How to Create a Concept Inventory: The Thermal and Transport Concept Inventory

Mary A. Nelson
University of Colorado

Monica R. Geist
University of Northern Colorado

Ronald L. Miller, Ruth A. Streveler, Barbara M. Olds
Colorado School of Mines

Presented at the Annual Conference of the
American Educational Research Association

Chicago, Illinois
April 9-13, 2007

How to Create a Concept Inventory:
The Thermal and Transport Concept Inventory

Mary A. Nelson
University of Colorado

Monica R. Geist
University of Northern Colorado

Ronald L. Miller, Ruth A. Streveler, Barbara M. Olds
Colorado School of Mines

Abstract

This paper will describe the entire process of developing a robust concept inventory from identifying the key concepts in a field, through developing multiple-choice questions and believable distracters. Results of reliability and validity testing will be provided. The process will be described for one specific question from the inventory, from how the question derived from the topics generated by experts, through the corrections and changes that were made during alpha and beta testing. We will describe how such an inventory can be used identify students' misconceptions so that instructors can facilitate deep understanding of important concepts.

Introduction

Creation of a concept inventory (CI) that is both reliable and valid is a long, arduous process. This paper will describe how a three-part Transport and Thermal Sciences Concept Inventory (TTCI) was written, revised and tested, and will follow the creation of one of its questions.

Engineering faculty members often comment that even students who can correctly solve problems still hold fundamental misconceptions about the concepts that underlie the formulas and procedures that they used (Olds, Streveler, Miller and Nelson, 2004). Some, for example, believe that heat flows like a substance or that processes stop when they reach equilibrium. These faculty observations are supported by evidence in the literature that suggests science and engineering students do not conceptually understand many fundamental molecular-level and atomic-level phenomena such as light, heat, or electricity. The literature also suggests that the problem is more than simply one of confusion or misunderstanding, but instead involves fundamental misconceptions by students about differences in the way that molecular-scale processes differ from observable, macroscopic causal behavior we experience in our daily lives (Chinn & Brewster, 1993; Chi, Slotta and deLeeuw, 1994; Chi & Roscoe, 2002).

Though many concept inventories have summative assessment as their primary goal (Allen, et al., 2004; Martin et al., 2003), the TTCI was created specifically as a formative assessment tool. Formative assessment is an evaluation of students' work that can be used to "shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning" (Sadler, 1989, p. 120). This formative assessment normally takes place in the classroom in the course of instruction and is intended to be a tool to move learning forward (Shepard, 2001).

One assessment method that has taken root over the past two decades is the concept inventory. The movement began with David Hestenes (1992) and his colleagues who created the Force Concept Inventory (FCI), an instrument to measure students understanding of the fundamental Newtonian force concepts. This instrument revealed that even college graduates failed to understand some of the most fundamental ideas of physics (Hake, 1998). The FCI is commonly used today in universities across the nation, but is often used to measure learning gains, administered as a summative pre/post test of the student learning gains in a given course. In contrast, the TTCI was created as a formative assessment tool designed to identify common student misconceptions in heat transfer, fluid mechanics and thermodynamics, so that those misconceptions can be addressed in classroom instruction.

The Delphi Study

Before faculty can hope to develop curricular interventions to repair student misconceptions, they must first identify which concepts their students do not understand and which misconceptions are prevalent. This list is lengthy, and therefore our objective was to identify the most important and poorly understood concepts for the TTCI. To achieve this goal, we conducted a Delphi study with approximately 30 well-respected engineering faculty experts and prominent engineering textbook authors.

