

This article was downloaded by:[Harvard University Graduate School]  
On: 30 June 2008  
Access Details: [subscription number 789066051]  
Publisher: Psychology Press  
Informa Ltd Registered in England and Wales Registered Number: 1072954  
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Measurement: Interdisciplinary Research & Perspective

Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title-content=t775653679>

### Assessing Elemental and Structural Validity: Data from Teachers, Non-teachers, and Mathematicians

Heather C. Hill <sup>a</sup>; Carolyn Dean <sup>b</sup>; Imani Masters Goffney <sup>b</sup>

<sup>a</sup> Harvard Graduate School of Education, Cambridge, Massachusetts, USA

<sup>b</sup> University of Michigan, Ann Arbor, Michigan, USA

Online Publication Date: 01 August 2007

To cite this Article: Hill, Heather C., Dean, Carolyn and Goffney, Imani Masters (2007) 'Assessing Elemental and Structural Validity: Data from Teachers, Non-teachers, and Mathematicians', *Measurement: Interdisciplinary Research & Perspective*, 5:2, 81 — 92

To link to this article: DOI: 10.1080/15366360701486999

URL: <http://dx.doi.org/10.1080/15366360701486999>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Assessing Elemental and Structural Validity: Data from Teachers, Non-teachers, and Mathematicians

Heather C. Hill

*Harvard Graduate School of Education, Cambridge, Massachusetts*

Carolyn Dean and Imani Masters Goffney

*University of Michigan, Ann Arbor, Michigan*

Validation efforts typically focus around what, exactly, is measured by an instrument(s), and whether what is measured corresponds to the theoretical domain(s) originally specified. In this paper, we conduct a first analysis into these issues. Our goal is building instruments focused around measuring the mathematical knowledge used in teaching: not only the content that teachers teach to students directly, but also the professional knowledge that helps support the teaching of that content. Following Kane (2001; 2004a) and as reported in Schilling & Hill (this issue), we developed two assumptions and related inferences to represent this thinking:

1. Elemental assumption: The items reflect teachers' mathematical knowledge for teaching and not extraneous factors such as test taking strategies or idiosyncratic aspects of the items (e.g., flaws in items).
  - A. Inference: Teachers' reasoning for a particular item will be consistent with the multiple-choice answer they selected.
2. Structural assumption: The domain of mathematical knowledge for teaching can be distinguished by both subject matter area (e.g., number and operations, algebra) and the types of knowledge deployed by teachers. The latter types include the following: content knowledge (CK), which contains

---

Correspondence should be addressed to Heather C. Hill, Harvard Graduate School of Education, Gutman Library, Room 445, 6 Appian Way, Cambridge, MA 02138. E-mail: heather\_hill@harvard.edu

both common content knowledge (CCK), or knowledge that is common to many disciplines and the public at large, and specialized content knowledge (SCK) or knowledge specific to the work of teaching; and knowledge of content and students (KCS), or knowledge concerning students' thinking around particular mathematical topics. Implications of this include:

- A. Inference: Items will reflect this organization with respect to both subject matter and types of knowledge in the sense that items reflecting the same subject matters and types of knowledge will have stronger inter-item correlations than items that differ in one or both of these categories. This will result in the appearance of multiple factors in an item factor analysis.
- B. Inference: Teachers can be reliably distinguished by unidimensional scores reflecting this organization by subject matter and types of knowledge. These scores are invariant with respect to different samples of items used to construct the scores.
- C. Inference: Teachers will tend to answer most problems (except those representing CCK) with knowledge specific to the work of teaching. Non-teachers will rely on test-taking skills, mathematical reasoning, or other means to answer these items.
- D. Inference: Teachers' reasoning for a particular item will reflect the type of reasoning (either CK or KCS) that the item was designed to reference.

Assessing these assumptions and interpretations will provide both practical and theoretical information about our measures. If the structural assumption is correct in identifying separate domains of knowledge, for instance, it provides a warrant for the construction of separate CK and KCS scores. It also implies that KCS, which has not yet been identified in large-scale survey research as a separate domain, is a genuine domain of measurement.

