that have been modified or adapted for youth or used with other youth age groups or contexts. Additionally, we learned about the culture of BSA based on the perspective of ten Scouts. Based on the findings and limitations of this study, we give practice recommendations for future studies and for conducting cognitive interviews with a youth sample.

Not surprisingly, the youngest participants had difficulty with item comprehension compared to older participants. This variability in comprehension is supported in the literature by differences in cognitive development between children and adolescents (Ginsburg & Opper, 1988). Most older adolescents are able to comprehend abstract ideas (Elkind, 2001) and, in our study, understood concepts such as "values." Children approximately 8 to 11 years old, however, think more literally and concretely, necessitating questions related to their direct experience (Borgers et al., 2000; Joffer et al., 2016; Scott, 2008). Younger Scouts in our study were not able to articulate their values. As a result, the interview question, "What are your core values?" was reframed in terms of their personal experience by adding the verbal probe, "What values do you live by?" This probe facilitated responses more aligned with our intended meaning of "values" and improved this interview question for use with younger Scouts.

The cognitive interviewing process revealed additional comprehension problems when participants gave multiple interpretations for the survey items that included the terms "background" and "cultural diversity." The survey items containing these terms were selected from a questionnaire that had been validated with a college student population (Moely et al., 2002). While these items were appropriate for college students, they proved difficult to interpret for our youth sample. Although applicable to all study populations, these findings exemplify the importance of making survey items simple, direct, and

unambiguous for children and adolescents (Borgers et al., 2000; Silva et al., 2019), due to their developing cognitive capacity (Woolley et al., 2004).

As it can be difficult to measure multidimensional concepts with children and adolescents (Kenny et al., 2017) in a simple and direct way, open-ended verbal probe questions were used to gain a better understanding of how the Scouts interpreted the abstract terms, “background” and “cultural diversity,” within their context or experience. Open ended questions are valuable for gaining in depth information on a topic and for obtaining information that evaluators may not have considered asking about (McDavid et al., 2013). With cognitive interviewing, we were able to identify problems with the interpretation of these terms, understand why they were problematic, and work toward rephrasing them with the aim of improving item validity.

Cognitive interviewing also helped us uncover errors related to response selection and judgment. For several survey items, we did not realize that a “None” category was an appropriate response option. The Scouts’ feedback was valuable for improving these response options, as without adding a “None” category, these Scouts would not have had an accurate answer choice to reflect their experience (Mulcahey et al. 2009; Silva et al., 2019). In addition, for one of the response scales, Scouts had difficulty correctly selecting the response category (“more than 2 hours”) that represented their answer (e.g., “three hours or more”). This type of response scale is complex and relates to response errors involving judgment and response selection (Tourangeau, 2000). These response errors may not have been discovered through pilot testing as the Scouts selected alternative responses from their verbalized answers. When participants are unable to match their answer to the response categories, it decreases data quality (DeCastellarnau, 2018). The use of verbal probes in the cognitive interview helped to reveal these errors, and the response options were revised and retested.

All of the items that we selected for testing were from measures that had not previously been used with the full 10-17-year-old age range or were adapted for use with youth. Our findings of comprehension, judgment, and

response selection problems indicate that it was important to test these items with younger and older Scouts and not assume that they would be understood as intended (Newcomer & Triplet, 2015). Similar to this study, adapting measures for use with a new population or in a different setting, such as Scouting, is a common reason certain items are selected for pretesting (Rebok et al., 2001; Silva et al., 2019; Woolley et al., 2004).

While the purpose of the cognitive interviews was to test the survey and interview items and ensure they were being understood by Scouts as intended by the evaluation team, we also learned about the Scouts BSA culture and the life experiences of these Scouts through the cognitive interviewing process. All of the youth spoke positively about their camping experiences with their troop and attending BSA summer camp. The older Scouts emphasized the troop being led by the Scouts and gave examples of Scouts planning service projects and leading the younger Scouts. The Scoutmaster, or the primary adult leader of the troop, was described as including, planning, teaching, helping, guiding, and allowing youth to pick outings and lead. All of these Scouts had been participating in BSA programs from a young age, beginning with Cub Scouts in their early elementary school years. While the participants were aware of differences in troop membership by race, culture, or religion, they emphasized everyone participating and working together to achieve troop goals. When asked about the value of individual differences, many of the older Scouts felt that what the person contributed to the troop was more important than individual differences. While there was a recognition of different experiences, beliefs, and cultures, the broader troop identity was emphasized. Similarly, Scouts felt valued by their troop because they had a role, were included, or felt involved. Learning about the experiences of your participants and the scope of your topic from their perspective is common with cognitive interviewing (Karabenick et al., 2007; Kenny et al., 2017). This information was valuable for revising items within the culture and context of these Scouts, and broadened our understanding of the target constructs as they applied to Scouting.

