Concept Inventories: Tools For Uncovering STEM Students' Misconceptions

Jim Richardson, University of Alabama, Tuscaloosa, AL

Concept inventories are receiving increased interest from STEM faculty. What are concept inventories, why the interest, and what do I need to know about concept inventories? This chapter answers these questions in the following order. In the first section, you will read a brief history of STEM concept inventory development, which should answer the question, "Why the interest in concept inventories?" In the next two sections, you will read first a short discussion on the theory of assessment as it applies to concept inventories and then a description of how to construct a concept inventory. Together, these two sections should answer the question, "What is a concept inventory?" And finally, you will read how others have used concept inventories and related tools to improve their teaching effectiveness.

An Abridged History of STEM Concept Inventories

The story of STEM concept inventories begins with the Force Concept Inventory (FCI) (1,2), which was developed in the late 1980s and early 1990s by David Hestenes and several of his graduate students at Arizona State University. Hestenes gave the following account about the origin of the FCI (3).

> *"One of my graduate students approached me with a set of questions related to Newtonian mechanics and asked me to give these questions to my Physics I students. I looked at the questions and told him, 'These questions are trivial. This is a waste of time.' He finally succeeded in convincing me to give the questions to my students. When I looked at the student responses, I was astonished. Large numbers of my students had failed to answer the questions correctly."*

Hestenes and his graduate students administered a Mechanics Diagnostic Test to students in Introductory Physics courses (4,5). Initial versions were not multiple choice, but required students to write out the answers. Common but wrong answers were used to construct the multiple-choice wrong answers (termed "distractors" by Hestenes) for later versions of the Mechanics Diagnostic Test and eventually the FCI.

While some may argue with the validity of the FCI (Does it really measure what it purports to measure?), everyone agrees that the FCI was the impetus for substantial innovation in physics instruction. Key in this process was the "conversion experience" of a highly regarded physicist at Harvard, Eric Mazur. The story is best told in his own words (6).

> *"For the past 8 years, I have been teaching an Introductory Physics course for engineering and science concentrations at Harvard University. I used to teach a fairly traditional course in an equally traditional lecture-type of presentation, enlivened by classroom demonstrations. I was generally satisfied with my teaching during these years—my students did well on what I considered pretty difficult problems, and the feedback I received from them was positive.*
>
> *About a year ago, however, I came across a series of articles by David Hestenes of Arizona State University, which completely and permanently changed my views on teaching. In these articles, Hestenes shows that students enter their first physics course possessing strong beliefs and intuitions about common physical phenomena. These notions are derived from personal experiences and color stu-*

dents' interpretations of material presented in the introductory course. Instruction does very little to change these 'common-sense' beliefs. Hestenes provides many examples in which students are asked to compare the forces of different objects on one another. When asked, for instance, to compare the forces in a collision between a heavy truck and a light car, a large fraction of the class firmly believes the heavy truck exerts a larger force on the light car than vice versa. My first reaction was, 'Not my students…!' I was intrigued, however, and [tested my own students]. The results of the test were undeniably eye-opening.

I spent many, many hours discussing the results of this test with my students one-on-one. The old feeling of satisfaction turned more and more into a feeling of sadness and frustration. How could these undoubtedly bright students, capable of solving complicated problems, fail on these ostensibly 'simple' questions? Slowly the problem revealed itself: many students concentrate on learning 'recipes,' or 'problem-solving strategies' as they are called in textbooks, without bothering to be attentive to the underlying concepts. Many pieces of the puzzle suddenly fell into place: The continuing requests by students to do more and more problems and less and less lecturing—doesn't the traditional lecture overemphasize problem-solving over conceptual understanding?

Just a year ago, I was entirely oblivious to this problem. I now wonder how I could be fooled into thinking I did a credible job teaching Introductory Physics. While several leading physicists have written on this problem, I believe most instructors are still unaware of it. A first step in remedying this situation is to expose the problem in one's own class. The key, I believe, is to ask simple questions that focus on single concepts. The result is guaranteed to be an eye-opener even for seasoned teachers."

Eric Mazur went on to develop a method for teaching physics concepts more effectively in large lecture classes, described in his book *Peer Instruction* (7). Another physics professor, Richard Hake, compiled FCI results from 6,000 students. The students took the FCI at the beginning and again at the end of the first physics course (8). He showed that students in classes using active engagement techniques had larger gains on the FCI than students in traditional lecture-style classes.

