International Education Research
Volume 2, Issue 1 (2014), 17-32
ISSN 2291-5273 E-ISSN 2291-5281
Published by Science and Education Centre of North America

What were They Thinking? Using Cognitive Interviewing to Examine the Validity of Self-Reported Epistemic Beliefs

Krista R. Muis[1*], Melissa C. Duffy[1], Gregory Trevors[1], John Ranellucci[1], & Michael Foy[2]

[1]Department of Educational and Counselling Psychology, Faculty of Education, McGill University, Montreal, Canada

[2]Department of Psychology, John Abbott College, Sainte-Anne-de-Bellevue, Canada

*Correspondence concerning this article should be addressed to Krista R. Muis, McGill University, 3700 McTavish Street, Montreal, Quebec, H3A 1Y2, or via e-mail at krista.muis@mcgill.ca

DOI: 10.12735/ier.v2i1p17

Abstract

We employed cognitive interviewing with a sample of secondary, college, undergraduate and graduate students to examine the cognitive validity of a popular epistemic beliefs self-report measure, the Discipline-Focused Epistemological Beliefs Questionnaire [DFEBQ] (Hofer, 2000). In addition, we examined cognitive validity across two domains. Analyses of interviews revealed that cognitive validity was good, wherein students' responses were typically within an expected range of interpretations. However, students' interpretations of items were not always consistent with researchers' intended meanings, interpretations sometimes differed across domains, and that the response option "3" as a neutral response was not always used as intended. To improve validity of self-report measures of epistemic beliefs more generally, we recommend that explicit anchors are used, such as "mathematician" instead of "expert," and that definitions of the dimensions are presented to respondents to ensure interpretations align with researchers' intended meanings. We end with broader methodological implications.

Keywords: epistemic beliefs, cognitive interviewing, domain differences

1. Introduction

"What is the truth?" When contemplating this question, an individual's answer may greatly vary depending on how it is interpreted. For example, one individual may believe the question concerns whether someone is telling the truth or telling a lie. In contrast, a second individual may ponder whether the question is of an epistemological nature. Is the statement merely someone's unsubstantiated belief, or is it a justified true belief? Clearly, interpretations and perspectives are important to consider when individuals respond to questions. This holds true for survey items wherein individuals are asked to express their beliefs about some abstract construct, particularly for beliefs about knowledge and knowing.

Over the past twenty years, research on individuals' beliefs about knowledge and knowing – epistemic beliefs – has become a prominent area in educational research. Various researchers have explored how beliefs develop (Kitchener & King, 1981; Kuhn, 1991; Perry, 1970), how different dimensions of epistemic beliefs influence learning, motivation, and achievement (Hofer & Pintrich, 1997; Schommer, 1990), and whether epistemic beliefs are domain general or domain specific (Buehl & Alexander, 2006; Muis, Bendixen, & Haerle, 2006). To measure individuals' beliefs, a number of methodologies have been used including interviews, vignettes, and self-report instruments. Unfortunately, methodological and conceptual issues have plagued research in this area (Buehl, 2008; Clarebout, Elen, Luyten, & Bamps, 2001; DeBacker, Crowson, Beesley, Thoma, & Hestevold, 2008; Hofer & Pintrich, 1997; Muis *et al.*, 2006). Specifically, self-report instruments have demonstrated low internal consistency and poor factor structure (e.g., DeBacker *et al.*, 2008). Conceptually, issues include the number and nature of the dimensions measured (Hofer & Pintrich, 1997), what "sophisticated beliefs" entail (Greene, Muis, & Pieschl, 2010a), the domain specificity of beliefs (Muis *et al*., 2006), and a lack of philosophical considerations in the development of theoretical frameworks (Chinn, Buckland, & Samarapungavan, 2011; Hofer & Pintrich, 1997; Muis *et al*., 2006).

By far, the vast majority of research that has been conducted on epistemic beliefs has relied primarily on self-reports given the ease with which they can be administered (DeBacker *et al*., 2008). Despite the prevalence of self-reports used to explore relations between epistemic beliefs and other constructs, research is needed that addresses the aforementioned issues. In particular, although several studies have explored the validity and reliability of many of the self-report instruments used today, only one study has been conducted to examine the cognitive processing that occurs when individuals respond to questionnaire items (Greene, Torney-Purta, Azevedo, & Robertson, 2010b). Given the measurement issues that currently plague research on epistemic beliefs, it is paramount that more validity research is conducted to further explore why reliability and validity are problematic with current self-report instruments. As such, following Karabenick and colleagues (Karabenick *et al.*, 2007), we applied the cognitive interviewing technique to more rigorously evaluate the cognitive validity of the most popular self-report instrument used to measure students' epistemic beliefs, the *Discipline-Focused Epistemological Beliefs Questionnaire* [DFEBQ] (Hofer, 2000). Through rigorous interviewing, recommendations can be made to improve current self-report instruments. More specifically, the purpose of this study was to explore whether students' interpretations of questionnaire items align with researchers' theoretical assumptions and construct definitions of epistemic beliefs, whether memories elicited to respond to those items fell within an acceptable range of responses, and whether responses chosen on the Likert scale were coherent with their elicited memories. A second purpose of this study was to examine whether cognitive validity varied across domain. Prior to detailing our study, to situate our analyses, we first briefly review Hofer's (2000) framework and then describe the cognitive interviewing technique.

