Using the Cognitive Pretesting Method to Gain Insight Into Participants’ Experiences: An Illustration and Methodological Reflection

Kristin L. K. Koskey

Abstract
Quantitative researchers traditionally use cognitive pretesting methods such as think-alouds to determine whether participants are interpreting close-ended survey items as intended. I applied this method to evaluate the score validity of quantitative survey instruments with a middle school and high school sample of students. Through these applications, the cognitive pretesting method also revealed to be a useful tool for cueing participants’ reflection on their experiences related to the phenomenon of interest. Qualitative researchers traditionally use more open-ended interviews guided by general “how” and “why” questions, which might not serve as sufficient prompts to elicit participants’ reflection on their experiences, especially younger participants. To date, there are no illustrative examples focused on how having participants think-aloud as they complete a quantitative survey can inform qualitative-focused or mixed research questions. Therefore, the purpose of this article was to illustrate how the cognitive pretesting method, where participants think-aloud as they complete a survey, can be used as an additional technique to gain insight into their experiences of a phenomenon. The phenomenon in this illustration was defined as the participants’ experiences with the science content that they were learning in school. Limitations and methodological challenges and recommendations for overcoming these barriers are also discussed.

Keywords
cognitive pretesting, cognitive interviewing, think-alouds, concurrent think-alouds

Qualitative researchers traditionally use more open-ended interviews guided by general “how” and “why” questions to gain insight into participants’ experiences. These questions might not serve as sufficient prompts to elicit participants’ reflection on their experiences, especially younger participants. To date, there are no illustrative examples focused on how having participants think-aloud as they complete a quantitative survey can be used as a tool for gaining insight into participants’ experiences beyond evaluating their experiences with the survey itself. The illustration in this article advances the development of qualitative research methods by providing an adaptation of a cognitive pretesting protocol as an alternative method for gaining insight into participants’ experiences using the familiar context of responding to a survey.

Cognitive pretesting, as applied in survey research, is used as a method for investigating whether participants interpret close-ended items as intended to inform revisions in item wording prior to larger administration of a survey or test (Tourangeau, Rips, & Rasinski, 2000; Willis, 2005). Although cognitive pretesting and cognitive interviewing are not new methods, they continue to evolve. Karabenick et al. (2007) contributed to the literature through outlining a specific procedure for conducting cognitive interviewing and analyzing data produced by think-alouds to inform the validity of inferences drawn from quantitative surveys. This procedure has since been applied in research evaluating the score validity of previously established and widely adopted instruments (e.g., Koskey, Karabenick, Woolley, Bonney, & Ammon, 2010). Briefly, this procedure involves having participants think-aloud as they respond to a survey. Participants are asked to share, in their
own words, what they believe the survey item is asking, what they would choose as their answer or rating, and why they chose that response. The researcher rates participants’ responses in terms of how closely their interpretations of the survey item aligned with what the survey developer intended to assess.

In this article, I focus on how researchers can move beyond using the cognitive pretesting method for informing survey item revisions. I invite you to join me in the journey that I share of my realization of the wider application of this method. In this article, cognitive interviewing as it relates to cognitive pretesting and its traditional uses are briefly overviewed. Within this description, I review the cognitive pretesting think-aloud procedure outlined by Karabenick et al. (2007). Next, I describe how a more structured prescribed procedure also can be used to cue participants’ reflection on their experiences and, then, I provide an illustration using data from a previous study. Finally, a reflection is provided on the limitations and methodological barriers that might be faced when employing this method.

Epistemological Frame

Prior to addressing the purposes of this article, it is necessary to discuss the lens from which I write. My formal training is in psychometrics or measurement and statistics. Thus, my initial tendency is to have a more postpositivist worldview described as “a replacement [of positivism] that is still bound to the quantitatively oriented vision of science” (Teddlie & Tashakkori, 2009, p. 69). As a result of this worldview, my research in the past was strictly quantitative. In fact, it is admittedly unusual for me to write in first person and to reflect on my subjectivity, as my formal training is to avoid doing so.

This being stated, I never aligned with being a purist and agreed with Reichardt and Rallis (1994) that “any given set of data can be explained by many theories” (p. 88). As I focused my research on constructing attitudinal measures and evaluating the psychometric properties of established measures, I found a number of patterns observed in quantitative data that could not be explained by numbers alone. Over the past few years, I aligned more with the pragmatist view, recognizing that both postpositivism and constructivism are valuable worldviews, and that the method adopted is driven by the research problem and question (Feilzer, 2010; Johnson & Onwuegbuzie, 2004; Teddlie & Tashakkori, 2009). Over the past 7 years, my colleagues and I have conducted a handful of studies employing cognitive pretesting methods to evaluate the quality of attitudinal measures and their associated rating scales to complement and to inform the data yielded from quantitative techniques for assessing score reliability and score validity. The call for this special issue was timely in that a few months prior to the call, I completed coding data elicited by cognitive think-alouds and experienced this realization of the greater potential use of this method in qualitative research.

Overview of Cognitive Interviewing in Pretesting Surveys

Cognitive interviewing is defined as an “intensive, extended, think-aloud, or laboratory pretest interview” (Willis, 2004, pp. 23–24). It is one of many cognitive pretesting methods used to evaluate whether participants interpret close-ended items on a survey as intended by the survey designer to provide validity evidence or to inform item revisions. Close-ended refers to survey items rated on a predefined or fixed scale such as a Likert-type scale ranging from strongly disagree to strongly agree. Cognitive interviewing falls under the Cognitive Aspects of Survey Methodology (Alwin, 2001; Sudman, Bradburn, & Schwarz, 1996; Willis, 2004), otherwise known as CASM. An information-processing framework (Shiffrin & Schneider, 1977) is applied in CASM to conceptualize participants’ unobservable thought processes in an attempt to understand causes of measurement error. Measurement error refers to “inaccuracy resulting from a flaw in a measuring instrument” (Bond & Fox, 2007, p. 312). These errors might be due to the item wording, response scale, reading difficulty level, or other factors associated with the instrument.

Cognitive interviewing has a long history of being applied to improve the survey design such as to inform the presentation of items and rating scale adopted (for a review, see Sirken et al., 1999; Sudman et al., 1996; Tourangeau et al., 2000). Nevertheless, as a method, it is typically “not carried out primarily for the purpose of developing questionnaire design, but rather, to evaluate targeted survey questions, with the goal of modifying these questions when indicated” (Willis, 2004, p. 23). These modifications are indicated when misalignment emerges between participants’ interpretations and a researcher’s intension. Participants’ qualitative responses then are used to guide revisions to individual items. One of the challenges identified in using cognitive interviewing for eliciting participants’ thought processes is “to gain information about the associated thought states without altering the structure of the course of the naturally occurring thought sequences” (Ericsson & Simon, 1998, pp. 180–181). The goal is to elicit verbalizations from participants who represent the thought processes that they would have if completing the task under normal circumstances.