The Delphi method is a structured process that collects and distills knowledge from a group of experts (ref to qualifications) by means of a series of questionnaires interspersed with controlled opinion feedback. The method takes its name from the Oracle at Delphi, an ancient Greek soothsayer able to predict the future, and was originally developed at the RAND Corporation by Dalkey and Helmer (1963) as a tool for forecasting likely inventions, new technologies and the social and economic impact of technological change (Adler & Ziglio, 1996). The technique has obvious advantages over the use of a single expert. Though efficient, a single expert almost ensures bias due to the expert's experience and specialty. Using focus groups has the potential that less vocal and more insecure members acquiesce to the more prominent and forceful members of the group. The Delphi method allows each person's views to be heard, and provides the benefit of anonymous access to the views of other members of the group.

Delphi participants

We began by identifying practitioners willing to participate in each step of the Delphi survey. Our invitations to participate were sent to 35 experts across the country with an explanation of the Delphi method and the requirements of participation. Though some new faculty were invited, the average number of years of teaching experience was 23. Five of the participants had authored books in thermodynamics, fluid mechanics or heat transfer, and all participants were tenured or tenure track professors. The institutions represented were both research universities and undergraduate institutions. Though not every member of the panel participated in every round, for each round we had at least 26 responses on which to base our decisions.

Generative round

The Delphi method does not require a generative round, but we felt that it was essential to our project. At the start of the Delphi, the 31 experts who agreed to participate were asked to individually generate a list of what they considered the most important and least understood concepts in the study's topical areas. We subsequently coded their answers and identified those concepts cited by at least two participants. These concepts were then used as the basis of Rounds One through Three.

The table below provides two responses that were then coded as "heat versus temperature," and they were included in the Delphi as one of the 28 concepts to be rated.

Table 1: Responses of Delphi participants in Round one.

Confusion about the difference between heat and temperature. How can a process occur where heat is added but the temperature drops?
Temperature is a measure of energy. Example: Students often believe that if you add energy, heat for example, to any system, the temperature must go up.

Round one. The practitioners were given a list of the 28 concepts mentioned at least twice in the Generative Round. Each expert was asked to rate every concept on a scale of 1 to 10 for understanding (1 = students had no understanding of the concept; 10 = students have complete understanding) and 1 to 10 on importance (1 = the concept is not at all important; 10 = the concept is extremely important). The interquartile ranges (IQRs) and the medians were computed from the ratings of the Delphi members for each concept on both understanding and importance scales.

Round two. Practitioners were then given the IQRs and the medians which had been computed from Round One. In the second round, the Delphi participants were asked to rate each concept again, and were required to provide a written justification for any question where their rating was outside the IQR computed for the group in Round One.

Round three. In the third round, the IQRs and medians calculated in Round Two were provided to the practitioners, along with any justifications provided by participants for ratings outside the IQR. The justifications were given anonymously to prevent undue influence. The experts were then asked to rate each concept a third time, and a third set of IQRs and medians were computed. In this final round, practitioners were not required to justify ratings outside the interquartile ranges. Studies show and our project confirmed that there is little movement of median or interquartile ranges after three iterations of a Delphi study (Linstone & Turoff, 1975).

TABLE 2: Results of Thermal and Transport Concepts Delphi One Study
(I*talicized* concepts refer to low understanding/high importance rankings.)