1.1 Hofer's Multidimensional Framework

Based on an extensive review of the literature, Hofer and Pintrich (1997) delineated commonalities and differences across a variety of theoretical frameworks that explored, in one way or another, some form of epistemological thinking or reasoning (e.g., Belenky, Clinchy, Goldberger, & Tarule, 1986; Kitchener & King, 1981; Kuhn, 1991; Perry, 1970) or set of independent epistemological beliefs (e.g., Schommer, 1990). Although the number and nature of the dimensions differ across each of the frameworks, as Hofer (2000) proposed, there are commonalities that, when combined, reflect four belief dimensions about knowledge and knowing that vary along a belief continuum. These include: 1) the certainty of knowledge, ranging from knowledge is unchanging to knowledge is evolving; (2) the simplicity of knowledge, ranging from knowledge is organized as isolated bits and pieces to knowledge is organized as highly interrelated concepts; (3) the source of knowledge, ranging from knowledge is handed down by authority to knowledge is acquired through reason or logic; and, (4) the justification for knowing, which refers to how individuals consider how a proposition or belief becomes justified knowledge. The first two dimensions reflect individuals' beliefs about knowledge, whereas the last two dimensions reflect individuals' beliefs about knowing.

To extend previous work, Hofer (2000) further suggested that beliefs about knowledge and knowing were domain specific, as opposed to domain general as initially assumed. Based on empirical evidence that supported the notion of domain specificity, Hofer developed a theoretical framework to take into consideration differences that may arise in individuals' beliefs across a variety of domains. To test this theoretical framework, she developed the DFEBQ, which has become one of the most widely used instruments today. Unfortunately, difficulties have been noted with regard to establishing the reliability and validity of this particular instrument, despite its widespread use (see Muis *et al.*, 2006). As such, more work is needed to evaluate precisely what aspects of this instrument are problematic to make recommendations for moving forward with developing new items or a new instrument. Our study responds to this need by exploring the cognitive validity of this self-report instrument using the cognitive interviewing approach, which is described next.

1.2 The Cognitive Interviewing Approach

Given the widespread use of self-report instruments in psychological and educational research, survey methodologists have developed numerous ways in which to better understand the cognitive processing involved when individuals respond to survey items (Tourangeau, & Rasinski, 1988). The purposes of these approaches are to better inform survey development and to provide an additional means by which to rigorously evaluate the validity of existing surveys. Importantly, information garnered from cognitive interviewing goes beyond the common types of measurement validity assessed in educational research including face, content, factor, criterion, convergent and divergent, and construct validity (Shadish, Cook, & Campbell, 2002). As Shadish *et al.* (2002) note, researchers cannot assess only one type of validity to evaluate a measuring instrument. It is necessary to assess as many aspects of validity as possible to help establish whether the interpretation of scores is valid. Cognitive interviewing is one additional method by which validity can be established. In particular, cognitive validity refers to the degree to which respondents in a defined population construe items in a manner similar to researchers' intended meanings.

Unfortunately, as Karabenick *et al.* (2007) highlight, educational researchers have not generally considered whether the information gathered about individuals' thoughts, feelings, or beliefs are valid in the context of the self-report instruments designed to elicit them. As they argue, measurement validity is particularly vital in the assessment of more abstract constructs, such as epistemic beliefs, as responding to items requires individuals to engage in self-reflection and abstraction, which is particularly challenging. Because of the measurement issues that epistemic beliefs researchers have had with self-report instruments (Buehl, 2008; Clarebout *et al.*, 2001; DeBacker *et al.*, 2008), we deemed that the use of cognitive techniques would be particularly relevant to better understand what students are thinking when they respond to items designed to measure epistemic beliefs.

As previously noted, to date, only one study has applied the cognitive interviewing technique to examine the cognitive processes involved as individuals respond to survey items designed to measure epistemic beliefs (Greene *et al*., 2010b). Specifically, as a preliminary step in the development of his theoretical framework and corresponding measuring instrument, Greene and colleagues interviewed a small sample of elementary (four) and secondary school students (three) to assess their interpretations of items designed to measure beliefs about the certainty and simplicity of knowledge as well as the justification for knowing. Based on analysis of the cognitive interviews, Greene *et al*. found that students had difficulty responding to items that included words such as "truth" and "believe" and were unable to respond to items for domains in which they had little to no prior knowledge or experience. Given that the items Greene and colleagues used were specific only to his theoretical framework, which did not include items more commonly used in the literature, we deemed it pertinent to explore whether similar issues arose with samples across a broader educational span with a more commonly used instrument. As such, we employed the cognitive interviewing technique recommended by Karabenick *et al*. (2007).