Verbal probing or questioning is one method used in cognitive interviewing to elicit these processes. A think-aloud task is another method used in cognitive interviewing requiring participants to verbalize their thought process, which is the focus of this article. Traditionally, think-aloud tasks require limited interchange between the interviewer and participant (Ericsson & Simon, 1980, 1993). In a think-aloud task, participants are asked to verbalize their thought processes either during (concurrent verbalization) or immediately after (retrospective verbalization) completing a task. The interviewer simply instructs a participant to “say what you are thinking” as he or she completes the task or to “say what you were thinking” after completing the task. Some researchers provide a video of the participant completing the task to assist them in recalling what he or she was thinking if using retrospective verbalization.
Minimal reactivity effects have been found when participants think aloud while completing a survey (Ericsson & Simon, 1980; Fowler, 1995; Low, 1999; Willis, 2004). Specifically, verbalization during a self-administered survey requires what Ericsson and Simon (1980) classified as Level-2 verbalizations, wherein participants transform their thoughts into words to verbalize but without changing the sequence of those thoughts. In other words, the verbalizations expressed during the think-aloud are close to representing the thoughts as they naturally occur. Level-2 verbalizations differ from Level-1 verbalizations in that Level 1 does not require the thoughts to be transformed before they are verbalized. Boren and Ramey (2000) provide an example of participants “who verbalize sequences of numbers while solving a math problem” as producing Level-1 verbalizations “because numbers can be verbalized in the same form as they were originally encoded” (p. 262). In addition to Level-2 verbalizations, the task requires participants to process information in their consciousness as opposed to using automatic or nonconscious processing. Finally, the task involves verbal processing that takes at least a few seconds but not more than 10 s, further indicating that information is processed in the consciousness but not to the point of automatic processing (Ericsson & Simon, 1980).

Measures to minimize reactivity during a think-aloud task are recommended such as the use of minimal verbal probing to allow participants to “maintain undisturbed focus on completion of the presented task” (Ericsson & Simon, 1998, p. 181). If too much verbal probing is employed, then what Ericsson and Simon referred to as Level-3 verbalization might occur, requiring participants to retrieve information that they are not initially attending to prior to the verbal prompt. This retrieval could result in interference and an inaccurate representation of the thought processes that they would experience when completing the task under authentic conditions. Such retrieval is problematic if using cognitive pretesting for the purpose of determining what participants are thinking as they complete a survey to inform whether they interpret and interact with the items as intended. The reason that it is problematic is that the researcher does not want to prompt participants to reflect more than they normally would when responding to a survey item under normal conditions (e.g., sitting completing the survey on their own).

Imagine you are completing a survey asking you to rate how you feel about a series of statements. You most likely will read the first survey item, consider what it is asking, retrieve the needed memory, and then respond using the scale provided and move to the next item. You probably do not spend more than 30 seconds responding to each item. Too much verbal probing as to why you chose the answer or rating that you selected could lead you to retrieve information about your experiences that you normally would not think about if taking the survey without thinking aloud. This being stated, Hughes (2004) found differences in the usefulness of data produced when comparing different cognitive pretesting methods employing varying levels of verbal probing. Specifically, cognitive interviewing and respondent debriefing were more likely to identify problems with the survey wording or design compared to behavior coding of the interaction. Cognitive interviewing and respondent debriefing both provided for probing of participants’ thoughts through preestablished probes or follow-up questions, respectively. I propose that using verbal probing to access Level-3 verbalizations is not problematic when the purpose is to gain deep insight into participants’ experiences, attitudes, and feelings. Thinking aloud while completing a survey can purposefully be used to cue the retrieval and expression of participants’ experiences related to a phenomenon of interest.

**Karabenick et al.’s Cognitive Pretesting Method**

Karabenick et al. (2007) proposed a systematic procedure for implementing a cognitive think-aloud task and analyzing the verbalizations produced. Multiple methods for conducting cognitive interviews and analyzing the verbal data produced exist (DeMaio & Landreth, 2004; Hughes, 2004). Several researchers (e.g., Hastie, 1987; Strack & Martin, 1987; Sudman et al., 1996; Tourangeau, 1984) have proposed models applying an information-processing model to participants’ thought processes as they complete a survey. Similarly, we applied six information-processing steps to this method with the goal of accessing participants’ short-term memory databases. The short-term memory being accessed was the immediate interpretation of an item meaning. The long-term memory being accessed was the experiences on which participants reflect while explaining their answer choice to a survey item. The six information-processing steps are outlined in detail in Karabenick et al. (2007); a brief overview is provided here. Steps 1 and 2 entail reading and interpreting the survey item and storing that interpretation in working memory. Step 3 involves accessing memories, experiences, attitudes, and associations relevant to the survey item. Steps 4 through 6 involve reading and interpreting the rating scale to select a response option, using the memories elicited to inform which number or label best aligns with the experiences accessed.

Participants are guided through this six-step process during a one-to-one interview where they are asked to think-aloud as they complete the survey with an interviewer who follows a structured script. Prior to beginning the task, participants engage in a practice task with one or more items until they feel comfortable thinking aloud. Rather than simply asking participants to “think aloud,” the script includes three core prompts that provide guidance and structure on how to verbalize their thoughts. The three core prompts comprise: (a) what is this question trying to find out from you, (b) which number do you choose as your answer or which number represents how you feel, and (c) explain why you chose that answer. Limited verbal probing is used unless participants have difficulty verbalizing their thoughts. In these cases, an interviewer selects from a number of predefined probes such as “can you give me an example?” or “can you share a little more about why you chose that number?” This script is repeated for each item. After participants engage with 1–3 items, limited prompting is
needed to remind them to think-aloud. In fact, participants
often go through the steps for each item without needing these
guiding questions from the interviewer due to the repetitious
nature of what is requested as they think-aloud for each item.

The verbalizations then are transcribed for coding ideally by
multiple raters using three criteria, defined by Karabenick et al.
(2007) as: item interpretation congruency, coherent elaboration
congruency, and answer choice congruency. The three
criteria are rated for each participant for each survey item; an
a priori coding scheme is applied to inform the three compo-
ments. Item interpretation congruency relates to whether the
verbalization supported an interpretation aligned with the
intended meaning of the item. Coherent elaboration con-
gruency indicates whether the memories, experiences, and atti-
udes accessed and verbalized by a participant align with what
the item intends to elicit. This criterion might be broken down
into elaboration on the appropriate content and context to
reflect coherent elaboration concept and coherent elaboration
context ratings. Concept refers to the main idea of the item,
wheras context refers to any specific location (e.g., while in
mathematics class, outside of school, at work, and at home)
referred to in the item to indicate whether a participant reflects
on his or her experiences in the appropriate context. Finally,
answer choice congruency relates to whether the rating that a
participant selected aligns with his or her elaboration during the
verbalization of his or her thoughts. All three criteria are rated
as yes, no, or insufficient prompting. Insufficient prompting
refers to insufficient verbalization by a participant to make a
judgment as to whether the interpretation, elaboration, or
answer choice was congruent.