The elemental assumption is designed to address a possible critique of our items: that no test cast in a multiple-choice format could measure a complex and judgment-laden practice such as teaching. The critique is noted by Haertel (2004) in his response to Kane's original article; Haertel writes, "the activities of reading and responding to a multiple-choice question are quite unlike the activities required in professional practice." (p. 176). He concludes that this renders problematic Kane's connection between the test domain and the knowledge, skills, and practice domain the test is meant to represent.

This critique is also backed, in the scholarly literature, by a wave of criticisms of multiple-choice assessment in the early 1990s, including allegations that such assessments focused narrowly on basic skills (in mathematics, recall, and procedures) and did not measure students' ability to solve complex real-world problems (Boodoo, 1993). A review of the larger literature on test validity by Martinez (1999)

suggests that although multiple-choice assessments are not limited by their format to assessing lower level cognitive functions, evidence from existing exams suggest many do. Martinez also notes an additional challenge: that scores on multiple-choice exams may reflect, in part, “test-wiseness”—examinees’ ability to recognize cues, deploy response elimination strategies, work backwards, or utilize other information in the stem to arrive at a correct answer without “true” knowledge of the underlying content being assessed. Both these potential threats to validity are of concern to the measures development effort described here, because test authors wanted to measure knowledge used in teaching rather than basic skills or low-level cognitive activities, and because the use of test-taking strategies is a serious threat to the interpretation of scores.

To assess these interpretations of scores, we engaged a set of teachers, non-teachers, and mathematicians in cognitive interviews around a subset of items sampled from our larger pool. Cognitive interviews are advocated by survey research methodologists such as Sudman, Bradburn, & Schwarz (1996) as a method for ascertaining how respondents interpret and respond to survey items. Though much of this work, including recent work in the field of education (Camburn & Barnes 2004; Desimone & Le Floch, 2004) has focused on survey rather than test-like items, the logic is similar.

Below we briefly describe the methods we employed in collecting and analyzing data. We then analyze interview data using two different methods. We address the elemental validity by examining whether, for each item, a teacher’s thinking was consistent with the multiple-choice answer she selected, and in particular whether correct thinking was accompanied by a correct answer (and vice versa). This is a first step in establishing whether teachers’ scores are reflective of their level of mathematical knowledge for teaching, and whether multiple choice is a viable format for measuring such knowledge.

Second, we also examined responses to determine whether teachers drew on the type of knowledge that items were intended to tap. This allows us to examine the elemental assumption—in particular, whether test-taking or other types of thinking or knowledge were prevalent. It also helps us address the structural assumption by determining whether teachers draw on different types of teaching-specific knowledge to answer these items.

For this analysis, we focus mainly on distinctions between CK and KCS items. Although the structural assumption describes further divisions within CK (common versus specialized CK), reasoning in both cases looks purely mathematical. As a result, we cannot use the analysis of respondents’ thinking to make a determination about the existence of sub-domains. We *are* able to examine, from this interview data, whether non-teachers’ thought process indicates this subdivision, and do so briefly at the end of the results section. We also take the common versus specialized distinction up at length in the Schilling (this issue) paper.

## METHODS

### Interview Sample and Protocol

Twenty-seven teachers, 18 non-teachers in similar professions (nursing, social work), and 18 professional mathematicians participated in this study. We included non-teachers and professional mathematicians in the sample specifically to address the question of dimensionality; if there is knowledge for teaching separate from content knowledge itself, teachers, and not non-teaching professionals, should be the primary users of this knowledge. Further details on response rate and sample characteristics are available in Hill, Dean, & Goffney (2005).

Respondents were selected based on their mathematical knowledge<sup>1</sup>, took the multiple-choice items on their own, then engaged during the 50–75-min interview in what Sudman, Bradburn, & Schwarz (1996) would term a “retrospective think-aloud,” in which they reported on how they determined their answers. Following advice from Ericsson and Simon (1984), we provided respondents with instructions to track their thinking process during their independent work. We chose retrospective, as opposed to concurrent (reporting thinking at the same time they were solving items), think-alouds in order to maximize the number of items covered in the interview, and because evidence suggests that information that is heeded during the performance of a task (i.e., the solution process) can be reliably retrieved (Ericsson & Simon, 1984). One drawback of retrospective interviews, however, is that the data generated may not represent the full array of problem-solving strategies by respondents—particularly those initially used and then discarded in favor of other methods or answers.