According to Karabenick *et al*. (2007), from an information processing perspective, there are six cognitive tasks involved when an individual is asked to respond to a survey item. First, the respondent must be able to read the item and retrieve definitions of words from memory. Second, the retrieved meaning must be interpreted and stored in working memory. Third, the respondent must then search memories for information that is relevant to what the item is asking, such as experiences, thoughts, beliefs, or emotions. In the fourth step, the respondent must read and interpret the response options and, in the last two steps, must hold the item meaning, retrieve relevant information from memory, and process response options in working memory to evaluate all information and then select an appropriate response. As Karabenick *et al*. note, the three critical points at which individuals can provide information to assess the validity of an item are during item interpretation, retrieval of information, and selection of a response option to assess whether the response is congruent with the information retrieved from memory.

1.2.1 *Cognitive Posttesting*

The approach we chose to examine the validity of self-report items was cognitive posttesting. This approach is a concurrent, interviewing technique designed to gather information about cognitive processing of survey items (Willis, 2005). As Karabenick *et al*. (2007) propose, there are three critical steps of cognitive processing which can provide information to judge the validity of questionnaire items. In the first step, individuals can be probed for *item interpretation*. Item interpretation is considered valid if items are interpreted as researchers intend them to be interpreted. For the second step, *coherent elaborations*, questions are asked to assess whether respondents elicit acceptable memories of their experiences or beliefs. Finally, for the last step, questions target whether respondents provide a *coherent response choice* that is consistent with their coherent elaborations. To assess validity of responses, Karabenick *et al.* suggest the following: (1) clearly identify the validity coding criteria that operationalize the intended meanings; (2) conduct semi-structured interviews with targeted populations; (3) apply coding criteria to interview transcripts to quantify item validity; (4) analyze validity, and establish inter-rater reliability, and (5) modify items or make recommendations for modification based on validity performance.

1.3 The Current Study

Using the cognitive posttesting approach, we sought to examine how individuals interpreted items on a widely used self-report instrument, the DFEBQ (Hofer, 2000). This questionnaire was chosen for four reasons. First, as noted previously, the DFEBQ is the most widely used self-report instrument designed to measure individuals' epistemic beliefs. Second, the DFEBQ was developed to be domain specific and allows for comparisons to be made across more than one domain. That is, the domain of focus can be easily switched by referencing the domain in the instructions. Third, we chose the DFEBQ as researchers have previously reported low internal consistency estimates for each of the dimensions it is designed to measure (see Muis *et al*., 2006). Finally, we chose this questionnaire as it is representative of other instruments designed to measure similar dimensions, includes similarly worded items, and was adapted from previously developed domain-general questionnaires (e.g., the Perry Checklist of Educational Values, and Schommer's epistemological beliefs questionnaire [1990]) (Hofer, 2000). In this regard, the DFEBQ can also be used to assess how individuals might interpret items on more domain-general scales.

Secondary, college, undergraduate and graduate-level students were interviewed to assess the cognitive validity of the survey items across two specific domains: mathematics, and psychology. Our research questions were as follows: (1) Are students' interpretations of items on the DFEBQ consistent with their intended meanings? (2) Do students' elicited memories fall within an acceptable range of responses? (3) Are students' responses coherent with their elicited memories? (4) Are there differences in item interpretation as a function of domain? Given the exploratory and qualitative nature of this study, we did not develop any specific testable hypotheses.

2. Method

2.1 Participants

A sample of 34 students (N = 24 females) from four levels of education within various educational institutes volunteered to participate. Ten participants were secondary school students (N = 5 females, age range 15-16 years of age), seven were post-secondary collegiate[1] students (N = 5 females, age range 16-19), nine were undergraduate students (N = 7 females, age range 18-21), and eight were graduate students (N = 7 females, age range 23-30). The undergraduate students were sampled from four majors including biology (N = 1), economics (N = 2), psychology (N = 5), and physical therapy (N = 1). The graduate-level students were all pursuing advanced degrees in education. All students had some level of exposure to psychology and mathematics (e.g., courses taken, content covered in social sciences classes, etcetera).

2.2 Materials

Participants completed two versions of the DFEBQ (Hofer, 2000) adapted for two specific domains: mathematics, and psychology. Instructions at the beginning of the questionnaire prompted students to reflect on their beliefs about knowledge for psychology or mathematics. For both domains, a brief definition was provided to ensure that students' points of reference were similar. Students rated their agreement to each item using a 5-point Likert scale ranging from 1 (completely disagree) to 5 (completely agree). Lower scores on this scale represent more constructivist beliefs, with the exception of the justification dimension (higher scores indicate more constructivist beliefs).

2.3 Procedure

Participants were recruited through online and in-class advertisements. There were two components to the study: (1) completing questionnaires, and (2) participating in a cognitive interview. The session was scheduled for 1 hour (15 minutes to complete the questionnaires and 45 minutes allotted for the interview). At the beginning of the session, participants were informed that there were two different versions of the same questionnaire. The participants were advised to carefully read the instructions and to reflect on their beliefs about knowledge in the domain specified at the top of the questionnaire. In contrast to the protocol that Karabenick *et al*. (2007) recommended for cognitive interviewing (i.e., showing one item at a time and asking questions about that item before moving on to the next item), participants in our study completed each questionnaire first prior to asking them questions about specific items. We took this approach given that we wanted to compare individuals' answers across the domains and ask specific questions regarding similarities and differences in their responses on the rating scale, as described in more detail below. The interviewer then administered one questionnaire at a time. The order of the mathematics and psychology questionnaires was counterbalanced across participants.