In the late 1990s, others began emulating Hestenes and developed concept inventories for electromagnetic waves (9,10), signals and systems (11), strength of materials (12–14), thermodynamics (15), materials science (16,17), statistics (18), heat transfer (19), fluid mechanics (20), chemistry (21), electromagnetics (21), and circuits (21).

A team of researchers at Colorado School of Mines is currently developing a concept inventory for thermal and transport science (22,23). This team includes assessment experts and psychologists as well as engineering professors. One of the unique aspects of this concept inventory is that the developers identified a fundamental and pervasive misconception before beginning work on the inventory. Ron Miller, a project principal investigator, had an "aha" experience while listening to a talk by Michelene Chi, a noted cognitive psychologist at the University of Pittsburgh. Micki Chi (as she's known) was describing how students have difficulty understanding "emergent" processes in science, frequently confusing them with "causal" processes. She explains the difference between causal and emergent processes below (24).

"As it turns out, many science processes that students fail to understand (such as heat transfer, natural selection) tend to be an emergent kind of process and not the causal kind. In causal processes, the process itself can be explained by identifiable actions, with temporal and spatial contiguity, and with an identifiable beginning and ending. For example, a baseball game is a causal event. There is a beginning and an ending, with temporal and spatial sequence. A team's wins or losses can be attributed to certain causal actions, such as how well the pitcher pitches.

An emergent kind of process, on the other hand, does not have any of these properties. Instead, an emergent process is caused by the collective aggregation of multiple, independent causal events. For example, birds flocking in a V-formation may look like a causal event (in that the lead bird shows the others where to fly). It is actually an emergent process, in that each bird tends to fly in a location that has the least resistance and drag. If each bird independently follows such a simple rule, it will result in the flock flying in a V-formation."

The thermal and transport science concept inventory currently under development at the Colorado School of Mines is one of the best-planned and most thoroughly documented engineering concept inventories to date. The project website, http://www.mines.edu/research/cee/Misconceptions.html, provides links to the project proposal (funded under the National Science Foundation's CCLI Assessment of Student Achievement program) and other useful information.

Elements of a Concept Inventory

The book *Knowing What Students Know* (25) by the National Research Council describes three interconnected components of assessment: 1) a cognitive model of student learning, 2) observations of student responses, and 3) interpretation of those responses in light of the cognitive model. These components of assessment (represented in Figure 1) are not a recipe for constructing a concept inventory, but rather serve as criteria against which concept inventory developers can check their progress toward developing a quality assessment.

The cognitive model for a concept inventory is much more than a list of topics or concepts. The cognitive model must incorporate knowledge of the student learning process. It is the cognitive model that allows one to meaningfully interpret the results from the observations. Interpreting results from a concept inventory that lacks a good cognitive model is analogous to a doctor interpreting a set of X-rays without a good understanding of human physiology.

The bottom-left corner of the assessment triangle in Figure 1 is observations. Observations of student performance in a concept inventory are the concept questions. Desirable qualities of assessment observations are described in the following excerpt from *Knowing What Students Know* (25).

> *"The tasks to which students are asked to respond on an assessment are not arbitrary. They must be carefully designed to provide evidence that is linked to the cognitive model of learning. …In assessment, one has the opportunity to structure some small corner of the world to make observations. The assessment designer can use this capability to maximize the value of the data collected, as seen through the lens of the underlying beliefs about how students learn in the domain."*

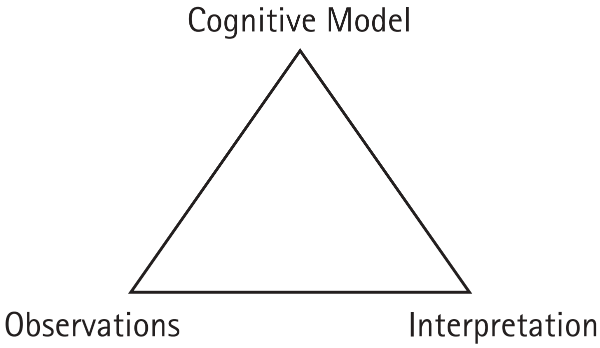

Figure 1. Assessment triangle.

The third and final corner of the assessment triangle is interpretation. Interpretation of the concept inventory is performed at two levels. First, the developers of the concept inventory perform statistical analyses on a large number of student responses to establish the validity (Does the inventory measure what it says it does?) and the reliability (Are the results repeatable?). The next paragraph describes an example of the use of one statistical technique, factor analysis, to help determine the validity of a concept inventory.