We selected items for interviews based on our interest in understanding how the item “worked,” variation in IRT results, and variation in construct. In the area of content knowledge, we selected 18 number/operations items lodged beneath eight stems and five geometry items lodged beneath one stem.<sup>2</sup> These content knowledge items varied in the extent we believed they would require the user to invoke common and specialized mathematical knowledge. In the area of KCS, we selected one open-ended and five multiple-choice items designed to explore teachers’ knowledge of common student errors in number and operations.

---

<sup>1</sup>In order to ensure variation in knowledge level, we selected teachers and non-teachers to participate in cognitive interviews based on scores on an initial survey, conducted through the mail with a larger sample of each group (50 teachers, 71 non-teachers). Raw scores were divided into quartiles from which we selected 10 individuals earning the lowest raw score and 10 individuals earning the highest raw score.

<sup>2</sup>A stem is a problem situation, and an item is a selection made by a teacher. Where some stems have only one item, others require teachers to make more than one selection (in psychometric terms, they are “testlets”).

## Analyses

### *Consistency*

As stated in the first inference above (1A), teachers' reasoning for a particular item should be consistent with the multiple-choice answer they selected. To elaborate, consistency occurs when interviewees who know the mathematics or information about students invoked by the item choose the correct answer, and those who do not know choose an incorrect answer. Inconsistency occurs when an individual gets the right answer on the basis of faulty thinking (e.g., a respondent who answers correctly that a cube never has eight edges, then elaborates that this is because all cubes have six edges [in fact, cubes have 12 edges]) and also when an individual with a sound understanding of the underlying problem answers incorrectly (e.g., when an individual has correctly diagnosed a student's difficulty as a misunderstanding about place value, but chooses the answer "I'm not sure" because one can really never know what a hypothetical student is thinking). Codes for inconsistency were also assigned to interview passages where respondents answered correctly by guessing, mistakenly circled the wrong answer choice, misread the item, and those passages where the respondent originally gave the wrong answer on the written pre-interview survey, then self-corrected, demonstrating sound mathematical understanding, during the course of the interview. We reached 89% interrater agreement in trials before completing coding individually. There were no subsequent checks for interrater reliability.

### *Thinking*

Our structural assumption and its inferences (2A–D) asserts that scores reflect specific types of mathematical knowledge for teaching. Analyzing the thinking behind answers can provide one insight into whether this is the case; comparing across the groups in this study allows us to see whether any knowledge is unique to the work of teaching. Finally, learning more about the characteristics of respondents' thinking processes can help us answer critiques of the multiple-choice format, particularly allegations that these items tap low-level knowledge and skill and may be prone to test-taking strategies and/or guessing.

After we received and cleaned our data, we constructed a set of codes that reflect patterns in respondents' thought processes. Six of these codes are mathematical in nature:

- **Mathematical justification:** This code reflects correct mathematical reasoning about an item, usually with recourse to definitions, definitive examples or counter-examples, or the consideration of unusual cases (e.g., zero or negative numbers).
- **Memorized rules or algorithm:** Respondent refers to memorized rule or algorithm for primary justification for answer.

- **Definitions:** Correct uses of definitions are included in mathematical justification, above. This code captures inaccurate or incomplete uses of definitions (e.g., “I can’t remember the definition exactly”).
- **Examples/counterexamples or pictures:** This code reflects a respondent choosing numbers, cases, figures or shapes (including drawing figures and shapes) to assist in reasoning through a problem. If an example is used as proof (e.g., by counterexample, by unusual case) it is coded as *mathematical justification* above. This code mainly comprises examples that do not rise to level of justification, or examples used to support incorrect answers.
- **Other mathematical reasoning:** Mathematical thinking that does not fall into the category of justification, definitions, examples or pictures. The respondent may use mathematical deduction, inference, or other type of thinking to support her answer (e.g., “looking at these numbers listed in the problem, (c) must be true.”) The response can either reflect correct thinking or incorrect thinking; the main feature is that there is something mathematical about the thought process, rather than a non-mathematical process such as test-taking, guessing, etc.