Additionally, there is an emphasis in the literature on research with youth increasing our understanding of their experiences (Karabenick et al., 2007). This process is also described as consulting with children and adolescents as experts on their own experience, language, and culture (Kenny et al., 2017), exploring topics from the youth perspective (Hall et al., 2016; Joffer et al., 2016; Katz et al., 2017), and consulting with youth as partners in research (Kenny et al., 2017). Furthermore, youth participatory evaluation aims to include youth as partners in all steps of an evaluation (Sabo, 2003). Consulting with youth using cognitive interviewing is consistent with the principles of youth participatory evaluation including actively engaging with youth to gain an understanding of their perspective, treating youth as experts in their own experience, and empowering youth to be part of the evaluation process rather than simply research subjects (Hall et al., 2016; Roberts, 2017; Zeldin et al., 2012).

Limitations

There were several limitations to this study including the small sample, testing only a subset of items, limitations specific to the number of testing rounds in the iterative process, and the reliance on the evaluators' judgment for modification decisions. A challenge to using cognitive interviewing with small samples is evaluators have to decide if a problem is unique to a participant or would also be relevant to other participants (Beatty & Willis, 2007). While the use of small samples is common in cognitive interviewing due to the intensive nature of the interviews (Willis, 2005), the same findings with additional participants in a larger sample would offer stronger support for item modifications, or yield different potential modifications. There has been a strong positive relationship found between sample size and problem identification, with more problems found as the sample size increases (Blair & Conrad, 2011). Newcomer and Triplet (2015) recommend a pretesting sample size of 20 to 25 participants and Willis (2005) recommends 5 to 15 participants per round for sufficient testing of the items. Although we were able to

recruit both younger and older Scouts from all four regions of the country, a larger sample may have revealed additional problems with the items, elicited further recommendations for modifications, and enabled us to test revisions with more participants.

A second limitation of the study was that we tested only a subset of survey and interview items. Many evaluators use cognitive interviewing to test an entire survey (Buers et al., 2014; Irwin et al., 2009; Mulcahey et al., 2009). While doing so may have strengthened the study, testing all of the survey items was not feasible due to study time constraints and sample availability. Instead, we established criteria for testing items that were potentially problematic, such as items not previously tested with the full 10-17-year-old age range, and items adapted for youth or modified for a Scouting context, and potentially confusing interview questions. While the subset of items provided valuable findings, it would have been beneficial to know if there were response errors related to other survey and interview items prior to field testing.

For the items that were selected, the iterative process was important for retesting modified items. A limitation of the iterative process, however, is determining a sufficient number of testing rounds (Beatty & Willis, 2007). In the current study, if re-tested items did not result in the same error from a previous testing round, the evaluation team deemed the item to be improved and the change supported. Due to study time constraints, cognitive interviewing ended after three rounds of testing when the desired number of 10 cognitive interviews was reached. Several difficult items remained in the survey and interview protocols for further testing in the pilot phase of the evaluation. While it is unclear if these remaining items would have been resolved with further rounds of cognitive interviewing, the iterative process raised our awareness that these items continued to be problematic and needed further revision or elimination from the measures.