Further, raters make a more holistic judgment of the overall
alignment of a participant’s verbalization by assigning an over-
all cognitive validity rating. This rating represents to what
degree the overall verbalization when responding to an item
was congruent with the intended meaning of the survey item
(Karabenick et al., 2007). Karabenick et al. (2007) and Koskey,
Karabenick, Woolley, Bonney, and Ammon (2010) applied a
scale ranging from 0 to 4 in past research to rate the overall
cognitive validity. A score of 0 was used to represent when the
entire verbalization was inconsistent with the intended mean-
ing of the item. A rating of 1 was assigned if a participant
interpreted the concept or context of the item incorrectly across
the entire verbalization for an item. For instance, a participant
interviewed interpreted the item concept correctly when asked
to respond to the item, In my math class, trying hard is very
important. However, the participant reflected on his or her
swimming teacher rather than on his or her mathematics
teacher and, thus, interpreted the context differently than
intended. A rating of 2 or 3 was used if there were instances
of inconsistency across the verbalization with the intended
meaning but some evidence of consistent interpretation with
the intended meaning of the survey item. Finally, a rating of 4
was assigned if the item interpretation, elaboration, and answer
choice throughout the verbalization support an interpretation
aligned with the intended item meaning. As is often the case
when applying holistic rubrics, in applying Karabenick et al.’s
(2007) scale, I found it difficult to implement reliably, finding
higher interrater reliability when collapsing to a 0- to 3-point
scale whereby \(0 = \text{no consistency across item interpretation,}
\text{elaboration, or answer choice with intended meaning, }1 =
\text{consistency for one of the three criteria, }2 = \text{consistency for}
two of the three criteria, and }3 = \text{consistency for all three}
criteria.

The criteria for what constitutes a congruent interpretation,
elaboration, and answer choice are defined using coding sheets
created by the original survey developer or another expert on
the construct or phenomenon under study. A coding sheet is
created for each item and used to support raters as they judge
how closely a participant’s interpretation aligns with what the
survey developer designed the item to assess. An example
coding sheet is provided in Appendix A. The detailed pro-
dure for applying the codes is outlined by Karabenick et al.
(2007). Further, an example application has been provided by
Koskey et al. (2010) where the method was applied to evaluate
the quality of a widely used and previously established measure
of mastery goal structure (Patterns of Adaptive Learning Sur-
vey; Midgley et al., 2000).

After completing the coding independently, interrater reli-
bility then is computed across the raters and disagreements in
ratings assigned are resolved. The final ratings for the three
criteria for each item across participants then are aggregated.
These findings are used to indicate which items have a low
congruency rating for each criteria and an overall low congru-
ent validity rating to flag items needing revision. Specific revi-
sions are guided by the qualitative verbalizations collected.
Also, during the coding process, raters take notes as to any
observations in misinterpretations, which are, in turn, used to
inform revisions needed. For example, a word or term that is
commonly misinterpreted or misread might be highlighted in
the raters’ notes. Participants’ verbalizations with high cogni-
tive validity ratings can be referred to for ideas on how to
reword the item or to revise it to more simplistic terms.

Repurposing the Cognitive-Pretesting Method

Through my experience employing this method and coding the
verbalizations across a number of studies, I found having par-
ticipants think-aloud as they complete a survey could elicit rich
data on their experiences with the construct being measured,
beyond their interpretation of the item meaning. The traditional
purpose(s) of each component of the cognitive-pretesting
method and the proposed alternative purpose(s) are outlined
in Table 1.

As outlined in Table 1, the procedure for conducting the
think-aloud remains, with each component reframed for the
purpose of gaining insight on participants’ experiences, atti-
udes, or feelings related to the phenomenon or topic of interest.
Item interpretation serves as an initial cue then to instigate the
retrieval of memories, experiences, and associations related to
the phenomenon. Selecting an answer choice indicates the
degree or direction of the experience, attitude, or feeling. It is
important to note here that it is not being suggested that participants’ experiences can necessarily be quantified or be represented by a single rating or label. On the contrary, the choices provided may be numeric, qualitative labels, or even illustrations if interviewing a younger sample. The purpose of the answer choice is to gain general insight on the direction of their experiences or feelings (e.g., positive or negative, less to more, and weak or strong) to cue further elaboration and to provide an opportunity for a participant to share that the statement does not represent his or her experiences.

Asking a participant to elaborate on why he or she chose a particular answer or rating cues Level-3 verbalizations described earlier (Ericsson & Simon, 1998). It is at the point of elaboration wherein I find that the traditional and alternative purposes diverge the most. Traditionally, the purpose of elaboration is to assess the degree of validity of the verbalizations with intended item meaning and congruency with the answer choice selected (e.g., if a participant chose “strongly disagree,” his or her elaborations support that he or she strongly disagrees). Limited additional probing is used in the traditional method so as not to alter what they would be thinking about if they took the survey in an authentic context without thinking aloud. Also, at this point in the think-aloud, the interviewer typically has sufficient information to make a judgment as to whether a participant interpreted the item as intended and can move forward to the next item. Conversely, the proposed alternative purpose is to instigate deeper reflection in an effort to elicit deeper-level memories or experiences. To achieve this goal, additional probing such as requesting to share further examples from the participant’s own life might be needed, which is a departure from the traditional method. In my experience, the prompt asking “why did you choose that as your answer?” cues participants to provide examples to support their ratings; however, further prompting might be needed to gain a deeper understanding of their rationale.

An interesting phenomenon related to the latter purpose of instigating deeper reflection occurred during a think-aloud interview that I conducted with a middle school science student. The interview was conducted to inform the development of a measure focused on assessing the engagement of students with content that they were learning in the classroom related to properties of matter. The student had completed a unit on properties of matter a week before. In the course of the interview, the student was presented with the item I can’t help but see objects in terms of the properties of matter. Initially, the student responded that he did not think about seeing objects around him in terms of his properties of matter. When asked to justify his answer choice, the student followed this up claiming, “I can’t not see it because I’m taking this [survey] . . . .” The response was problematic for the score validity of the survey item because the student selected strongly agree and justified his choice by stating that he saw the things around him right at that moment in terms of properties of matter, but this was due to the task. He shared that he did not see things in terms of properties