One code referred to knowledge of students:

- **Knowledge of students and content:** Respondent invokes knowledge of students as partial or complete explanation for selecting their answer (e.g., “my students do this all the time.”)

And three codes referred to non-mathematical and non-student methods for answering problems:

- **Guessing:** Respondent reports guessing.
- **Test-taking skills:** Respondent uses information in the stem, matches response choices to information in the stem, or works to eliminate answers as method for solving problem (e.g., “I knew it wasn’t (a), thus (d) ‘all of the above’ couldn’t be true.”).
- **Other:** Other non-mathematical thinking.

When no reasoning was apparent from the transcript, a code of “not present” was applied. Two or more codes could be applied to a single response, if necessary; in practice, however, we tried to keep the number of double-coded responses to a minimum. All responses, whether correct or incorrect, were coded.

This coding scheme helps provide information on both the elemental and structural assumptions. In the former case, we can estimate the amount of test-taking, guessing, and other non-mathematical thinking that occurred in response

to our questions. In the latter case, we can assess whether respondents appeared to draw on different knowledge bases when answering CK and KCS items. We can also determine whether non-teachers also “have” KCS and, to a more limited degree, specialized content knowledge.

## RESULTS

### Consistency of Thinking and Answers

As Table 1 shows, average inconsistency rates are quite low for the CK items. Among teachers, the primary target of these items, only 1 in 20 answers to items failed to capture respondent’s underlying knowledge. These rates of inconsistency range from zero for several items to 19% for a particularly problematic item asking teachers to represent  $1\frac{1}{2} \times \frac{2}{3}$  using a diagram (item 2 in the Appendix); however, most other items had low inconsistency rates. Inconsistencies in the KCS domain were, on average, higher for teachers but comparable for the other groups. The high teacher inconsistency rate reported for KCS is largely due to problems with one item, which showed a 40% inconsistency rate.

An analysis of inconsistent responses shows that 44% percent result from an answer changed during the interview; in some cases, respondents reported they misread the item in their initial work while in others, respondents changed their answer in response to probing by the interviewer. Both types of inconsistency seem endemic to either the process of taking any written survey/questionnaire

TABLE 1  
Consistency Rates

		<i>Inconsistent Responses</i>
Overall rates	Teachers	8.1%
	Mathematicians	5.2%
	Non-teachers	7.9%
	Average	7.3%
CK items only	Teachers	6.3%
	Mathematicians	5.0%
	Non-teachers	8.1%
	Average	6.6%
KCS items only	Teachers	15.6%
	Mathematicians	6.1%
	Non-teachers	6.0%
	Average	10.0%

*Note.* The unit of analysis is respondent comments about one item. Percentages show the proportion of units falling into each category.



without interaction or feedback, or the social process that occurs during interviews. Inconsistencies also occurred because some items simply failed to accurately capture the mathematical knowledge held by individuals. The cube problem described above is one example. In another example, some respondents incorrectly agreed with statements such as “you can’t subtract a number from zero,” but in interviews, revealed that they had consciously limited the domain to whole numbers, for which this statement is true. The stem urged respondents to consider whether this statement was true, “not actually true, or [...] not true for all numbers.” Yet many respondents answered incorrectly while displaying what we could consider to be relatively nuanced understanding of mathematical domains.

A third category of inconsistent answers can be attributed to more traditional survey design difficulties. In most cases, we believe that the main issue is the language of the item itself. Questions that asked respondents to identify the *least* likely student errors were often mistakenly answered as if they asked for the *most* likely student error. In several items, we asked respondents to infer what students or teachers knew in solving mathematics problems; some respondents balked, epistemologically, at making such an inference.

## Respondents’ Thinking

### *Evidence for the Elemental Assumption*

Tables 2 and 3 show the types of thinking that occurred in answers to the CK and KCS items, respectively. Results look promising for the CK items, where the most commonly used processes were mathematical in nature: justification, other mathematical thinking, and the recall of memorized rules or procedures. To answer our items, in other words, respondents relied on mathematical knowledge or thinking. Guessing was virtually non-existent, at least in the face-to-face interviews<sup>3</sup>. Test-taking strategies were also relatively minimal. These findings suggest that the CK items do capture some aspects of respondents’ mathematical thought process.