When a participant returned a completed questionnaire, the interviewer recorded the participant's responses to five items that were selected specifically for interviewing. These items were chosen as they were representative of each of the four dimensions. The items were as follows: "Sometimes you just have to accept answers from the experts, even if you don't understand them" (source; item 3); "Most of what is true is already known" (certainty; item 9); "Ideas are really complex"

[1] In our country in one particular province, students who wish to pursue an undergraduate degree must first complete two years of college. Secondary school begins in grade 7 (age 12) and ends at grade 11 (age 17). College begins immediately following secondary school (ages 18-19).

(simplicity; item 10); "Correct answers in this field are more a matter of opinion than fact" (justification; item 12); and, "Experts can ultimately get to the truth" (attainability of truth; item 17). The interviewer recorded each participant's responses on the Likert scale for each of these items across both questionnaires using the format shown in Table 1 (the same items were used for all interviews to allow for comparisons in responses across individual cases).

Table 1. Example interviewer record sheet for participant Likert responses across two domains and five dimension

	Mathematics	Psychology	Notes
Item 3	1	5	D – M & P
Item 9	5	1	D - M & P
Item 10	4	4	S – M & P
Item 12	3	4	O – Interview only about mathematics
Item 17	2	3	O – Interview only about psychology

Note: D = very different responses to an item across the domains, S = very similar responses to an item across the domains, O = Focus on only one domain for interview questions. M = mathematics, P = psychology

The interviewer used this table to identify similarities, differences, and unique trends in response patterns by comparing each participant's responses for these items across the two questionnaires. First, for two items the interviewer focused on one survey each. There were several types of occurrences that prompted the interviewer to record a particular item response for questioning in the cognitive interview. For instance, the interviewer might have noted that a participant selected a strong response (i.e., a response that fell at either end of the Likert scale), or a neutral response (i.e., a response that falls in the middle of the Likert scale), or the same response across the two domains. Examples are shown in Table 1.

Next, the interviewer selected three items that showed the greatest contrast or similarity between the two surveys. For example, as shown in Table 1, the interviewer noted a contrast between one participant's responses for mathematics compared to psychology for item 3. This pattern was also found when comparing the participant's response for item 9, but in reverse order. Although there were several combinations of patterns that were identified for interviewing, the goal was to select response patterns that provided sufficient insight into participants' thought processes when completing the epistemic beliefs questionnaires across the two domains. This approach provided the interviewer with specific examples to guide the interview process.

2.3.1 *Cognitive Interviewing*

The cognitive interview was conducted after the participant completed all questionnaires. During the interview, participants were asked questions about their overall experience completing the questionnaires, their interpretation of specific items, and their answer choices for these items. A mix of open-ended questions (e.g., "Can you explain to me why you chose that as your response?") and structured questions with yes/no responses (e.g., "Did you have a specific example in mind when you chose this response?") were used throughout. One word or brief responses were typically followed by probes to elicit further details (e.g., "Why/why not? When? How? What?" "Can you tell me more?"). Domain-specific probes were also used (e.g., "Were you thinking about a specific type of math problem or math in general?").

Following the protocol outlines by Karabenick *et al*. (2007), there were three components to the cognitive interview. During the first phase, the interviewer asked questions about participants' response choice to an item on one survey. For example, to assess item interpretation, the interviewer would start by reading an item aloud and then asked: "When you read this item, what did it mean to you?" The interviewer would then ask the participant to explain his or her answer choice for that item. Next, the interviewer would ask whether the participant found anything difficult or confusing about the questionnaire item. This protocol was then repeated for a second item.

In the second phase of the interview, the interviewer asked participants about their answer choices for the same item across two domains. The item was first read aloud for each domain and then, depending on the answer choices, the participant was asked to explain the reason why he/she selected different or similar responses. For each participant, this comparison process was conducted for three items. All interviews were recorded using a digital recorder and transcribed verbatim.

2.3.2 *Cognitive Validity Coding*

First, we developed a coding guide similar to the one developed by Koskey, Karabenick, Woolley, Bonney, and Dever (2010). Twice in each interview students were asked about their cognitive processing on individual questionnaire items. Their responses to these interview questions were coded for cognitive validity. In particular, we first rated each student's response as acceptable, partially acceptable, or not at all acceptable (a rating of 2, 1, or 0, respectively) for item interpretation, coherent elaboration (e.g., acceptable memory), and congruent answer choice. Item interpretation was considered acceptable if the student offered an explanation of what the item was asking according to its intended meaning. For coherent elaboration, students' responses had to mention the domain of focus, and their experiences or perceptions had to be related to the item's intent. Finally, for coherent response choice, students' choice on the Likert scale had to be consistent with their elaboration. To guide our analysis, we used the definitions and intended meanings presented in Table 2. A global validity score was also calculated by summing the three separate cognitive validity scores. Three trained raters then independently rated each student's responses for all items based on the coding criteria.