CONCEPT	Understanding Data Median (interquartile range)			Importance Data Median (interquartile range)		
	Round 1	Round 2	Round 3	Round 1	Round 2	Round 3
1.Adiabatic vs. Isothermal Processes	7.5 (6-8)	8 (6-8)	8(6.75-8.25)	9 (8-10)	9 (9-10)	9 (9-10)
2.Bernoulli Equation	7 (4-8)	6 (5-7)	6 (5-7)	9 (7-10)	9 (8-9)	9 (8-9)
3. Compressible vs. Incompressible Flow	5 (3-7)	6 (4-6.5)	6 (5-7)	7.5 (6-8)	7 (7-8)	7.5 (7-8)
4. Conservation of Linear Momentum	5 (3-6)	5 (4-6)	5.5 (5-6)	9 (8-10)	9 (8-10)	9(8-9.25)
5. Differential vs. Integral Analysis	4.5 (3-6)	4 (3-5.25)	4 (4-5)	7 (6-9)	8 (6-8)	8 (7-9)
6. Dimensional Analysis	6 (4-7)	5.5 (4.25-7)	6 (5-6.25)	7 (5-7)	6 (5-8)	7 (5-8)
7. Entropy & 2nd Law of Thermodynamics	4 (2-6)	4 (3-5)	5 (3-5.25)	8 (7-9)	9 (8-9)	9 (8-10)
8. Extensive and Intensive Properties	8 (6-9)	8 (7-8)	8 (7-9)	7 (6-9)	8 (7-9)	8 (7-9)
9. First Law of Thermodynamics	8 (7-9)	8 (7-9)	8 (8-9)	10 (10-10)	10 (10-10)	10 (10-10)
10. Fluid vs. Flow Properties	7 (5-8)	6 (5-7)	6 (5-6)	7 (5-9)	7 (5-8)	7 (5-8)
11. Heat Transfer Modes	8 (6-9)	8 (6.25-8)	8 (7-9)	9 (8-10)	9 (9-10)	9 (9-10)
12. Heat vs. Energy	6 (5-8)	6 (5-7)	6.5 (5-7)	9 (8-10)	9 (8-10)	9 (8-10)
13. Heat vs. Temperature	6 (4-8)	6.5 (5-8)	7 (6-8)	9 (8-10)	10 (9-10)	10 (9-10)
14. Ideal Gas Law	8 (7-9)	8 (8-9)	8 (8-9)	9 (8-10)	9 (9-10)	9 (9-10)
15. Internal Energy vs. Enthalpy	6 (3-7)	5 (4-6)	6 (5-6.25)	8 (7-9)	9 (8-9)	9 (8-9)
16. No-slip Boundary Conditions	8 (6-9)	8 (7-9)	8 (8-9)	8 (7-9)	9 (8-9)	9 (8-9)
17. Nozzles and Diffusers	6 (5-8)	6 (6-7.5)	7 (6-7)	7 (5-9)	7 (6-8)	7 (6-8)
18. Pressure	8 (6-9)	8 (7-8)	8 (7.75-9)	9 (8-10)	10 (9-10)	10(9.75-10)
19. Reversible vs. Irreversible Processes	5 (4-7)	5 (4-6)	5 (5-6)	8 (8-9)	9 (8-9)	9 (8-9)
20. Spatial Gradient of a Function	4 (3-7)	5 (4-6)	5 (4-5)	7 (3-9)	7 (6-8)	7 (6-8)
21. Specific Heat Capacity	7 (6-8)	7 (6-7)	7 (6-8)	8 (7-10)	9 (8-9)	9 (8-9)
22. Steady-state vs. Equilibrium Process	5 (3-8)	5 (3-6)	5 (4-5.25)	8 (5-10)	9 (7-9)	9 (8-9)
23. Steady-state vs. Unsteady-state Process	8 (7-8)	8 (7-8)	8 (7-8)	9 (8-10)	9.5 (9-10)	9.5 (9-10)
24. System vs. Control Volume	7 (4-8)	6 (5-7)	6 (6-7)	8 (6-10)	9 (8-10)	9 (8.5-10)
25. Temperature Scales	7 (5-9)	8 (8-9)	9 (8-9)	8 (6-10)	9 (8-10)	9 (9-10)
26. Thermal Radiation	6 (4-8)	5 (5-6)	5 (5-6)	7 (5-9)	8 (6.75-8)	8 (7-8.25)
27. Thermodynamic Cycles	7 (5-8)	7 (6-7)	7 (7-8)	8 (8-10)	9 (8-10)	9 (8-9.25)
28. Viscous Momentum Flux	5 (3-7)	4 (3.75-5)	4 (3-4)	7.5 (6-9)	8 (7-8)	7 (7-8)

Understanding Scale	**Importance Scale**
0 = no one understands the concept	0 = no at all important to understand the concept
10 = everyone understands the concept	10 = extremely important to understand the concept

Phase Two: From Concepts to Open-Ended Questions

Once the Delphi study was completed, 12 topics fell into the region of interest: topics which were very important, but poorly understood. Scatter diagrams with “importance” on the horizontal axis and “understanding” on the vertical axis were created using the median results computed in Round Three. A “cluster” of concepts emerged as being the most important and least understood. Ten of those concepts represented by the points in the triangular cluster below became the target concepts that eventually were developed into TTCI questions. Two of the twelve were identified as mathematical concepts which we elected not to include in the inventory.