Finally, an important limitation not unique to the present study is the use of evaluator judgment to make decisions about the revision or elimination of items. If one or more participants has a problem with an item, it is the evaluators' decision as to whether or not

that item would present a problem for others and how it should be modified or eliminated (Beatty & Willis, 2007; Fowler et al., 2016; Presser et al., 2004). McDavid, Huse, and Hawthorn (2013) note, however, that no program evaluation can be done without professional judgment, or the evaluator's experiences, values, beliefs, and expectations having an important role. In our study, we strengthened the support for these decisions by using a consensus approach among our multi-disciplinary evaluation team, including the survey and interview designers, interviewers, interview raters, and subject matter experts. With consensus among the evaluation team after each testing round, items were modified, retained in the cognitive interview for further testing, or removed from the protocol. Hsu and Sandford (2007) note that feedback from an iterative approach allows for the reevaluation of previous judgments and helps to build consensus. Although the decisions ultimately relied on our informed judgments, the feedback received from testing the items with Scouts helped to build consensus among the evaluation team and strengthened support for the changes made.

Despite the limitations to the study, cognitive interviewing with Scouts was valuable to improving the survey and interview measures and for understanding Scouts' perspective on the items. Through cognitive interviewing, we included a sample of the participant population in the survey and interview development process. While testing a subset of survey and interview items, we essentially consulted with these Scouts regarding their opinions on the format of the survey, their ability to locate an answer choice, the meaning of terms and items, and solicited suggestions for improvement. These Scouts identified missing response choices and critiqued complex response options. This participant perspective assisted us with gaining a better understanding of the study population, shed light on why certain terms or items were problematic, and helped us to contextualize the results in the BSA Scouting experience. These observations reflect the expanding relevance of cognitive interviewing for examining how participants' life experiences and culture influence the response process, as well as how items are

understood (McCoy 2014; Norris et al., 2014; Vreeman et al., 2014; Willis & Miller, 2017).

Practice Recommendations for Future Studies

Based on our experiences in this study, we make the following recommendations for evaluators interested in conducting cognitive interviews with child and adolescent populations:

1. Test as many items from the measure as possible, not only the items evaluators believe may be problematic. Youth offer a unique perspective and will likely identify problems that adult evaluators do not recognize or anticipate;
2. Cognitive interviewing can be used with youth to pretest qualitative interview items as well as quantitative survey items. While cognitive interviewing has primarily been used as part of survey development, useful feedback was received from participants on interview items that resulted in revisions to the questions prior to pilot testing, and informative descriptions of constructs that are challenging to measure with youth;
3. Use an iterative process and retest item and response option revisions with an adequate sample size as determined by your evaluation team. A larger sample may offer stronger support for confirming problem items and allow for additional testing rounds to determine if revisions have improved the items;
4. Use the youths' challenges and feedback that result from cognitive interviewing to build consensus among the evaluation team on whether to retain, modify, or eliminate items. A consensus approach can help to minimize individual judgment;
5. Treat the cognitive interview as a consultation with your study population, as children and adolescents are experts in their own experience. Hearing their perspectives can help evaluators determine if survey items, response options, and interview questions have validity, or are interpreted as intended.

Conclusion

Cognitive interviewing was used to test and improve the validity of youth survey and interview items for the BSA BEST Study.

Although a small sample was recruited, purposive sampling ensured participation from younger and older Scouts from different regions of the country. Based on the results of cognitive interviewing, survey and interview items were revised prior to the actual evaluation. The use of a verbal probing technique and feedback from youth participants helped the evaluators not only understand that an item or response option was problematic, but also why it was problematic. Scouts' responses confirmed if items were being understood as intended, needed revision, further testing, or to be replaced or eliminated. While cognitive interviewing is typically used in survey development, this process was also useful for understanding how Scouts understood qualitative interview items and to learn about the life experiences and perspectives of the participating Scouts. Learning about the Scouting context helped us to further refine items. Similar to other evaluation studies, the data collected for the BSA BEST Study may be used to inform programmatic decisions. As a result, it is important that the data accurately reflect what evaluators intended to measure. Cognitive interviewing is a critical pretesting step in the process of measurement development that evaluators can use to improve item validity in the context of the youth perspective.

Funding

Funding for this research was provided by the S.D. Bechtel, Jr. Foundation under Grant 8325.