Results in Table 2 also help provide information on whether our multiple-choice items lead respondents to try to recall rules and procedures, or whether they ask respondents to use more sophisticated forms of mathematical reasoning. For both teachers and non-teachers, reliance on memorized rules was a not uncommon (15%) strategy, but also failed to make up the majority of explanations for answers. It is worth noting, also, that the majority of recall/procedures codes were applied to three stems: one invoking the formula for area of a

<sup>3</sup>One reason might be that respondents are less likely to admit guessing in face to face interviews. However, we have also seen no psychometric results that suggest guessing is a prevalent problem for most items, either.

TABLE 2  
Explanations for Content Knowledge Items

	<i>Mathematical Justification</i>	<i>Mathematical Reasoning</i>	<i>Memorized Rules</i>	<i>Definition</i>	<i>Examples and Counter examples</i>	<i>Admitted Guessing</i>	<i>Test-Taking</i>	<i>Other</i>	<i>Not Present</i>
Teachers	26.0%	35.2%	15.3%	0.8%	2.3%	1.8%	3.8%	0.4%	11.7%
Mathematicians	39.7%	36.8%	11.4%	0.4%	0.0%	0.0%	2.7%	0.0%	7.9%
Non-teachers	25.0%	36.5%	15.4%	1.6%	2.5%	1.8%	2.5%	0.2%	9.2%
Average	29.5%	36.0%	14.2%	1.0%	2.0%	1.2%	3.1%	0.2%	9.9%

*Note.* The unit of analysis is respondent comments about one item. Percentages show the proportion of units falling into each category.

TABLE 3  
Explanations for KCS Items

	<i>Knowledge of students</i>	<i>Mathematical reasoning</i>	<i>Admitted guess</i>	<i>Test-taking skills</i>	<i>Other</i>	<i>Not present</i>
Teachers	40.5%	40.0%	0.0%	16.1%	0.0%	1.5%
Mathematicians	1.8%	50.9%	0.9%	30.7%	0.0%	4.4%
Non-teachers	15.5%	58.1%	3.9%	21.7%	0.8%	0.0%
Average	23.4%	48.0%	1.3%	21.4%	0.2%	1.8%

*Note.* The unit of analysis is respondent comments about one item. Percentages show the proportion of units falling into each category.

circle, and two showing standard and non-standard but mathematically correct algorithms for operations with rational numbers. In one of the latter cases, recall of the standard algorithm was a common explanation for the *wrong* answer; as one respondent said, “[I] knew that [the non-standard] method was wrong even though he happened to get the right answer for this problem . . . . [respondent recites standard algorithm]. I remember that rule. So that’s what I went by.” [Loepp, 117–120]<sup>4</sup> Although this non-standard method was correct, this respondent relied only on the memorized rule and thus answered the item incorrectly.

Reasons for the KCS responses, however, were more troubling. Table 3 shows that while 40% of teachers did express familiarity or knowledge of the student error or strategy we sought to assess, an equal number used other mathematical thinking to arrive at their answer. Often, these two codes were used in combination to describe respondents’ thought process. To some degree, this makes sense; an inspection of Appendix items #3 and #4, for instance, suggests that some amount of mathematical thinking is necessary to solve the item correctly—if only, for instance, being able to put decimals in order accurately. However, a large number of correct responses to these items by non-teachers and mathematics, who relied *mainly* on mathematical thinking in formulating their answers, suggests that such problems can be answered with no knowledge of students present. In addition, the use of test-taking strategies occurred at a fairly high level with KCS items—at a rate of over 16% for teachers. This constitutes a threat to the validity of the items.