Factor analysis is a statistical technique to detect structure in the relationships between variables. Imagine that a concept inventory was intended to probe student understanding on five distinct concepts (concepts 1 through 5). And for each concept, four questions were written to probe student understanding on that concept only (questions 1.1, 1.2, 1.3, and 1.4 for concept 1, etc.). The order of the questions is randomized, and the concept inventory is given to a large number of students. A factor analysis is performed on the student responses and returns five factors (factors A, B, C, D, and E). For a perfect concept inventory administered in a perfect world, factor A would be composed of questions 3.1, 3.2, 3.3, and 3.4, and factor B would be composed of questions 5.1, 5.2, 5.3, and 5.4, and so on, with each factor mapping exclusively to those questions intended to probe a single concept.

The results of a concept inventory are also interpreted by the instructor. For example, Hestenes and Halloun relate the following experience in their article, "Interpreting the FCI" (2).

"However, as a placement exam for accelerated or advanced courses, the FCI may be very useful. We expect students testing below the Newtonian entry threshold (60%) to have difficulty with such courses. Our limited data support this expectation. We have used the FCI several times as a pretest for an Honors section of University Physics, wherein all the students have exceptional academic records. There is a wide distribution of their FCI scores, however. Without exception, the students testing below the 60% threshold had difficulty with the course, while those above did not.

Process for Constructing a Concept Inventory

The steps for constructing a concept inventory are listed below.

1. Determine the concepts to be included in the inventory.
2. Study and articulate the student learning process regarding those concepts.
3. Construct several multiple-choice questions for each concept.
4. Administer the beta version of the inventory to as many students as possible. Perform statistical analyses on the results to establish validity, reliability, and fairness.
5. Revise the inventory to improve readability, validity, reliability, and fairness.

Construction of a concept inventory will involve iteration on the steps outlined above. For example, the list of concepts to be examined by the inventory may change after gaining insight to the student learning process. And student answers to concept questions without multiple-choice answers will likely provide insight regarding how students learn the concepts. Each of the five steps is expanded in a paragraph below.

1. Determine the concepts to be included. Because most concept inventories are designed to be administered in about 30 minutes, the number of concepts examined by the inventory is limited. The developers of the thermal and transport science concept inventory followed a process called the Delphi method, in which they asked experienced teachers to list important concepts with which students frequently struggle. "The Delphi method is based on a structured process for collecting and distilling knowledge from a group of experts by means of a series of questionnaires interspersed with controlled opinion feedback" (22). Selecting the content of the concept inventory based on expert opinion is important to help establish the concept inventory's validity.
2. Articulate the student learning process. This is a difficult step, one that requires extensive observations of the student thought process and thoughtful synthesis of those observations into a cognitive model. The developers of the thermal and transport science concept inventory incorporated cognitive psychologist Micki Chi's cognitive model of emergent versus causal processes. In work on a strength of materials concept inventory, Paul Steif identified two underlying mental constructs (see Table 1) that are taken for granted by most experts, but that are at the root of many student learning difficulties in the subject (14). Concept inventory developers have used several methods for probing student thought processes, including personal interviews, focus groups, and written open-ended questions.
3. Construct multiple-choice questions. This is a two-step process. In the first step, questions without multiple-

Table 1. Strength of Materials "Root" Constructs

Concept	Description
Internal versus external force	Internal rather than external forces determine stress and strain.
Displacement versus deformation	Relative rather than absolute displacements determine deformation and strain.

choice answers are generated following Eric Mazur's advice, "The key, I believe, is to ask simple questions that focus on single concepts" (6). This is harder than it sounds, but is very important because questions based on more than one concept will make it difficult or impossible to determine which concept the student did not understand when he or she answered the question. The second step of writing multiple-choice questions is to generate the multiple-choice answers. Many concept inventory developers use the student responses from the previous step to construct false answers based on common student misconceptions (1,21,23). Student interviews and focus groups are also frequently used to identify misconceptions and to ensure that the questions are worded clearly.

4. Administer the beta version of the concept inventory and analyze the results statistically to establish validity, reliability, and fairness. Validity addresses the question, "Does the concept inventory measure what it purports to measure?" Validity is established in several ways, including the following: incorporating the knowledge of experts in the field during the development process, checking that questions intended to probe the same concept illicit similar student responses, and comparing students' concept inventory scores with observations of their ability to use concepts on traditional exam-style problems.