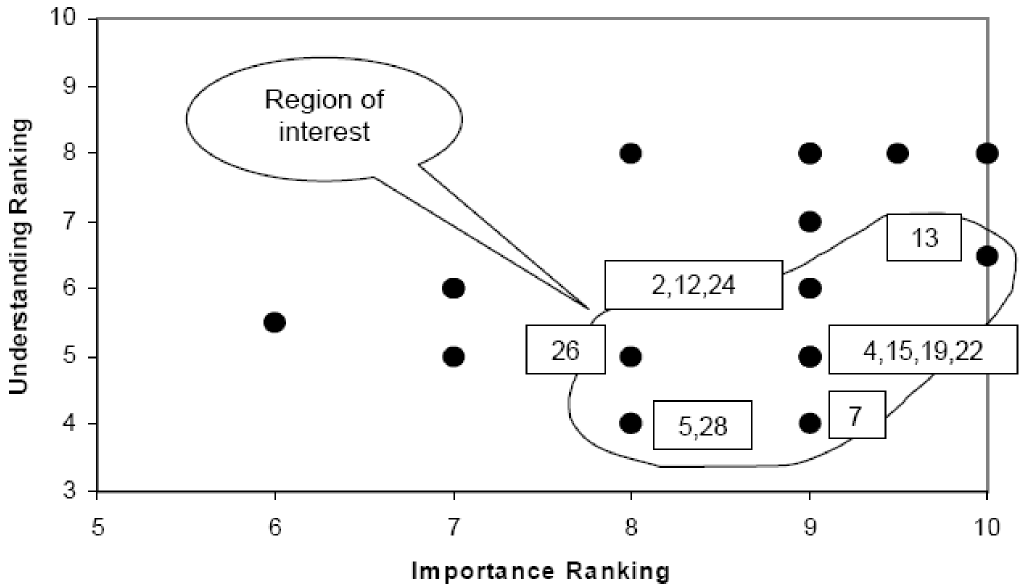


Figure 1 – Understanding versus Importance (Numbers refer to concept numbers listed in Table 1, where concepts in the triangle are highlighted)

Our next objective was to create appropriate concept questions in the ten areas of interest. This was achieved in several ways. Our content expert created CI questions himself, used outside sources such as engineering textbooks or previously developed CIs, or elicited contributions from engineering faculty around the country. This resulted in an initial list of at least one open-ended question on each of the targeted concepts.

Phase Three: Open-Ended Questions to Multiple-Choice Questions

Undergraduate chemical and mechanical engineering students were invited to answer the open-ended questions generated in Phase Two. The think aloud method was used (van Someren, Barnard, Sandberg, 1994) to discover students’ mental models of the target concepts. The think aloud sessions were taped, transcribed, and coded to help us create a thermal and transport concept inventory similar to the Physics Force Concept Inventory (Hestenes,D., Wells, M. & Swackhamer, G., 1992). We used these initial interviews to confirm that the concepts identified in the Delphi were indeed poorly understood by junior and senior engineering students, and to help in the writing of multiple choice questions which would be based on these open-ended questions.