To address our other research questions, a coding guide was developed through an iterative process of reviewing transcribed interviews. Although the coding scheme for this approach was primarily theory-driven, we also followed Creswell's (2013) recommendations of allowing additional codes to emerge during analyses. The unit of analysis for segmenting each transcript was based on naturalistic segmentation using the response to each interview question as an indicator of a separate idea (Chi, 1997). Segments from transcripts that were considered relevant to these categories were marked; the remaining responses were excluded from analyses. Responses were then coded across both domains using categories in the coding scheme. After independently coding transcripts for a particular education level, three coders met to discuss categories, themes, and relevant quotations to verify consistency, following a similar procedure outlined in Creswell (2013). Emerging themes and interpretations were verified with the first author to establish credibility (Creswell, 2002). Inter-rater reliability was established as 88% agreement. Disagreements were resolved through discussion with the first author.

Table 2. Researchers' definitions and range of acceptable answers for each of the dimensions

Category	Definition	Range of Beliefs
Certainty	**Certainty** is the degree to which knowledge is more *fixed* versus more *fluid.*	*Fixed* – Knowledge cannot be doubted. Everyone would come up with the same answer to a question and the answer to a question will probably never change. *Fluid* – Knowledge is not certain. Answers to questions are the best we know now but knowledge evolves as more information is gathered.
Simplicity	**Simplicity** is the degree to which knowledge is more a *set of facts* versus more a network of *highly interrelated concepts.*	*Set of facts* – Knowledge is one fact after another and facts are unrelated. Answers are straightforward. *Highly interrelated concepts* – Knowledge is interconnected concepts. The meaning of one concept is complex, relative to others, and depends on the situation where the concept is used.
Source	S**ource** focuses on whether people acquire knowledge from *outside the self* versus by actively *constructing it within the self*.	*Outside the self* – people believe knowledge is handed down by an authority (like a teacher or expert) and should not be questioned. *Within the self* – people believe they can construct knowledge by interacting with others, and using logic and/or evidence provided by experience.
Justification	**Justification** focuses on how claims become accepted as knowledge by relying more on *authority and experts* versus relying more on *personal experience and evaluation*.	Relying on *authority and experts* – people accept claims made by experts if experts reach consensus or agreement. Relying on *personal experience and evaluation* – people question experts, and evaluate information based on logic and/or evidence provided by experience.
Attainability of Truth	**Attainability of Truth** concerns the degree to which an individual believes that ultimate truth is *obtainable* versus *unobtainable.*	*Obtainable* – people believe it is possible to ultimately figure out the correct answer to any question and that there is a solution to every problem. *Unobtainable* means people believe that some things will never be known and some problems have no solution.

3. Results

The following sections describe quantitative and qualitative results from the cognitive interviews with graduate, undergraduate, college, and secondary school students. We divide the results related to cognitive validity across the three main questions that guided the coding process: (1) item interpretations; (2) elicited memories; and, (3) coherence of responses. We also highlight similarities and differences in each of these three categories across domain. Overall, the number of students who expressed unintended word interpretations, unacceptable memories, and inappropriate response choices is presented in Table 3. To provide a richer analysis, we separate out examples as a function of educational level. Sample interpretations and unacceptable memories from students are shown in Table 4. Inter-rater reliability, estimated using intra-class correlation coefficients (Shrout & Fleiss, 1979), for each of the three categories was within an acceptable range: item interpretations (.81), elicited memories (.62), and coherent response choice (.71).

Table 3. The number of students across educational levels who expressed beliefs related to each variable

Variable	Educational level			
	Secondary School	College	Undergraduate	Graduate
Unintended word interpretations				
Truth	1/10	4/7	3/9	0/8
Expert	1/10	2/7	3/9	1/8
Complex	3/10	2/7	4/9	2/8
Elicited memories				
Lack of specificity	5/10	4/7	6/9	3/8
Choice of 3	5/10	3/7	5/9	2/8

Table 4. Unintended interpretations of key words in DFEQB

Variable	Psychology	Mathematics
Truth		CS: What do they mean about the truth, do they mean like the right answer, the right method of going about answering the prob- solving the problem? So maybe that was a bit ambiguous. I: And how were you thinking about the truth when you answered that, how did you ultimately come down? CS: The right answer.
Expert	SSS: In psychology I was thinking of a therapist, analyzing the patient, what they're feeling.	CS: Mathematicians // Pythagorus
Complex	GS: A course that I took in psych I thought a lot of the ideas were complex and um not very easy for me to understand.	SSS: I thought it meant that ideas are really complicated. CS: I've always been very good at algebra so I don't think it's really complex UGS: For math, I just keep going back to some of the harder courses, I've taken math, which was very complex. GS: So the adjective is complex, so if I look at it from what the opposite of complex is simple, and something simple can be defined or stated in a few words and is easy to understand, easy for many people to understand, and my ideas of what psychologist work on isn't like that.