References

- Beatty, P.C., & Willis, G.B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), 287-311. <https://doi.org/10.1093/poq/nfm006>
- Blair, J. & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly*, 75(4), 636-658. <https://doi.org/10.1093/poq/nfr035>
- Boeiji, H. & Willis, G. (2013). The Cognitive Interviewing Reporting Framework (CIRF): Towards the harmonization of cognitive testing reports. *Methodology*, 9(3), 87-95. <https://doi.org/10.1027/1614-2241/a000075>
- Borgers, N., de Leeuw, E., & Hox, J., (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin of Sociological Methodology*, 66, 60-75. <http://pascal-francis.inist.fr/vibad/index.php?action=search&terms=1551123>
- Bos, N., Engel, M. F., van Rijswijk, N. J., Verheijden, J. M. A., Coster, W., Moed, R., & Ketelaar, M. (2019). Translation and cross-cultural adaptation of the PEDI-CAT: Dutch version. *Journal of Pediatric Rehabilitation Medicine: An Interdisciplinary Approach*, 12, 57-64. doi: 10.3233/PRM-180544
- Buers, C., Triemstra, M., Bloemendal, E., Zwijnenberg, N. C., Hendriks, M., & Delnoij, D. M. (2014). The value of cognitive interviewing for optimizing a patient experience survey. *International Journal of Social Research Methodology*, 17(4), 325-340. <https://doi.org/10.1080/13645579.2012.750830>
- Campanelli, P., Gray, M., Blake, M., & Hope, S. (2015). Cognitive interviewing as tool for enhancing the accuracy of the interpretation of quantitative findings. *Quality & Quantity*, 50(3), 1021-1040. <https://doi.org/10.1007/s11135-015-0188-y>
- Castillo-Díaz, M., & Padilla, J. (2013). How cognitive interviewing can provide validity evidence of the response processes to scale items. *Social Indicators Research*, 114(3), 963-975. <https://doi.org/10.1007/s11205-012-0184-8>
- DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity*, 52, 1523-1559. <https://doi.org/10.1007/s11135-017-0533-4>
- De Leeuw, E., Borgers, N., & Smits, A. (2004). Pretesting questionnaires for children and adolescents. In Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Singer, E. (Eds). *Methods for testing and evaluating survey*

- questionnaires. (pp. 409-430). John Wiley & Sons, Inc.
- Domanska, O. M., Firnges, C., Bollweg, T. M., SØresnsen, K., Holmberg, C., & Jordan, S. (2018). Do adolescents understand the items of the European Health Literacy Survey Questionnaire (HLS-EU-Q47) – German version? Findings from cognitive interviews of the project “Measurement of Health Literacy Among Adolescents” (MOHLAA) in Germany. *Archives of Public Health*, 76(46).
<https://doi.org/10.1186/s13690-018-0276-2>
- Elkind, D. (2001). Cognitive Development. In J. V. Lerner & R. M. Lerner (Eds.), *Adolescence in America: An encyclopedia* (Vol. 1, pp. 127-134). ABC-CLIO, Inc.
- Fowler, F. J. Jr., Lloyd, S. J., Cosenza, C. A., & Wilson, I. B. (2016). Coding cognitive interviews: An approach to enhancing the value of cognitive testing for survey question evaluation. *Field Methods*, 28(1), 3-20.
<https://doi.org/10.1177/1525822X14549921>
- Ginsburg, H. P., & Opper, S. (1988). *Piaget's theory of intellectual development* (3rd ed.). Prentice-Hall, Inc.
- Hall, S., McKinstry, C., & Hyett, N. (2106). Youth perceptions of positive mental health. *British Journal of Occupational Therapy*, 79(8), 475-483.
<https://doi.org/10.1177/0308022616632775>
- Hinds, J. T., III, Loukas, A., Chow, S., Pasch, K. E., Harrell, M. B., Perry, C. L., Delnevo, C., & Wackowski, O. A. (2016). Using cognitive interviewing to better assess young adult e-cigarette use. *Nicotine & Tobacco Research*, 18(10), 1998-2005.
<https://doi:10.1093/ntr/ntw096>
- Hofmeyer, A., Sheingold, B. H., & Taylor, R. (2015). Do you understand what I mean? How cognitive interviewing can strengthen valid, reliable study instruments and dissemination products. *Journal of International Education Research*, 11(4), 261-267.
<https://doi.org/10.19030/jier.v11i4.9460>
- Hsu, C. & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation*, 12(10), 1-8.
<https://doi.org/10.7275/pdz9-th90>
- Irwin, C. W., & Stafford, E. T. (2016). Survey methods for educators: Collaborative survey development (part 1 of 3) (REL 2016–163). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands.
<http://ies.ed.gov/ncee/edlabs>
- Irwin, D. E., Varni, J. W., Yeatts, K., & DeWalt, D. A. (2009). Cognitive interviewing methodology in the development of a pediatric item bank: A patient reported outcomes measurement information system (PROMIS) study. *Health and Quality of Life Outcomes*, 7(3), 1-10. <https://doi.org/10.1186/1477-7525-7-3>
- Joffer, J., Jerdén, L., Öhman, A., & Flacking, R. (2016). Exploring self-rated health among adolescents: A think-aloud study. *BMC Public Health*, 16(156), 1-10.
<https://doi.org/10.1186/s12889-016-2837-z>
- 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.
<https://doi.org/10.1080/00461520701416231>
- Karpen, S. C., & Hagemeyer, N. E. (2017). Assessing faculty and student interpretations of AACPE survey items with cognitive interviewing. *American Journal of Pharmaceutical Education*, 81(5), 1-9.
<https://doi.org/10.5688/ajpe81588>
- Katz, C., McLeigh, J. D., & szwec, B. E. (2017). Preschoolers’ perceptions of neighborhood environment, safety, and help-seeking. *Children and Youth Services Review*, 73, 197-204.
<http://dx.doi.org/10.1016/j.childyouth.2016.12.016>
- Kenny, U., Molcho, M., & Kelly, C. (2017). Young people’s perspectives in developing a survey item on factors that influence body image. *Young*, 25(4S), 54S-72S.