   Reliability addresses the question, "Does the concept inventory yield similar scores when repeated by the same student?" Reliability can be established using statistical procedures. Fairness addresses the question, "Does the concept inventory produce comparably valid inferences between students of different gender or ethnic groups?" Because the number of engineering students from underrepresented groups may be too small at a single university to yield statistically significant results, beta-testing the concept inventory should be performed at several universities. Concept inventory development teams need to have experts in assessment and test construction involved at every stage of the project, especially the planning stages. See the project website for the thermal and transport science concept inventory for a good example of incorporating assessment expertise in all phases of development (http://www.mines.edu/research/cee/Misconceptions.html).

5. Revise the inventory to improve readability, validity, reliability, and fairness. Analysis of the results from the beta version will prompt changes in the inventory to improve the qualities listed in the sentence above. The revised concept inventory will need to be reanalyzed, which may prompt further changes. This is a continual process for large, widespread, high-stakes assessments such as the SAT and ACT.

Using Concept Inventories to Improve Teaching

The first step in advancing skills and abilities in an area is to recognize the need for improvement. This is the principal contribution of the FCI. Eric Mazur articulately expressed his experience of realizing that his students weren't learning what he thought they were learning. His frank description of his awakening, coupled with his stature in the physics research community, encouraged many other physics instructors to reexamine their teaching effectiveness. Eric Mazur went on to develop his own teaching techniques (7) and is currently working on a physics textbook.

Once you are aware that your students struggle to learn the concepts in your course at a "deep" level, you can use a concept inventory to help you modify your teaching. Assessment experts never base their evaluation on a single assessment, but rather triangulate between several independent assessments. You too should base your evaluation of your students' conceptual knowledge on several assessments (one of which can be a concept inventory). In the next paragraph, the strengths of a concept inventory as a teaching improvement tool are discussed, and in the following paragraph, other assessment techniques that can complement a concept inventory are briefly discussed.

A properly constructed concept inventory is based on a cognitive model of student conceptual understanding in an area of focus. The developers should communicate that cognitive model to a teacher interested in using his or her concept inventory. Also, the concept inventory has been extensively tested to establish that it is a valid and reliable measure of a student's conceptual knowledge, as articulated in the cognitive model. And finally, a properly constructed database of concept inventory results can help a teacher interpret the significance of his or her students' performance. The database needs to include results from thousands of students on dozens of campuses if it is meant to reflect a national average. A concept inventory administered over the web would facilitate the collection of information about student demographics, class size, type of school, etc., and would allow teachers to compare their students' performance against the performance of similar students.

Concept inventories provide a narrow view of students' conceptual knowledge, but one that is anchored to a coherent theory of student learning and can be interpreted in relation to thousands of other students' performances. Assessments complementary to a concept inventory can provide a broader and more personal (albeit a less transferable) view of students' conceptual knowledge. One assessment technique emulated by many was originally developed by Eric Mazur. In this technique, he stops his lecture, he asks his students to discuss and then answer a multiple-choice question focusing on a single concept, and then he polls his students and records the number of students selecting each answer. Another assessment technique, described in the excerpt below from *Knowing What Students Know* (25), focuses on how to ask students questions. The technique, suitable for smaller classes, illuminates the change in perspective required to effectively probe our students' conceptual knowledge.

> *"The teachers altered their practice [of asking questions] to give students extended time to think about the question posed, often asking them to discuss their ideas in pairs before calling for responses. The teachers did not label answers as right or wrong, but instead asked a student to explain his or her reasons for the answer offered. Others were then asked to say whether they agreed and why. Thus, questions opened up discussion that helped expose and explore students' assumptions and reasoning. At the same time, wrong answers became useful input, and the students realized that the teacher was interested in knowing what they thought, not in evaluating whether they were right or wrong.*

The quality of the questions posed to the students is particularly important to the success of both techniques described above.

Summary and Conclusions

Concept inventories bring to light the disparity between what we want our students to learn and what they really learn. Spurred by the success of the FCI in communicating the existence of this disparity to physics teachers, education researchers are developing concept inventories for many different subjects in science and engineering. Constructing a concept inventory is a multi-year process involving a team of professionals with expertise in both teaching the subject and in educational assessment. Teachers interested in improving the effectiveness of their teaching can stand on the shoulders of the concept inventory developers and catch a glimpse of the typical student's thought process. Teachers can also use a concept inventory as a reference point for gauging their progress in closing the gap between actual and desired learning.

As helpful as concept inventories can be, they cannot alone provide sufficient guidance for changing one's teaching to improve students' conceptual knowledge. Such guidance must be based on a more intimate knowledge of students' thought processes. The teacher must become a student of his or her pupils' learning process. Time consuming? Yes. Rewarding? Definitely.