Note: SSS = Secondary School Student, CS = College Student, UGS = Undergraduate Student, and GS = Graduate Student

3.1 Item Interpretations

As shown in Table 5, participants' interpretations of questionnaire items approached acceptable levels, ranging from 1.33 to 1.65 out of a possible 2 points. However, there were several patterns that were noted that we deemed problematic. For example, student interview responses revealed that several words from items on the DFEBQ were interpreted in a manner that deviated from researchers' intended meanings. For example, when probed about their interpretation of the item "Ideas are really complex," across both domains, students typically interpreted *complex ideas* to be akin to knowledge that is difficult to learn or understand, as the following quotation from a graduate student illustrates: "… the opposite of complex is simple, and something simple can be defined or stated in a few words and is easy to understand." Similarly, when explaining why she felt knowledge in math was complex, an undergraduate student stated: "I don't understand, I just can't even begin to grasp what these things mean so I knew they can be really complex." These interpretations differ from the meaning intended by researchers, wherein "complex" denotes knowledge that has an intricate structure and is highly interconnected. In contrast to interpreting "complex" as being difficult to understand in mathematics, this same undergraduate student interpreted complex in psychology to mean complicated, as noted in the following quotation: "Humans are complicated and we come up with complicated explanations for things and they have to be complicated in order to describe the whole big wide world so yeah, stuff is complex." This variation suggests that not only do students espouse different interpretations of the word *complex* compared to researchers' intended meanings, but the meaning may shift across contexts as well.

Table 5. Mean cognitive validity scores by educational level

Education level	Item interpretation [a]	Appropriate memory [a]	Coherent response [a]	Global score [b]
Secondary	1.65	1.83	1.73	5.20
College	1.33	1.50	1.79	4.63
Undergraduate	1.57	1.86	1.75	5.18
Graduate	1.61	1.69	1.58	4.88
Overall	1.54	1.72	1.71	4.97

[a] Possible range of 0-2

[b] Possible range of 0-6

The word *truth* was also interpreted in an unexpected manner and varied according to domain. There were two prominent unintended interpretations of the word truth. In psychology, analysis of students' responses revealed an interpretation of truth that was less discernible and more refutable than mathematical truth. For example, as one undergraduate student compared math and psychology truths: "I think they're different. Because I, math I feel like math is something I can see on paper, anything complex they can work out for me, like something equals something, alright, it always comes out to that. Whereas psychology I feel like you can't really explain that, like nothing equals… even if you say something equals something you can always argue against it."..." Students also held distinct interpretations of *truth* while discussing mathematics. In this domain, *truth* was interpreted as being a "correct answer" to a mathematical problem or question. Confusion over the intended meaning of the word truth is demonstrated in the following response from a college student: "What do they mean about the truth? Do they mean the right answer, the right method of going about answering the--solving the problem? So maybe that was a bit ambiguous." These meanings of *truth* are not necessarily mutually exclusive from researchers' definition, but responses suggest that their understanding of these words may deviate from researchers' objectives.

Finally, students held unintended interpretations for the word *expert*; the word did not hold a clear or consistent definition for students across domains. In psychology, for example, definitions of *expert* ranged from historical figures, like Freud, to current practitioners. For mathematics, experts were defined as mathematicians. As one secondary student noted about psychologists, "Some [experts] deal with feelings, but other experts in psychology don't know much about feelings." Another college student described an expert in psychology as, "I was thinking about a behavioral psychologist…I even thought about pop psychology." This wide variation across students and domains shows no consistent interpretation for the word *expert*.

3.2 Elicited Memories and Coherent Responses

Interview transcripts were also analyzed to determine if students' memories elicited while responding to questionnaire items were appropriate and whether they were consistent with their response choice on the Likert scale. As shown in Table 5, the majority of students' elicited memories were considered within an acceptable range of responses (e.g., 1.50 – 1.86, overall = 1.72). Moreover, for most responses, students' choices on the Likert scale were considered coherent with their memories. Cognitive validity ratings ranged from 1.58 to 1.79 (overall = 1.71).

As an example of a coherent response choice, a secondary school student who believed that knowledge in mathematics was certain chose "5" on the Likert scale for his answer (which corresponds to "strongly agree") and then reported that: "In mathematics, it's more known. What's known is known. A lot of what they got is from ancient Greece, I guess. So, a really long time ago." In contrast to this coherent response, when asked whether there was a specific aspect of mathematics that came to mind when responding to a particular item, one respondent reported: "I was thinking of my experiences in theater because I do a lot of theater and doing carpentry and stuff like that because carpentry is a pretty straightforward field where like, is this two by four the right length?" As illustrated here, the individual's memory was unconventional, despite carpentry and measurement being related to mathematics.

Interestingly, although the majority of students' elicited memories were acceptable, the range of memories varied significantly from one respondent to the next, even within each respondent across domains. That is, individuals' responses varied across levels of specificity. For example, when respondents were asked about what specifically came to mind when responding to an item for psychology, one college student reported a very specific aspect about psychology: "I was thinking about Freud's theory, about the id" whereas one of the undergraduate students reported that nothing really specific came to mind. Similarly, for mathematics, one secondary school student responded, "Like math in general, like formulas, equations" as did one of the undergraduate students, "I thought about the subject as a whole, it was not like a situation." In contrast to these general examples, several students thought about something more specific about mathematics and, in some cases, contrasting notions of mathematics, such as the following quote from an undergraduate student: "I think I was looking at several, because math is incorporated in so many fields so it's kind of hard to limit math by itself, like you have math in physics, like I said there's real world math, there's math in finance, there's math in calculus, there's math in algebra… ."