- <https://doi.org/10.1177/1103308816674136>
- Koskey, K. L. K., Karabenick, S. A., Woolley, M. E., Bonney, C. R., & Dever, B. V. (2010). Cognitive validity of students' self-reports of classroom mastery goal structure: What students are thinking and why it matters. *Contemporary Educational Psychology*, 35, 254-263. <https://doi.org/10.1016/j.cedpsych.2010.05.004>
- Kramer, J. M., & Schwartz, A. (2017). Refining the Pediatric Evaluation of Disability Inventory – Patient-Reported Outcome (PEDI-PRO) item candidates: Interpretation of a self-reported outcome measure of functional performance by young people with neurodevelopmental disabilities. *Developmental Medicine and Child Neurology*, 59, 1083-1088. <https://doi.org/10.1111/dmcn.13482>
- McCoy, H. (2014). Using cognitive interviewing to explore causes for racial differences on the MAYSI-2. *Crime & Delinquency*, 60(5), 647-666. <https://doi.org/10.1177/0011128710388922>
- McDavid, J. C., Huse, I., & Hawthorn, L. R. L. (2013). *Program evaluation and performance measurement: An introduction to practice* (2nd ed.). Sage Publications, Inc.
- Moely, B. E., Mercer, S. H., Ilustre, V., Miron, D., & McFarland, M. (2002). Psychometric properties and correlates of the Civic Attitudes and Skills Questionnaire (CASQ): A measure of students' attitudes related to service-learning. *Michigan Journal of Community Service Learning*, 8(2), 15-26. <http://hdl.handle.net/2027/spo.3239521.0008.202>
- Mulcahey, M. J., Calhoun, C., Riley, A., & Haley, S. M. (2009). Children's reports of activity and participation after sustaining a spinal cord injury: A cognitive interviewing study. *Developmental Neurorehabilitation*, 12(4), 191-200. <https://doi.org/10.1080/17518420902998177>
- Newcomer, K. E., & Triplett, T. (2015). Using surveys. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 344-382). John Wiley & Sons, Inc.
- Norris, A. E., Torres-Thomas, S., & Williams, E. T. (2014). Adapting cognitive interviewing for early adolescent Hispanic girls and sensitive topics. *Hispanic Health Care International*, 12(3), 111-119. <https://doi.org/10.1891/1540-4153.12.3.111>
- Pepper, D., Hodgen, J., Lamesoo, K., Koiv, P., & Tolboom, J. (2018). Think aloud: Using cognitive interviewing to validate the PISA assessment of student self-efficacy in mathematics. *International Journal of Research & Method in Education*, 41(1), 3-16. <https://doi.org/10.1080/1743727X.2016.1238891>
- Powell, R. R. (2006). Evaluation research: An overview. *Library Trends*, 55(1), 102-120. <https://muse-jhu-edu.ezproxy.montclair.edu/article/202358>
- Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, 68(1), 109-130. <https://doi.org/10.1093/poq/nfh008>
- Rebok, G., Riley, C. F., Starfield, B., Green, B., Robertson, J., & Tambor, E. (2001). Elementary school-aged children's reports of their health: A cognitive interviewing study. *Quality of Life Research*, 10(1), 59-70. <https://www.jstor.org/stable/4037607>
- Reeve, B. B., Willis, G., Shariff-Marco, S. N., Breen, N., Williams, D. R., Gee, G. C., Alegria, M., Takeuchi, D. T., Stapleton, M., & Levin, K. Y. (2011). Comparing cognitive interviewing and psychometric methods to evaluate a racial/ethnic discrimination scale. *Field Methods*, 23(4), 397-419. <https://doi.org/10.1177/1525822X11416564>
- Roberts, H. (2017). Listening to children and hearing them. In P. Christensen & A. James (Eds.), *Research with children: Perspectives and practices* (3rd ed., pp. 142-159). Routledge.
- Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach* (8th ed). Sage Publications, Inc.