### Table 1. Traditional and Newly Proposed Purposes of Cognitive Pretesting Method.

| Criteria               | Example Generic Prompts* | Traditional Purpose                                                                 | Alternative Purpose                                                                 |
|------------------------|--------------------------|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| Item interpretation    | What is the question trying to find out from you? | Inform whether a participant interprets meaning of item and understand specific terms used in the item as intended. | Initial cue of personal- and/or context-specific terms to instigate retrieval memories, experiences, attitudes, and associations related to phenomenon, experience, or topic. |
| Answer choice          | Which number/label do you choose as your answer? What number best represents how you feel? What label best represents your experience? | Determine whether a participant interprets the rating scale as intended, using lower ratings to represent less and higher rating to represent more of the construct being measured. | Cue participant to reflect on degree or direction of experience, attitude, or feelings. Collect data to triangulate with elaboration on degree of experiences, attitudes, or feelings. Inform whether the statement, question, or scale aligns with a participant’s experience, attitude, or feeling. |
| Elaboration            | Why did you choose that number/label? | Assess the alignment between the answer choice and verbalization (do they converge or diverge?) and if the memories, experiences, attitudes, and context verbalized align with what the item intends to access. | Instigate reflection related to a participant’s experiences, feelings, and attitudes in relationship to the predetermined number/label. |
| Additional probing     | Can you give an example? What is another reason you chose that number/label? | Prompt verbalization for those who need further cuing. Clarify elaboration verbalized accurately to determine whether aligns with intended item meaning. Elicit further insight into background and context of the individual participant and his or her experiences, attitudes, or feelings. Clarify elaboration verbalized accurately to represent a participant’s experience, attitude, or feeling. |

*Prompts can be adapted depending on the topic and rating scale (numbers, labels, numbers and labels) used.
of matter before this exact point in time. The item was intended to assess whether students see things in terms of properties of matter after completing a unit on this topic, not while taking the survey. However, for the purpose of gaining insight on the student’s experience, the response is not problematic; the further elaboration provided insight into how students see (or do not see) objects in terms of the content that they are learning and what might instigate this experience.

For instance, another student elaborated that he or she does not notice examples of the properties of matter outside of class and gave an example when he or she plays in the snow: “Disagree because it’s really my brain’s relaxing time while I’m out there so I’m not really thinking ‘Oh look at all the solid, look at all the water.’ I’m saying, ‘I can build a snow man!’” This student’s response communicated that he does not have this experience of seeing things through the lens of properties of matter. Still, in the context of the think-aloud, he or she was able to describe accurately an experience in which he or she might have been able to see things in this way. The think-aloud provided the opportunity for this student to make such a connection. In comparison, the exchange below illustrates a student elaborating that he or she does have this experience outside of class when asked what rating he or she would choose when presented with the item I look for examples outside of class of the properties of matter. It is possible that this student would not have shared this experience if asked more generally to “Describe your experience with properties of matter outside of the classroom?”

Student: Agree. I find them without them coming to me. My brother challenges me to find different kinds of things.

Interviewer: Can you give me an example of what you mean when you say you “have a brother challenge you to find different things”? What does he do?

Student: Like, once he challenged me to find a gas inside of a solid, so I went to the freezer and I got an ice cube that had a little bubble of air in it.

I do not propose that this method be used to replace more in-depth interviewing, but rather to assist participants in verbalizing what might otherwise be difficult experiences to retrieve without prompting. The process could complement a more open-ended in-depth interview that follows. Figure 1 outlines the overall phases of the process, not meant to be prescriptive, but rather offered as a framework for applying the repurposed

---

**Figure 1.** Outline for adapting the cognitive pretesting method.

- **Identify Phenomenon of Interest**
  - Determine the phenomenon, topic, or attitude of interest guided by the research problem and aligned with the research question(s).

- **Generate Multiple Statements**
  - Develop statements with potential to elicit memories related to the phenomenon, attempting to represent the realm of potential experiences.
  - These statements can be based on prior research, existing surveys, or other resources similar to those used to guide the development of interview questions.

- **Select Answer Choices**
  - Select a rating scale or labels providing for varying degree of experiences, attitudes, or feelings to be represented.
  - Consider the age of participants and align the scale or labels with the item (e.g., if asking about confidence, the rating scale should relate to degrees of confidence).

- **Employ Protocol**
  - Implement the protocol outlined by Karabenick et al. (2007), with added verbal probing at the point of elaboration to gain deeper insight into a participant’s experiences, attitudes, or feelings. Document body language during the interview.

- **Analyze Verbalizations**
  - Analyze verbalizations within each item, but also across the entire think-aloud to gain a holistic understanding of a participant’s experiences, attitudes, or feelings.
  - Consider if the answer choices converged or diverged from the elaborations (e.g., Does a subject indicate he or she agrees with a statement but share an elaboration that seems to support otherwise?).

- **Triangulate Data and Findings**
  - Generate inferences from the multiple sources of data including, verbalizations transcripts, field notes on the participant’s body language during the interview, and analysis results of the data.
  - Employ member checking to accurately represent the participant’s verbalizations.
alternative method for the purpose of gaining insight into participants’ experiences, attitudes, and feelings.

Regarding the analysis phase, the coding scheme proposed by Karabenick et al. (2007) can be applied or adapted to gain surface-level insight as to whether a participant’s experience, attitudes, or feelings aligned with the statements, focusing mainly on the coherent elaboration ratings. However, it is recommended that a more in-depth analysis be conducted on the resultant verbalizations within each item and across the items. Verbalizations for a single item are not intended to represent a participant’s experience, attitude, or feeling. A holistic analysis of the verbalizations across the items is necessary to infer and to represent a participant’s experiences. Thus, it is essential to generate multiple and a realm of statements to present to participants. If too narrow of a realm of statements are presented to participants, then the experiences shared might be limited. If a researcher selects statements or questions that only represent his or her experiences, selection bias could occur with the items selected impacting the results. To minimize the degree of selection bias in the statements or questions presented to participants, researchers should use prior literature and theory to guide the selection.

Leech and Onwuegbuzie (2007) recommend applying multiple data analysis techniques such as content analysis, keywords-in-context analysis, and constant comparative analysis and then to triangulate the findings to inform the research question(s). The analysis approach depends on the research purpose(s) and research question(s) being addressed. For instance, if the research question relates to gaining insight on a particular group of participants’ perceptions about a phenomenon, then a thematic analysis might be applied, followed by constant comparative analysis through identifying the emergent phrases or chunks across the think-aloud to determine codes. Throughout this process, the emergent codes are continuously compared to those previously generated from the data.

Example Illustration

To illustrate the insight that can be gained from employing this method, two cases are presented using data collected to examine the phenomenon of transformative experience. Briefly, transformative experience is a construct nested in Dewey’s (1938, 1934/1980) theory of aesthetics and described by Pugh (2011) as representing a student’s engagement with the content that they are learning, extending beyond the classroom. Pugh conceptualized transformative experience as consisting of three qualities: behavioral (motivated use for the content), cognitive (expansion of perception), and affective (experiential value). Motivated use is defined as the application of the content in one’s everyday life without being required to do so. Expansion of perception refers to seeing the world in a new way, through the lens of the content learned. The third quality, experiential value, relates to holding a value for the utility of the content in everyday experiences. Pugh and his colleagues (Girod & Wong, 2002; Pugh, 2004) previously attempted to gain insight into students’ transformative experiences through traditional interviewing techniques and open-ended questioning. They later sought to construct a measure guided by the theory and their findings from qualitative research studies to attempt to measure this holistic construct to assess the degree to which students engaged with content from inside to outside the classroom.