Despite the general consistency between students' memories and their response choices, one aspect that we deemed problematic was the reasons students chose '3' on the Likert scale. Students selected '3' not because they endorsed a neutral stance for a statement, but because they held multiple conflicting examples in mind. In these instances, a choice of '3' represented a middle ground between two opposing beliefs as the following explanation from a college student demonstrates: "I put 3 because it's either sometimes I strongly agree for a certain scenario and sometimes I strongly disagree because there's so many different scenarios." A similar explanation was also provided by a graduate student: "I put it in the middle, because I can see, I can see both sides." An undergraduate student further explained that these types of diverging scenarios could represent beliefs about different subfields or perspectives within a domain, such as the social science versus natural science aspects of psychology.

In sum, analyses of interview transcripts for cognitive validity revealed that although students demonstrated coherent links between their interpretations of questionnaire items and memories elicited to make response choices, some words were interpreted differently than definitions used by researchers in the field. Furthermore, students reported that in some instances they held multiple contrasting examples in mind when responding to questionnaire items. In these cases, they reported a neutral stance to represent the conflict between two opposing ideas or beliefs. Implications of these findings are addressed next along with recommendations for the improvement in measuring epistemic beliefs.

4. Discussion

The purposes of this study were to explore the cognitive validity of a popular epistemic beliefs self-report instrument, the Domain-Focused Epistemological Beliefs Questionnaire (Hofer, 2000), and examine whether cognitive validity varied as a function of domain. Results from our study suggest that the DFEBQ is a valid instrument with regard to the three aspects of cognitive validity. However, there is certainly room for improvement. Specifically, students' interpretations of items were not always consistent with researchers' intended meanings, and interpretations differed across domains and educational levels. We elaborate each of these findings as a function of our research questions, and begin with interpretation of items. For each of our research questions, we make recommendations with regard to the improvement of instruments designed to measure epistemic beliefs.

4.1 Item Interpretations and Intended Meanings

Our first research question addressed the issue of whether students' interpretations of items were consistent with researchers' intended meanings. Results from the cognitive interviews revealed that three words were problematic: "complexity," "truth," and "expert." Rather than interpreting "complexity" of knowledge as being highly interconnected and related, students interpreted complexity to mean "difficult." This is particularly noteworthy for students' beliefs about mathematics, as students across all levels of education believe that mathematics is a difficult school subject (Muis, 2004). However, it should be noted that the dual meanings of "complex knowledge" – "difficult to understand" and "interconnected" – are not necessarily mutually exclusive. Indeed, knowledge that is interconnected would likely be perceived as requiring a greater degree of effort to understand and learn, which is exemplified by the following quotation from a secondary school student: "There's much more to it [mathematics], it's harder to understand. There's more to the understanding."

Similar to complexity, the words "truth" and "expert" were problematic from an interpretive standpoint, and interpretations differed as a function of domain. For truth, when responding to items for the domain of mathematics, students viewed "truth" as meaning that a question or problem in mathematics was either right or wrong. In contrast, "truth" was interpreted to mean factual or false, as opposed to the nature of reality, when responding to items about psychology. Interestingly, Greene *et al*. (2010b) found similar difficulties with the word "truth," but for their sample of students, truth was interpreted to mean opinion as opposed to fact. Finally, for our sample of students, interpretations of the word "expert" also differed across domains and, even within domains, students' reference points varied substantially.

Additionally, although our intentions were not to explicitly compare validity across the various educational levels, we found it noteworthy that the youngest of our sample, the secondary school students, had the highest validity scores, followed by the undergraduate students. In contrast, college, followed by graduate-level students, had the lowest validity scores. We have no plausible explanation for this pattern, but given that the majority of research is conducted with undergraduate students, we feel more confident in results from previous research. Moreover, this result boads well for researchers who wish to study younger students, who are underrepresented in current research.

Despite these patterns, improvements are needed. Given the problems we encountered with students' interpretations, we present several recommendations. First, items need to be reworded to improve clarity and interpretation. One way this might be achieved is to interview students with regard to what words they deem might be more representative of the meanings that researchers intend. For example, the item "Ideas are really complex" might be replaced with "Ideas [or knowledge] are highly connected [related]." For items that refer to "experts," replacing the word "expert" by a specific example relevant to the domain may improve item interpretation, such as using "mathematicians" for mathematics. By providing a specific anchor for what is meant by expert, students' elicited memories may be more homogeneous.

Additionally, to further ensure that respondents are interpreting items as intended, it may be helpful to present examples of how to interpret words embedded in items. Specifically, in the instructions, there may be definitions of each of the dimensions, with examples from each end of the continuum. For example, for the simplicity of knowledge, the following definition with examples may be provided. "Simplicity is the degree to which knowledge is more a *set of facts* versus more a network of *highly interrelated concepts* (complex)." The range of beliefs might then be described as follows: "*Set of facts* – knowledge is one fact after another and facts are unrelated. *Highly interrelated concepts* – knowledge is interconnected concepts. The meaning of one concept is complex, relative to others, and depends on the situation where the concept is used."