*Evidence for the Structural Assumption*

The results above suggest there is evidence for the multidimensionality we hypothesized under our validity argument. Table 3 shows teachers *did* refer to their experiences with students in answering KCS items, and in many cases, those

<sup>4</sup>Names are pseudonyms. Numbers refer to line numbers in NUDIST database.

experiences appear to enable the production of correct answers. For instance, one problem asked teachers to reflect on why a student might not be able to solve a problem such as  $8+4= \_+5$ ; research in student cognition suggests students interpret the equals sign to mean “compute now,” making the most common error on this problem an answer of 12 (Carpenter, Franke, & Levi, 2003). In discussing this item, a teacher commented:

I’ve seen this over and over as a 3rd and 5th grade teacher. Eight plus four equals something plus five. And, you know, as many times as you explained that this is a balance and that equals sign is the middle of the balance beam and both sides—it’s straight because they’re both perfectly equal. . . . And how many kids will write twelve there? Because they’re just looking at eight and four *equals something*. They’re not looking at this equation balances with that equation. (Thorn 237–241)

Teachers like Thorn are articulate about the mathematical problems their students face; non-teachers and mathematicians, by contrast, reasoned their way to the correct answer through other, often more circuitous means—only 2% of responses from mathematicians and 16% of responses from non-teachers invoked knowledge of students in their answers. Similarly, mathematicians were markedly less likely to get these items correct than content knowledge items (Sleep, Delaney, Dean, Ball, Hill, & Bass, 2005).

In related work (Sleep, Delaney, Dean, Ball, Hill, & Bass, 2005), we have also explored the extent to which specialized knowledge exists by examining whether mathematicians have difficulty answering any CK items. Most mathematicians answered most CK items correctly. However, there were a small number of items on which some mathematicians struggled. Our analyses of these items suggest some mathematicians lack flexibility with non-standard approaches. When shown non-standard multi-digit multiplication methods, for instance, some respondents could recognize that the student had arrived at the correct answer, but could not decipher the method. In a few extreme cases, we found that mathematicians indicated that an alternative approach to solving a problem was incorrect simply because it was non-standard. Mathematicians’ responses also reflected, in some cases, their “compressed” knowledge of the subject: for example, mathematicians freely interchanged equivalent representations or concepts, unaware of how such a seemingly simple substitution might impact the teaching of mathematics to children. On other items, mathematicians didn’t recognize that a concept or logical step might not be obvious to non-experts. We hypothesize that these two aspects of reasoning—flexibility and decompression—are aspects of specialized knowledge for teaching, and may be possessed by teachers expert in teaching mathematical content to students. This finding does not align precisely with our initial definition of specialized content knowledge, a topic we take up again in the Schilling paper (this issue) and in our concluding paper.

## CONCLUSION

Our elemental assumption, that these measures represent teachers' mathematical knowledge for teaching, is supported by evidence for the content knowledge items but not their companions, the knowledge of content and students items. Results from our examination of consistency and reasons suggest that it is mathematical processes, by far, that underlie answers to the CK items. Test-taking and guessing occurred at relatively low rates, and inconsistencies between individuals' thinking and answers were within what we would probably consider to be normal bounds. Results from the KCS items were more mixed, however. While some teachers did use what we coded as KCS in their answers to these items, other teachers, non-teachers, and mathematicians also relied on mathematical reasoning in generating their answers.

The evidence about KCS, however, does suggest that the answer to the structural assumption is a modified yes. As we shall see in our psychometric analysis, it does appear that there is not one unitary "mathematical knowledge for teaching" that underlies teachers' answers to these items, but at least two: CK and KCS. This suggests that, in practice, the construction of two separate scores to represent individuals' capacities is warranted. However, the multidimensionality found *within* the KCS items suggests that constructing a scale score for this domain will not be entirely straightforward, because KCS must be separated from the mathematical reasoning ability tapped by such items.

This analysis has allowed us to rule out common problems and critiques of multiple-choice items. It does not appear, for instance, that CK items draw mainly on respondents' ability to recall rules or algorithms. Instead, mathematical reasoning—and in some cases, justification—are required to come to an appropriate answer. Test-taking is also not a skill that is widely used in the CK domain, and if it is used, it does not bias the respondent toward a correct answer.

Finally, this study has provided a rich data source for learning about and improving our items. Some of the improvements have already occurred at the item level: for instance, not asking respondents to infer what a hypothetical teacher or student knew or was thinking. Others will guide future item writing efforts. "Choose the least likely," for instance, is no longer used as a format for items. And insights from the KCS item responses, in particular, will allow us to rethink and possibly entirely recast our measurement efforts in this domain.