The two cases illustrated here were drawn from a study focused on evaluating the psychometric properties of adaptations of the previously established Transformative Experience Questionnaire (TEQ; Pugh, Kleshinski, Linnenbrink, & Fox, 2004). The TEQ is a 28-item instrument assessing students’ degree of engagement with content that they are learning in the classroom, with each item rated on a 4-point scale ranging from strongly disagree to strongly agree. Items were developed based on the three qualities of a transformative experience. The item wording was adapted to target three different contents based on what the middle school and high school students who participated in the study were learning at the time of the study, including the topics of properties of matter, heredity and variation of living things, and human rights.

In this study, qualitative think-aloud data from a subsample of students were quantitized and used to inform the quality of the items (if students interpreted the items and used the rating scale as intended) to provide validity evidence and to guide further item revisions. The students were asked on the student assent form if they were willing to participate in the think-aloud task. Of those who responded “yes,” 20 students were randomly selected to participate. The researchers verified on the day of the interview whether each student was still willing to participate. A total of 19 students participated.

Alternative sampling methods should be considered depending on the research purpose. For instance, if the purpose were to represent the realm of experiences, an alternative method would be to apply a sequential mixed methods research design by asking participants first to respond to the survey to identify participants with low scores, midrange scores, and high scores to represent the continuum of responses in the sample. It is also possible that sampling a small number of participants will not represent the realm of experiences. Indeed, Blair and Conrad (2011) found that when conducting cognitive pretesting for the purpose of informing score validity, the probability of detecting issues with a survey increased as the number of participants increased. Literature on determining sample sizes for interviewing provides further guidance on sampling (e.g., Creswell, 2007; Graves, 2002; Guest, Bunce, & Johnson, 2006).

I chose to present the responses of two students’ verbalizations across 3 of the 28 items for three major reasons. First, the richness of the data yielded from the verbalizations could not be observed from the think-aloud for a single item alone. The students’ degree of transformative experience had to be examined holistically across the multiple items intended to represent a transformative experience. Second, it was too cumbersome to present a full transcript for the entire think-aloud task for all 28 items that would more accurately represent the richness of the data. Instead, 3 items focusing on engagement with the content outside of school were selected for illustrative purposes here.
Third, a case where a higher degree of transformative experience and a case where a lower degree of transformative experience were reported as evidenced by the answer choice ratings were selected. The purpose of selecting a case from both extremes was to demonstrate how this method might overcome the potential of statements to bias a participant’s responses. The second case also demonstrated that the task could provide the opportunity for a participant to disagree with the experiences in the statements as presented.

Table 2 presents the transcript for a ninth-grade, 15-year-old female exhibiting a high degree of transformative experience with the topic of human rights through agreeing or strongly agreeing to the statements. She self-reported her race as multicultural, her current grade in World History as an “A,” and her grade point average as 4.5 on a 5-point scale. Table 3 presents the transcript for a sixth-grade, 12-year-old male exhibiting a low degree of transformative experience with the topic of properties of matter through disagreeing to the statements. He self-reported his race as White and his current grade in Science as an “A.” This student did not report his grade point average. Both students were from a suburban school district in the Midwest in the United States. A researcher trained in cognitive pretesting conducted the interviews. The interviews took place in a study room in the school library and each were approximately 30 min in duration.

The text in italics indicates item wording. The text within brackets represents added notation for the purpose of clarification. Each “...” represents a 1-s pause in the verbalization. The last column in both tables is titled “My Reflection.” In this column, the black text represents notes that I made when coding the student’s response to reflect on if the verbalization supports a coherent interpretation of the item (the original purpose of applying this method). Traditionally, coding of whether the interpretation was coherent or not with the intended item meaning would be the extent of the reflection. In this alternative approach, additional reflection on what ways each participant’s verbalizations support or do not support a transformative experience is relevant. Thus, the bold bracketed text are notes added at a later point in time that highlight my interpretation of what aspects of a participant’s response support a transformative experience (or lack thereof) and in what ways. As seen in the transcripts, opportunities for additional probing certainly exist. Here, the cognitive pretesting protocol was followed for the purposes of informing the survey item quality and it is through examining the data that I came to the realization of the potential alternative use of this method.

Methodological Considerations and Concluding Thoughts

Limitations exist when applying this method for the purposes of gaining insight on participants’ experiences, attitudes, or feelings. As with all research, techniques for minimizing limitations must be planned for when implementing this method. A few of these limitations and recommendations for how to address each are highlighted next.

First, as noted earlier, bias can occur in the selection process of the statements to present to participants. Researchers could select statements that represent their own experiences or perceptions. This limitation also exists when formulating open-ended interview questions. Techniques for minimizing this degree of bias are to have other experts or participants from the target sample review and corroborate the statements. Also, researchers should allow prior literature to guide the generation of statements. A related limitation is that an item presented earlier during the interview could prime the response to a later item, potentially presenting a testing and an instrumentation threat to internal validity (Campbell & Stanley, 1966).

Second, the statements selected might not represent the full realm of experiences or the experiences of participants interviewed, which is why I recommend complementing this task with additional open-ended interview questions after the task or in a different phase of the study. Some participants might need additional time to reflect on their thoughts. The transcripts presented in this article were purposefully selected as exemplars of participants able to verbalize their thoughts clearly as reflected in the clarity of the transcripts. The exemplars presented do not represent the realm of experiences across all of the students in this example. Some of the students were not able to verbalize their thoughts clearly or felt comfortable completing the think-aloud task. Of the 19 students who completed the task, 1 student requested to end the interview after 11 min because he or she felt uncomfortable thinking aloud. Other students lacked elaboration in their responses. As an example, this sixth-grade male adolescent had a pattern of difficulty in elaborating on his answer choice:

Interviewer: Okay, how about number three.
Student: I talk outside of class about the properties of matter.
Interviewer: And what’s that trying to find out from you?
Student: Like, if I like … I don’t know. I’m not good at this.
Interviewer: You’re doing just fine. Would you like to keep going or stop?
Student: Um … like it’s asking if I like, I don’t know.
Interviewer: Okay, what is it about this one that is confusing?
Student: Like, I don’t know what it’s really asking.
Interviewer: Okay, go ahead and read it one more time.
Student: I talk outside of class about the properties of matter. It’s like asking if I like talk to my friends or parents about it.
Interviewer: Yep. Good. And how would your rate that one?
Student: Strongly disagree.
Interviewer: And why do you say that?
Student: Because … I don’t know.