REFERENCES

1. Hestenes, D., M. Wells, and G. Swackhamer. 1992. Force concept inventory. *The Physics Teacher.* 30: 141–158.

2. Hestenes, D., and I. Halloun. 1995. Interpreting the FCI. *The Physics Teacher.* 33: 502–506.

3. Personal communication, Tempe AZ, March 2000.

4. Halloun, I., and D. Hestenes. 1985. The initial knowledge state of college physics students. *Am J Phys* 53: 1043–1055.

5. Halloun, I., and D. Hestenes. 1985. Common sense concepts about motion. *Am J Phys* 53: 1056–1065.

6. Mazur, E. 1992. Qualitative versus quantitative thinking: are we teaching the right thing? *Optics and Photonics News* 3:38.

7. Mazur, E. 1997. *Peer Instruction.* Upper Saddle River, NJ: Prentice Hall.

8. Hake, R. R. 1998. Interactive-engagement vs traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Phys* 66: 64–74.

9. Roedel, R., S. El-Ghazaly, T. Rhoads, and E. El-Sharawy. 1998. *The Wave Concepts Inventory: An Assessment Tool for Courses in Electromagnetic Engineering.* In Proceedings of the 1998 FIE Conference. Champaign, IL: Stipes Publishing L.L.C.

10. Rhoads, T., and R. Roedel. 1999. The wave concept inventory: a cognitive instrument based on Bloom's taxonomy. In *Proceedings of the 1999 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

11. Wage, K., and J. Buck. 2001. Development of the signals and systems concept inventory (SSCI) Assessment Instrument. In *Proceedings of the 2001 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

12. Richardson, J., J. Morgan, and D. Evans. 2001. Development of an engineering strength of materials concept inventory assessment instrument. In *Proceedings of the 2001 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

13. Richardson, J., P. Steif, J. Morgan, and J. Dantzler. 2003. Development of a concept inventory for atrength of materials. In *Proceedings of the 2003 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

14. Issen, K., P. Steif, and J. Richardson. 2004. Probing conceptual knowledge of students in strength of materials. In *Proceedings of the 17th ASCE Engineering Mechanics Conference.* Newark, NJ: University of Delaware.

15. Midkiff, K., T. Litzinger, and D. Evans. 2001. Development of engineering thermodynamics concept inventory assessment instruments. In *Proceedings of the 2001 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

16. Krause, S., J. Decker, J. Niska, T. Alford, and R. Griffin. 2003. Identifying student misconceptions in introductory materials engineering classes. In *Proceedings of the 2003 ASEE Annual Conference.* Washington, DC: ASEE.

17. Krause, S., J. Decker, and R. Griffin. 2003. Using a materials concept inventory to assess conceptual gain in introductory materials engineering courses. In *Proceedings of the 2003 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

18. Stone, A., K. Allen, T. Rhoads, T. Murphy, R. Shehab, and C. Saha. 2003. The statistics concept inventory: a pilot study. In *Proceedings of the 2003 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

19. Jacobi, A., J. Martin, J. Mitchell, and T. Newell. 2003. A concept inventory for heat transfer. In *Proceedings of the 2003 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

20. Martin, J., J. Mitchell, and T. Newell. 2003. Development of a concept inventory for fluid mechanics. In *Proceedings of the 2003 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

21. Evans, D., G. Gray, S. Krause, J. Martin, C. Midkiff, B. Notaros, M. Pavelich, D. Rancour, T. Reed-Rhoads, P. Steif, R. Streveler, and K. Wage. 2003. Progress on concept inventory assessment tools. In *Proceedings of the 2003 FIE Conference.* Champaign, IL: Stipes Publishing L.L.C.

22. Streveler, R., B. Olds, R. Miller, and M. Nelson. 2003. Using a Delphi study to identify the most difficult concepts for students to master in thermal and transport science. *Proceedings of the 2003 ASEE Conference.* Washington, DC: ASEE.

23. Olds, B., R. Streveler, R. Miller, and M. Nelson. 2004. Preliminary results from the development of a concept inventory in thermal and transport science. In *Proceedings of the 2004 ASEE Conference.* Washington, DC: ASEE.

24. Michelene Chi's website: http://www.pitt.edu/~chi

25. National Research Council. 2001. *Knowing What Students Know.* Washington, DC: Board on Testing and Assessment, National Academies Press.