During the coding of the think aloud data, some common misconceptions emerged (Olds, et al., 2004), which were then used to develop “believable” distractors for the TTCI that correspond to common student misconceptions. Other distractors were taken from a review of literature. We were also able to identify difficulties encountered with the confusing wording, labeling of diagrams or other issues which would confound our results. With this information our expert was able to write one multiple-choice question for each concept and we alpha tested all ten at the Colorado School of Mines. Later at least two more questions were developed for each concept, and think alouds were used again to uncover common student misconceptions to use as distractors and to refine wording and diagrams.

Development of One Question

One of the questions which was written to examine student misconceptions about heat versus temperature was one we labeled *Hotplate.* The quotes in Table 1 represent only two of 11 responses by Delphi members who cited "heat versus temperature" as an important and poorly understood concept. The Hotplate question was designed to address that concept and the original open-ended version of the questions is shown below in Figure 2.

> Two identical beakers contain equal masses of liquid at a temperature of 20^o. One beaker is filled with water and the other beaker is filled with ethanol (ethyl alcohol). The temperature of each liquid is increased from 20^oC to 40^oC using identical hotplates. It takes 2 minutes for the ethanol temperature to reach 40^oC and 3 minutes for the water to reach 40^oC. Once a liquid has reached 40^oC, its hot plate is turned off. To which liquid was more energy transferred during the heating process? Why did you answer the way you did (i.e. explain your reasoning)?

Six junior and senior engineering majors participated in the think aloud sessions on this question. All were chemical and mechanical engineering majors. The question was meant to probe students' understanding of the relationship between energy, heat, temperature and heat capacity. Two interviewers asked the students explain if they thought one liquid received more energy and then to defend why that was the case.

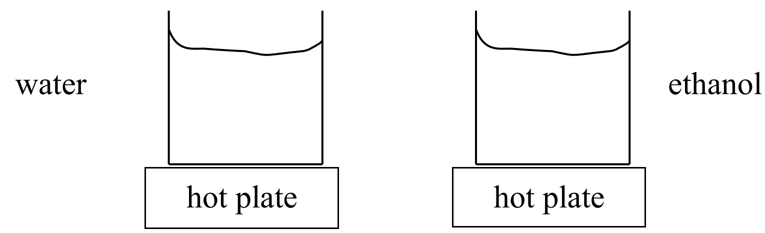
As hypothesized, all six students had notable difficult in explaining the concepts being addressed.. Many believed, for example, that crucial information had been withheld, saying "We can't tell because we don't know the [fluid] heat capacities." Others were able to deduce important information about the heat capacities, saying, "Just because ethanol gets hotter faster does not mean it gains more heat. Just that the ethanol has a lower heat capacity." These think alouds confirm what the research literature tells us about student misconceptions about heat versus temperature, and we used student misconceptions to write plausible distractors for the multiple choice question that resulted. The resulting question is Figure 2 below.

Figure 2: Initial Hotplate multiple choice question for the TTCI.

Hotplate TTCI Test Item

Two identical beakers contain equal masses of liquid at a temperature of 20 ^{O}C as shown below. One beaker is filled with water and the other beaker is filled with ethanol (ethyl alcohol). The temperature of each liquid is increased from 20 ^{O}C to 40 ^{O}C using identical hot plates.

It takes 2 minutes for the ethanol temperature to reach 40 ^{O}C and 3 minutes for the water temperature to reach 40 ^{O}C. Once a liquid had reached 40 ^{O}C, its hot plate is turned off.



To which liquid was more energy transferred during the heating process?

a. Water because more energy is transferred to the liquid that is heated longer.

b. Alcohol because more energy is transferred to the liquid that heats up faster (temperature rises faster).

c. Both liquids received the same amount of energy because they started at the same initial temperature and ended at the same final temperature.

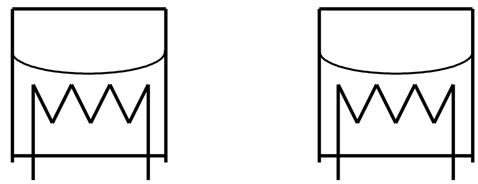
d. Can't determine from the information given because heat transfer coefficients for water and ethanol are needed.

e. Can't determine from the information given because heat capacities of water and ethanol are needed.