Similarly, the answer choices presented might be too limiting or not provide a representation of how a participant feels. Thus, the researcher needs to be prepared for participants to share alternative numbers or labels as their answer choices from those provided. For instance, when I applied the same method for investigating the score validity of a survey...
Table 2. Exemplar of a Transcript Illustrating Task-Eliciting Student Memory of Transformative Experience.

| Student/ Interviewer Transcript | My Reflection |
|---------------------------------|--------------|
| Student (Reading Item # 6) **Outside of class, I think about human rights.** | Coherent item interpretation—understands concept (thinking about without being required) and context (outside of class). |
| **Interviewer:** What do you think that item is trying to find out from you? | | Example illustrates grasps concept. [Example of expansion of perception and motivated use, thinking about the concept outside of class. Sees discrimination through lens of human rights. Applies human rights to case of discrimination.] |
| **Student:** Do you think about human rights not when you’re learning about it, but maybe you’re rethinking about what you learned? You’re not thinking about it in class, you think about it somewhere else. It doesn’t have to be right in front of you to think about it. | | \*Example from outside of class showing grasps item focusing on that context. [Example from outside of social studies class, supporting transfer of concept and high rating.]* |
| **Interviewer:** And what number do you choose as your answer? | | Coherent item interpretation evident in example provided—concept (“do you think immediately?”). |
| **Student:** A four (strongly agree). | | \*Example illustrates grasps concept of seeing through new lens. [Example of expansion of perception and experiential value. Interprets case through lens of human rights.]* |
| **Interviewer:** Tell me a little bit about why you chose a four. | | \*Example of experiencing expansion of perception—seeing discrimination of African Americans through a new lens; seeing discrimination through lens of human rights.* |
| **Student:** Well, I read books a lot. I read One Flew Over the Cuckoo’s Nest and they were talking about how the inmates were discriminated against because some of them weren’t totally all there. Some of them were not very sane and sometimes I think about how people are treated today just because they’re different. | | \*Specific example supports grasps concept of understanding the world through lens of human rights, and high rating selected. [Example of experiencing expansion of perception—seeing discrimination of African Americans through a new lens; seeing discrimination through lens of human rights.]* |
| **Interviewer:** And did you read that book for social studies class? | | \*Example supports answer choice congruency, sharing does experience this. [Example of experiencing expansion of perception. Seeing war through lens of human rights, supporting high rating.]* |
| **Student:** No. | | \*Coherent interpretation of “can’t help but see.”* |
| **Student:** (Reading Item # 15) **I can’t help but see people, places, and events in terms of human rights now.** That’s trying to ask if you see a story on the news, do you think immediately if someone’s human rights have been violated or something like that? | | \*Example illustrates grasps concept of “better understanding” by stating now see those in past as being discriminated on when occurs again. [Example of experiencing expansion of perception.]* |
| **Interviewer:** Okay. And what answer do you choose for that question? | | \*Coherent item interpretation. Example illustrates grasps concept of “better understanding” by stating now see those in past as being discriminated on when occurs again. [Example of experiencing expansion of perception—seeing discrimination of African Americans through a new lens; seeing discrimination through lens of human rights.]* |
| **Student:** A four. Because often I see on the news, like recently there was a case of an African American man being beaten up by police unnecessarily and I thought that they violated his human rights, like he got pulled over for a speeding ticket, but they treated him like he had done something much worse and it was unnecessary and I think about how it had something to do with the fact that he was Black and they were White. | | \*Example supports answer choice congruency, sharing does experience this. [Example of experiencing expansion of perception. Seeing war through lens of human rights, supporting high rating.* |
| **Interviewer:** Okay, good example. What does it mean I can’t help but see things this way? | | \*Coherent item interpretation. Example illustrates grasps concept of “better understanding” by stating now see those in past as being discriminated on when occurs again. [Example of experiencing expansion of perception—seeing discrimination of African Americans through a new lens; seeing discrimination through lens of human rights.]* |
| **Student:** Like no matter what that’s the first thing to come to mind. | | \*Coherent interpretation of “can’t help but see.”* |
| **Interviewer:** Can you give me an example? | | \*Example supports answer choice congruency, sharing does experience this. [Example of experiencing expansion of perception. Seeing war through lens of human rights, supporting high rating.* |
| **Student:** Like whenever I see, sometimes on the news they show what’s going on in Afghanistan or Iraq right now, people might think the most important thing is getting the Taliban or Osama Bin Laden, but I think about the people who have gotten killed indirectly because of this conflict. | | \*Coherent item interpretation. Example illustrates grasps concept of “better understanding” by stating now see those in past as being discriminated on when occurs again. [Example of experiencing expansion of perception—seeing discrimination of African Americans through a new lens; seeing discrimination through lens of human rights.]* |
| **Student:** (Reading Item # 20) **Knowledge of human rights helps me to better understand the world around me.** That question is asking if you learned about human rights in the past. You might see the same groups being discriminated against today and you might think, “Oh, that’s what happened in the past and it’s happening again because these people were discriminated against.” | | \*Example supports answer choice congruency, sharing does experience this. [Example of experiencing expansion of perception. Seeing war through lens of human rights, supporting high rating.* |
| **Interviewer:** And what number would you choose for your answer? | | \*Coherent item interpretation. Example illustrates grasps concept of “better understanding” by stating now see those in past as being discriminated on when occurs again. [Example of experiencing expansion of perception—seeing discrimination of African Americans through a new lens; seeing discrimination through lens of human rights.]* |
| **Student:** A four. Because I see people of the Islamic nation being discriminated in our country and I think about how it’s like they’re the new African Americans of the 60s being discriminated against. People just don’t know much about them and their discrimination is based off of ignorance and bias. | | \*Specific example supports grasps concept of understanding the world through lens of human rights, and high rating selected. [Example of experiencing expansion of perception—seeing discrimination of African Americans through a new lens; seeing discrimination through lens of human rights.]* |
Table 3. Exemplar of a Transcript Illustrating Task-Eliciting Student Memory of Lack of Transformative Experience.