- Rothgeb, J., Willis, G., & Forsyth, B. (2007). Questionnaire pretesting methods: Do different techniques and different organizations produce similar results? *Bulletin of Sociological Methodology*, 96, 5-31. <http://pascal-francis.inist.fr/vibad/index.php?action=search&terms=19560247>
- Ryan, K., Gannon-Slater, N., & Culbertson, M. J. (2012). Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation*, 33(3), 414-430. <https://doi.org/10.1177/1098214012441499>
- Sabo, K. (Ed.). (2003) Youth participatory evaluation: A field in the making. *New Directions in Evaluation*, 98, 1-12.
- Scott, J. (2008). Children as respondents: The challenge for quantitative methods. In P. Christensen & A. James, (Eds.). *Research with children: Perspectives and practices* (2nd ed., pp. 87-108). Routledge/Taylor & Francis Group.
- Sebastian, R. A., Ramos, M. M., Stumbo, S., Mcgrath, J., & Fairbrother, G. (2014). Measuring youth health engagement: Development of the Youth Engagement with Health Services Survey. *Journal of Adolescent Health*, 55(3), 334-340. <https://doi.org/10.1016/j.jadohealth.2014.02.008>
- Silva, J. R., Fera, B., Sudula, S., Koetzle, D., & Schwalbe, C. (2019). Cognitive interviewing to improve questionnaires for justice-involved youth. *International Journal of Offender Therapy and Comparative Criminology*, 63(10), 1931-1951. <https://doi.org/10.1177/0306624X19839597>
- Tatian, P. (2016). Performance measurement to evaluation. Urban Institute. <http://www.urban.org/sites/default/files/alfresco/publication-pdfs/2000555-Performance-Measurement-to-Evaluation-March-2016-Update.pdf>
- Teal, R., Enga, Z., Diehl, S. J., Rohweder, C. L., Kim, M., Dave, G., Durr, A., Wynn, M., Roman Isler, M., Corbie-Smith, G., & Weiner, B. J. (2015). Applying cognitive interviewing to inform measurement of partnership readiness: A new approach to strengthening community-academic research. *Progress in Community Health Partnerships: Research, Education, and Action*, 9(4), 513-519. doi:10.1353/cpr.2015.0083
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. B. Jabine, M. L. Straf, J. N. Tanur, & R. Tourangeau, (Eds.). *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73-100). National Academy Press. <http://doi.org/10.17226/930>
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- Urban, J. B., Linver, M. R., Moroney, D., Nichols, T., Hargraves, M., Roberts, E. D., Quinn, J., Brown, M., Gama, L., Gorsky, D., Doubledee, R., & Cox, M. (2019). Developing and testing a theory of change for Boy Scouts of America. *Applied Developmental Science*. <https://doi.org/10.1080/10888691.2020.1844008>
- Urban, J. B., & Trochim, W. (2009). The role of evaluation in research-practice integration: Working toward the "Golden Spike." *American Journal of Evaluation*, 30(4), 538-553. <https://doi-org.ezproxy.montclair.edu/10.1177/1098214009348327>
- Vreeman, R. C., Nyandiko, W. M., Ayaya, S. O., Walumbe, E. G., & Inui, T. S. (2014). Cognitive interviewing for cross-cultural adaptation of pediatric antiretroviral therapy adherence measurement items. *International Journal of Behavioral Medicine*, 21(1), 186-196. <https://doi.org/10.1007/s12529-012-9283-9>
- Willis G. B. (1999). Cognitive interviewing: A "How To" guide. Proceedings from the 1999 Meeting of the American Statistical Association. <https://www.hkr.se/contentassets/9ed7b1b3997e4bf4baa8d4eeced5cd87/gordonwillis.pdf>
- Willis, G.B. (2005). *Cognitive interviewing: A tool for improving survey design*. Sage Publications, Inc.
- Willis, G. B., & Artino, A. R. Jr. (2013). What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. *Journal of*