The use of Likert scales to measure epistemic beliefs has also been long debated (see Muis *et al*., 2006), and results from our study provide further evidence of the challenges researchers face when interpreting item responses. For example, students occasionally elicited several memories to respond to items but memories themselves were conflicting. When conflicts occurred, students opted for '3,' which is intended for a neutral response as opposed to a conflicting response. This pattern was in contrast to the pattern that Greene *et al*. (2010b) found with their sample of elementary and secondary school students. Rather than choosing '3' as an option when there were conflicting memories, students in their study chose the neutral position when they were unfamiliar with the domain of focus.

To address this issue, as Fowler (2001) recommends, one option might be to eliminate the middle 'neutral' position, which then forces individuals to select one side over the other. Another possibility might be to rewrite items such that they are more reflective of the frequency of students' thoughts or behaviors as opposed to their agreements with specific statements and arrange response options from "none of the time" or "not at all like me" to "all of the time" or "very much like me." For example, the item "Ideas are really complex" might be rewritten to the following: "When I think about knowledge in mathematics, I think about how ideas are interconnected" and response options could range from "none of the time" to "all of the time." Alternatively, we propose that perhaps individuals' beliefs can be measured in action; that is, it may be possible to measure enacted beliefs, or epistemic cognition, coupled with individuals' expressed beliefs to develop a better understanding of individuals' beliefs and how they may or may not be related to specific cognitive processes.

We also found interesting patterns of responses when students responded to items in the context of little prior knowledge. In particular, when students had limited exposure to psychology, they based their responses on their everyday experiences, as indicated in the following quote by one secondary school student: "[The psychologist] is taking what the patient tells him and he sees where he takes him… It might be two different things, two different disorders that it could be, so the psychologist matches it to this one or this one. He thinks which one it's more right, but it's just his opinion." For mathematics, however, students rarely elicited memories that were not affiliated with some classroom or school experience. This finding is consistent with previous research (e.g., Schoenfeld, 1988) wherein students reported not seeing the inherent value of mathematics beyond the classroom, or considering its relevance in a real-world context. Given this, we suggest that not only are students' beliefs influenced by the classroom and school experiences, but they are also shaped by experiences in everyday life.

References

[1] Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice, and mind*. Basic Books, New York.

[2] Buehl, M. M. (2008). Assessing the multidimensionality of students' epistemic beliefs across diverse cultures. In M. S. Khine (Ed.), *Knowing, knowledge and beliefs: Epistemological studies across diverse cultures* (p.65-112). New York: Springer.

[3] Buehl, M. M., & Alexander, P. A. (2006). Examine the dual nature of epistemological beliefs. *International Journal of Educational Research, 45*(1-2), 28-42.

[4] Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*(3), 271-315.

[5] Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist*, *46*(3), 141-167.

[6] Clarebout, G., Elen, J., Luyten, L., & Bamps, H. (2001). Assessing epistemological beliefs: Schommer's questionnaire revisited. *Educational Research and Evaluation, 7*(1), 53-77.

[7] Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Thousand Oaks, CA: Sage.

[8] Creswell, J. W. (2002). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

[9] DeBacker, T. K., Crowson, H. M., Beesley, A. D., Thoma, S. J., & Hestevold, N. L. (2008). The challenge of measuring epistemological beliefs: An analysis of three self-report instruments. *The Journal of Experimental Education, 76*(3), 281-312.

[10] Fowler, Jr. F. J. (2001). *Survey research methods (Applied social research methods series No. 1)* (3rd ed.). CA: Sage, Thousand Oaks.

[11] Greene, J. A., Muis, K. R., & Pieschl, S. (2010a). The role of epistemic beliefs in students' self-regulated learning with computer-based learning environments: Conceptual and methodological issues. *Educational Psychologist*, *45*(4), 245-257.

[12] Greene, J. A., Torney-Purta, J., Azevedo, R., & Robertson, J. (2010b). Using cognitive interviewing to explore primary and secondary students' epistemic and ontological cognition. In L. D. Bendixen & F. C. Haerle (Eds.), *Personal Epistemology in the Classroom: Theory, Research, and Implications for Practice* (pp.368-406). New York: Cambridge University Press.

[13] Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology, 25*(4), 378-405.

[14] Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88-140.

[15] 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*(3), 139-151.

[16] Kitchener, K. S., & King, P. M. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology, 2*(2), 89-116.

[17] 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*(4), 254-263.

[18] Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

[19] Muis, K. R. (2004). Personal epistemology and mathematics: A critical review and synthesis of research. *Review of Educational Research, 74*(3), 317-377.

[20] Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review, 18*(1), 3-54.

[21] Perry, W. G., Jr. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Hold, Rinehart and Winston.

[22] Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist, 23*(2), 145-166.

[23] Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82*(3), 498-504.

[24] Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (2nd ed.). Boston: Houghton-Mifflin.

[25] Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428.

[26] Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in Attitude Measurement. *Psychological Bulletin, 103*(3), 299-314.

[27] Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.

Author Note

Support for this research was provided by a grant to Krista R. Muis, recipient of the Early Career Research Award, from the American Psychological Association.

Copyrights