| Student/Interviewer | Transcript | My Reflection |
|---------------------|------------|--------------|
| Student:            | (Reading Item # 2): Question number 2. I think about the properties of matter when I see things like people cooking, grass being cut, leaves changing color, or ice melting. And this one is pretty much saying . . . It’s trying to find out that uh well after like leaving the class if you think about it. If you apply it to life. Like if you see a paper burning and think, “Oh, that’s a chemical change.” I’d have to say on this one actually disagree. I mean, sometimes I do, but really, it’s uh, it’s rarely, it’s not like I see my mom cooking and like I usually think, “Oh, what’s for dinner” not “Oh, she’s probably doing a chemical change to that toast.” | Coherent item interpretation of concept (applying concept to objects or events) and context (outside of class). Example aligns with disagreement rating. [Example illustrates can think this way but does not freely do so.] |
| Student:            | (Reading Item # 14) If I see a really interesting object (either in real life, in a magazine, or on TV), then I think about it in terms of the properties of matter. | Coherent item interpretation. Example of “wow” indicates grasps concept of interest. Example indicates differentiates between seeing an object just as the object vs. through lens of topic (properties of matter). [Example of how does not see objects through lens of properties of matter. Second example shows can see through this lens, but student shares does not view this way in reality.] Response option aligns with elaboration—gives example does not experience this. |
| Student:            | (Reading Item # 17) I notice examples outside of class of the properties of matter. | Coherent interpretation of concept (noticing examples on own) and context (outside of class). [Example illustrates can identify example from outside of class.] |
| Student:            | (Reading Item # 18): Question number 18. I think about the properties of matter when I see things like people cooking, grass being cut, leaves changing color, or ice melting. And this one is pretty much saying . . . It’s trying to find out that uh well after like leaving the class if you think about it. If you apply it to life. Like if you see a paper burning and think, “Oh, that’s a chemical change.” I’d have to say on this one actually disagree. I mean, sometimes I do, but really, it’s uh, it’s rarely, it’s not like I see my mom cooking and like I usually think, “Oh, what’s for dinner” not “Oh, she’s probably doing a chemical change to that toast.” | Coherent item interpretation of concept (applying concept to objects or events) and context (outside of class). Example aligns with disagreement rating. [Example illustrates can think this way but does not freely do so.] |

assessing attitudes toward statistics using a 10-point scale, a participant responded that he or she wanted to rate an item 100 to represent how strongly he or she agreed with a statement.

Additionally, the verbalizations could only yield a surface-level understanding of a participant’s experience, attitude, or feeling requiring the need for follow-up. However, again, the purpose here is to provide cueing of the memories, experiences, associations, and so forth related to the phenomenon of interest and then to probe for further elaboration. A participant’s interview is minimally guided by the few prompts and, thus, he or she does have a degree of freedom to verbalize information that is unexpected by the researcher.

Third, participants might not be comfortable in the situation or have experience thinking aloud, potentially limiting the amount of rich data. Providing practice thinking aloud and verbalizing one’s thoughts as they view a given statement is recommended prior to beginning the task. Additionally, there is potential for observation or reactivity effects, if participants act differently than they normally would due to the presence of an observer or interviewer. This effect could result in a threat to the internal validity of the data (Onwuegbuzie, 2003). For instance, providing socially desirable answers when responding to the survey items is a common phenomenon and potentially more likely to occur if responding aloud. One way to minimize the reactivity effect is to spend time establishing a sense of trustworthiness before beginning the task (Glesne, 2006), especially if asking about topics of a sensitive nature such as one’s experience with peer rejection. Also, I find that participants tend to demonstrate more relaxed body language and elaborate after engaging with a few items, supporting the importance of practicing responding aloud to a set of items before responding to the items of interest.
Fourth, divergence might occur between an answer choice selected for an item and the elaboration provided by a participant. This divergence can cause a barrier when analyzing the data and drawing inferences about participants’ experiences, attitudes, or feelings. For example, this sixth-grade student rated his agreement low at disagree regarding whether he sees objects around him in terms of the properties of matter but I perceived his elaboration as supporting that he does indeed see objects this way: “I’m tending to see simple things like a fan as a solid and Purell in the form of a liquid or photographic solution.” Divergence can indicate that follow-up probing is needed during the interview or after. Also, it could indicate that the answer choice options need revision. Further, divergence can raise interesting questions about how a participant perceives his or her own experiences, attitudes, and feelings. In this case, the student seems to have underrated or was too hard on himself when it comes to reflecting on whether he applies the content that he is learning at a higher level through seeing objects through the lens of properties of matter. Attending to whether a participant’s self-evaluation is consistently severe or not is important because it might serve not only to inform the items but also to provide insight on a participant’s perception.

Even when accounting for the limitations inherent in this method, this adaptation of the cognitive pretesting protocol can provide an additional data collection tool for qualitative researchers to consider. This method of presenting statements in the form of a survey has the potential to instigate and to cue participants’ memories that might not be otherwise elicited when using more general open-ended interview questions. Further, this method has the potential to assist participants in verbalizing their thoughts for those who are less vocal when presented with more open-ended prompts. As I generated this idea, I questioned the unique contribution, asking myself, “How is what I am proposing different from the traditional think-aloud method used for decades?” My answer to this question is that the difference is the notion of using prompts and presenting statements in a familiar format, a survey. Then, I questioned how my idea was unique from what Karabenick et al. (2007) already outlined.

The response to my own question was that the protocol is the same but the purpose of the method and approach to analysis differs, advancing the development of qualitative methods. The purpose shifts from having participants think-aloud as they complete a survey to evaluate score validity to using the think-aloud protocol when responding to a survey as a medium for eliciting participants’ experiences with a phenomenon. Analyzing the data shifts from focusing on how close the verbalizations at an item level agreed with the intended meaning to analyzing the verbalizations as a whole across the survey.

I encourage others to consider whether the cognitive pretesting method traditionally used for informing the score validity of survey items might be applied for alternative research purposes and to test further the proposed application with populations beyond adolescents. One example alternative purpose in research or in the clinical practice of teaching is to apply this method to gain insight into students’ misconceptions about the concept they are learning. Questioning strategies and think-aloud tasks are common methods used to reveal misconceptions held by students. A unique alternative approach might be presenting students with a survey consisting of a series of statements (or other form of stimuli) representing a realm of common misconceptions and facts related to the topic under study. By asking students to think-aloud as they rate how strongly they disagree or agree with the statements in a survey form, a researcher or practitioner might gain deeper insight into how deeply a student holds a misconception.

Appendix A

Example Coding Sheet Used to Rate the Congruency of a Participant’s Verbalization With the Intended Item Meaning

| Item 11: I look for chances to use my knowledge of the properties of matter in my everyday life. |
|-----------------------------------------------|
| **Item Interpretation:** Student actively looks for or seeks out opportunities to apply/makes use of knowledge about the properties of matter to solve a problem; to understand a situation; to answer a question; to understand objects or solids, liquids, or gases, and so on. |
| **Coherent elaborations** |
| Rating A—Item Content: Student refers to using or applying the ideas. Such use may involve solving a problem, understanding a situation, answering a question, understanding living things, completing an assignment, or similar entities. |
| Rating B—Item Context: Student references a process of actively looking for or seeking out an opportunity to apply the ideas. Student may reference anticipation about applying the ideas. Student may reference trying to apply the ideas. Student specifically refers to everyday life situations—not science class or school assignment situations. |
| **Congruent Answer Choice congruent with student’s interpreted item meaning may differ based on student interpretation. What the student selects as answer agrees with what expresses verbally. Students do not have to provide examples to be rated as congruent answer choice.** |
| 1 (Strongly disagree): |
| Student confidently disagrees that he or she engages in the item content—student expresses does not do this or feel this at all. (Student can’t provide example of supporting engages in item content or provides examples of how does not do or feel this.) |
| 2 (Disagree): |
| Student generally disagrees that he or she engages in the item content—student expresses does or feels this rarely, if at all. |
| 3 (Agree): |
| Student generally agrees that he or she engages in the item content—student expresses does or feels this to some extent. (Student might be able to provide examples supporting engages in item content to some extent.) |
| 4 (Strongly agree): |
| Student strongly agrees he or she engages in the item content—student expresses does or feels this to a good extent. (Student might be able to provide examples supporting engages in item content to a good extent.) |