- Graduate Medical Education, 5(3), 353-356. <https://doi.org/10.4300/JGME-D-13-00154.1>
- Willis, G. B., & Miller, K. (2011). Cross-cultural cognitive interviewing: Seeking comparability and enhancing understanding. *Field Methods*, 23(4), 331-341. <https://doi.org/10.1177/1525822X11416092>
- Woolley, M. E., Bowen, G. L., & Bowen, N. K. (2004). Cognitive pretesting and the developmental validity of child self-report instruments: Theory and applications. *Research on Social Work Practice*, 14(3), 191-200. <https://doi.org/10.1177/1049731503257882>
- Zeldin, S., Bestul, L., & Powers, J. (2012). Youth-adult partnerships in evaluation (Y-AP/E): A resource guide for translating research into practice. ACT for Youth Center of Excellence, Cornell University. http://actforyouth.net/resources/n/n_yap-eval-guide.pdf

Appendix A: Summary of Cognitive Interviewing Results: Items, Reviewer Observations, Item Revisions, Response Errors, and Observation Codes

Item	Observation	Revision	Response Error Type* & Code
What Scouting ranks have you achieved?	<ul style="list-style-type: none"> One participant responded "None" 	<ul style="list-style-type: none"> Added a "None" category to response options 	RS Easy
Please select the trainings you've attended.	<ul style="list-style-type: none"> One participant did not attend any of the trainings listed 	<ul style="list-style-type: none"> Added a "None" category to response options 	RS Easy
If I am not naturally good at something, I will never do it well.	<ul style="list-style-type: none"> One participant recommended a "don't know" response choice 	<ul style="list-style-type: none"> Added "don't know" to the response options 	J/RS Easy
Do you serve as a mentor to a Cub Scout pack? If yes, how frequently do you attend Cub Scout meetings?	<ul style="list-style-type: none"> One participant recommended adding "three times a month" because it was the closest to how often he was going to Cub Scout meetings 	<ul style="list-style-type: none"> Added "three times a month" to response options 	RS Easy
What is the typical number of days you participated in this activity?	<ul style="list-style-type: none"> Participant answered "one day a week" but did not see answer listed. Selected "one to three times a month" instead 	<ul style="list-style-type: none"> Changed format of response options from three columns to one column and listed response choices in order from most to least frequent to make it easier to read 	RS Moderate
		<ul style="list-style-type: none"> Changed written "one" into numeric format to match the numeric format of the other response choices Dropped "only" from the response choices. For example, "only 2 days a week" changed to "2 days a week" 	RS Easy
I prefer the company of people who are very similar to me in background and expressions.	<ul style="list-style-type: none"> "Expressions" was defined by participants in different ways including "activities and interests"; "verbal communication, terms, vocabulary;" and one participant did not know what it meant One participant said he can get along with and work with anyone, but selected "agrees" from the response set. Seemed to have difficulty with the intended meaning of "prefer" as answer choice did not match explanation "Background" was defined in different ways, such as "they do the same things, they have the same religion, and they come from the same part of the earth." 	<ul style="list-style-type: none"> Removed "and expressions" from the item Added probe: "What do you think it means when people are similar to you in background?" Added probe, "What do you think about the word 'prefer' in this item?" 	C Difficult C Difficult
		<ul style="list-style-type: none"> The word "background" remained under review for pilot testing. 	C Difficult
Cultural diversity within a group makes the group more interesting.	<ul style="list-style-type: none"> Participant selected "sometimes true" and his explanation was consistent with response. When asked about the 	<ul style="list-style-type: none"> Added "don't know" answer choice 	RS Easy

	response choices, the same participant recommended adding "don't know."		
What summer or other notable Scouting experiences have you attended? (How frequently? —once, every year)	<ul style="list-style-type: none"> Several participants had not attended any of the events Participant interpreted question to mean only summer experiences; talked about BSA camp and camping outings in general; was not familiar with response choices like "Jamboree" Several participants did not recognize "Summer Resident Camp" to mean overnight camp and chose "Other" instead 	<ul style="list-style-type: none"> "None" option added Item changed to "What Jamboree or Summer Scout Camps have you attended?" Response options expanded to include camp options "Summer Resident Camp" response choice changed to "BSA Summer Camp (Overnight)" 	<p>RS Easy C Moderate</p> <p>RS Moderate C/RS Moderate</p>
During the past 6 months, what after school activities did you participate in?	<ul style="list-style-type: none"> One participant did not include what he did on weekends as part of "afterschool." Participant suggested additional response choices, such as "summer school," but he was not sure if it was included as "afterschool." One participant described answer format as "jumbled" One participant unsure where cultural/religious school fit in. From research team discussion, "service club" not selected and included as another response option. 	<ul style="list-style-type: none"> Item wording changed to "During the past 6 months, what out of school time activities did you participate in?" Summer school not included, not an after-school activity. Response option format changed from three overlapping columns to two distinct columns Religious activities added to response set to include religious youth groups, religious school, religious service attendance Probe added, "What do you think religious activities includes?" Service club response choice eliminated 	<p>J Moderate</p> <p>J/RS Moderate</p> <p>RS Easy</p> <p>RS Difficult</p> <p>C Difficult RS Easy</p>
What is the typical number of hours a day you participated in this activity?	<ul style="list-style-type: none"> Two participants requested "3 hours or more." And one participant answered "up to 5 hours." None of these participants recognized "more than 2 hours" as including their answer One participant answered one hour 30 minutes, but chose 30 minutes but less than 1 hour 	<ul style="list-style-type: none"> Response options organized as a vertical list to make it easier to read These response options remained under review for pilot testing 	<p>C/J/RS Difficult</p> <p>C/J/RS Difficult</p>
It is hard for a group to work well when the people involved come from very diverse backgrounds.	<ul style="list-style-type: none"> Participants were unclear as to what was meant by backgrounds. One participant answered, "Like some people don't like socialize well or some people like love to socialize or something like that." 	<ul style="list-style-type: none"> Item modified to: <i>It is hard for a group to work well when the people involved come from different racial, ethnic, or cultural backgrounds.</i> 	<p>C Difficult</p>
Interview item: Values are important in Boy Scouts. What are your core values?	<ul style="list-style-type: none"> Younger participants struggled with the word "values." One participant did not know what was meant by values and another had difficulty explaining core values stating, "Basically what can you 	<ul style="list-style-type: none"> Probe added to help participants understand the intent of the item.: <i>"What values do you live by?"</i> 	<p>C Difficult</p>

	<p>help if you're in a team with or what can you help yourself with in these kind of values."</p>	
<p>Interview item: Do you feel that your troop values your background, attributes, and experiences (for example your race, gender, culture, or orientation)? Why or why not?</p>	<ul style="list-style-type: none"> • Younger participants struggled with this item. One participant explained the item as asking, "Do you think your troop helps you or trust you and like will like trust you with a leadership position? Or something like that." 	<ul style="list-style-type: none"> • Probes added to assist with comprehension of item: "What kinds of things does your troop do to make you feel valued?" "What parts of your background, attributes, and experiences do they value?" • Item remained under review for pilot testing <p style="text-align: right;">C Difficult</p>

Note: * RS = Response Selection, J = Judgment, C =Comprehension