Note. Coding criteria written by Dr. Kevin Pugh who developed the theory of transformative experience (see for a description of the theory of transformative experience, Pugh, 2011).
Acknowledgment
The author would like to thank you to Dr. Victoria C. Stewart for her ongoing feedback on the draft of this article. Her expertise in qualitative research and recommended edits resulted in a stronger article.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Data presented in this article are from research supported by an internal grant funded by The University of Akron, LeBron James Family Foundation College of Education.

References
Alwin, D. F. (2001). Research on survey quality. *Sociological Methods and Research, 20*, 1–39. doi:10.1177/00491241020201001
Blair, J., & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly, 75*, 636–658. doi:10.1093/poq/nfr035
Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
Boren, M. T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication, 43*, 261–278. doi:10.1109/47.867942
Campbell, D. T., & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago, IL: Rand McNally.
Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among the five approaches*. Thousand Oaks, CA: Sage.
DeMaio, T., & Landreth, A. (2004). Cognitive interviews: Do different methodologies produce different results? In S. Presser, J. Rothsgeb, M. Couper, J. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Questionnaire development, evaluation and testing methods* (pp. 89–108). New York, NY: Wiley Interscience.
Dewey, J. (1938). *Experience and education*. New York, NY: Macmillan.
Dewey, J. (1980). *Art as experience*. New York, NY: Perigee. (Original work published 1934)
Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251. doi:10.1037/0033-295X.87.3.215
Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: *Verbal reports as data* (Revised ed.). Cambridge, MA: MIT Press.
Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity, 5*, 178–186. doi:10.1207/s15327884mca0503_3
Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research, 4*, 6–16. doi:10.1177/1558689009349691
Fowler, F. J. (1995). Presurvey evaluation of questions. In F. J. Fowler (Ed.), *Improving survey questions: Design and evaluation* (pp. 104–137). Thousand Oaks, CA: Sage.
Gird, M., & Wong, D. (2002). An aesthetic (Deweyan) perspective on science learning: Case studies of three fourth graders. *Elementary School Journal, 102*, 199–224. doi:10.1086/499700
Glesne, C. (2006). *Becoming qualitative researchers: An introduction* (3rd ed.). Boston, MA: Pearson Education.
Graves, T. (2002). *Behavioral anthropology: Toward an integrated science of human behavior*. Walnut Creek, CA: Roman & Littlefield.
Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*, 59–82. doi:10.1177/1525822X05279093
Hastie, R. (1987). Information processing theory for survey researchers. In H. J. Hippler, N. Schwartz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 42–70). New York, NY: Springer-Verlag.
Hughes, K. A. (2004). Comparing pretesting methods: Cognitive interviews, respondent debriefing, and behavior coding (Research Report No. #2004-02). Retrieved from http://www.census.gov/srd/papers/pdf/rsm2004-02.pdf
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*, 14–26. doi:10.3102/0013189X033007014
Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., . . . Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*, 139–151. doi:10.1080/00461520701416231
Koskey, K. L. K., Karabenick, S. A., Woolley, M. E., Bonney, C. R., & Ammon, B. V. (2010). Cognitive validity of students’ self-reports of classroom mastery goal structure: What they are thinking and why it matters. *Contemporary Educational Psychology, 35*, 254–263. doi:10.1016/j.cedpsych.2010.05.004
Leech, N. L., & Onwuegbuzie, A. J. (2007). An array of qualitative data analysis tools: A call for data analysis triangulation. *School Psychology Quarterly, 22*, 557–584. doi:10.1037/1045-3830.22.4.557
Low, G. (1999). What respondents do with questionnaires: Accounting for incongruity and fluidity. *Applied Linguistics, 20*, 503–533. doi:10.1093/applin/20.4.503
Midgley, C., Maehr, M. L., Hruda, L., Anderman, E., Anderman, L., Freeman, K. E., . . . Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor: University of Michigan. Retrieved from http://www.umich.edu/~pals/PALS%202000_V13Word97.pdf
Onwuegbuzie, A. J. (2003). Expanding the framework of internal and external validity in quantitative research. *Research in the Schools, 10*, 71–90.
Pugh, K. J. (2004). Newton’s laws beyond the classroom walls. *Science Education, 88*, 182–196. doi:10.1002/sce.10109
Pugh, K. J. (2011). Transformative experience: An integrative construct in the spirit of Deweyan pragmatism. *Educational Psychologist, 46*, 107–121. doi:10.1080/00461520.2011.558817
Pugh, K. J., Kleshinski, O., Linnenbrink, E. A., & Fox, C. M. (2004, April). *Transformative experience in science: Using Rasch to develop a quantitative measure*. Paper presented at the meeting of the American Educational Research Association Annual Conference, San Diego, CA.
Reichardt, C. S., & Rallis, S. F. (1994). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. In C. S. Reichardt & S. F. Rallis (Eds.), The qualitative-quantitative debate: New perspectives (pp. 85–92). San Francisco, CA: Jossey-Bass.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review, 84, 127–190. doi:10.1037/0033-295X.84.2.127

Sirken, M. G., Hermann, D. J., Schechter, S., Schwarz, N., Tanur, J. M., & Tourangeau, R. (1999). Cognition and survey research. New York, NY: John Wiley.

Strack, F., & Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H. Hippler, N. Schwarz, & S. Sudman (Eds.), Social information processing and survey methodology (pp. 123–148). New York, NY: Spring-Verlag.

Sudman, S., Bradburn, N., & Schwarz, N. (1996). Thinking about answers: The application of cognitive processes to survey methodology. San Francisco, CA: Jossey-Bass.

Teddlie, C., & Tashakkori, A. (2009). Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences. Thousand Oaks, CA: Sage.

Tourangeau, R. (1984). Cognitive science and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), Cognitive aspects of survey methodology: Building a bridge between disciplines (pp. 73–100). Washington, DC: National Academy.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). The psychology of survey response. Cambridge, England: Cambridge University Press.

Willis, G. B. (2004). Cognitive interviewing revisited: A useful technique, in theory? In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), Methods for testing and evaluating survey questionnaires (pp. 23–43). New York, NY: John Wiley.

Willis, G. B. (2005). Cognitive interviewing: A tool for improving questionnaire design. Beverly Hills, CA: Sage.