This multiple choice version of Hotplate was included in a set of concept questions administer to nearly 90 students at four engineering institutions. We found that nearly 50% of the students chose the correct answer "a". The remaining responses were distributed among c, d and e with only 2 students choosing b. This data tells us that "b" may not be a plausible distractor and could either be replaced or eliminated. At this juncture, we are waiting for more data before changing it.

Our next step was to elicit expert review of each question and we received valuable input on the Hotplate questions. Experts suggested that we address their concerns about evaporation losses due to open beakers, effect of fluid properties on actual heat transfer rates for the two fluids, and the wording of both the question and some of the distractors. Having addressed each of these concerns our new Hotplate question actually became an "Immersion Heater" question. The latest version is in Figure 3.

Figure 3: Revised Hotplate question on the TTCI:



Ignoring evaporation losses, to which liquid was more energy transferred during the heating process?

a. Water because it is heated longer

b. Alcohol because it heats up faster (temperature rises faster)

c. Both liquids received the same amount of energy because they started at the same initial temperature and ended at the same final temperature

d. Can't determine from the information given because heat transfer coefficients from the water and alcohol beaker surfaces are needed

e. Can't determine from the information given because heat capacities of water and ethanol are needed

Along the way several types of validity and reliability testing were conduced which contributed to decisions about each of the questions. Those efforts are described below.

Validity and Reliability

Preliminary validity and reliability checks have been done on the TTCI, based on approximately 100 students from 6 universities. As the instrument becomes more widely used, more extensive validity and reliability testing will be conducted. Although there are many types of validity to consider, at this stage we have focused on construct and content validity. After the instrument is more widely used, we will conduct analyses to test for external validity and bias. We presently have data from six universities, but taken separately there are not enough data from each university to determine if statistical differences exist among the student scores at these institutions.

Construct and content validity. (Are we measuring what we think we are measuring and have we covered the domain of what we want to measure?). We have worked to ensure construct and content validity in a variety of ways, consulting experts about the concepts chosen for the TTCI, alpha testing each question in open-ended form, and reviewing the 10 most popular textbooks in the areas covered by our instrument. In addition, we consulted students and experts about the wording of questions and the labeling of diagrams to ensure that the questions are phrased and illustrated accurately and clearly. We used student think-alouds to develop appropriate distracters for the multiple-choice questions.

Presently we do not have enough data to conduct invariance testing, which will indicate if items are biased. (Does the instrument measure different groups of the same ability differently?) However, we do have preliminary validity evidence which comes from an initial item analysis using Classical Test Theory (CTT). CTT measures item difficulty and item discrimination as described below, and proved valuable in deciding which questions were truly performing as desired.

Item difficulty. In CTT, item difficulty is defined as the proportion of examinees who answered the item correctly. If 90% of the student got an item correct, the item difficulty would be 0.90. Item difficulty is a misnomer, as high values mean low difficulty, and low values mean high difficulty.

Item discrimination. In CTT, item discrimination is a measure of how well an item differentiates between examinees. The Index of Discrimination for each item is computed by finding the difference between the percentage of students in the top one-third of the class who correctly answered that test item and the percentage in the bottom one-third who correctly answered that item. The Index of Discrimination ranges from -1 to 1. The closer it is to 1, the more discriminating the item. The maximum discrimination of 1 occurs when all examinees of the upper group got the item correct, and all the examinees of the lowest group got the item incorrect.

When more examinees from the upper group get the item correct than those of the lower group, the item is said to be positively discriminating. If, on the other hand, more examinees from the lower group get the item correct than those of the upper group, the item is said to be negatively discriminating, indicating the question is measuring something other than what the rest of the test is measuring (Thorndike,1997), measuring nothing, or simply poorly worded or confusing. Our preliminary item analysis revealed only one item with a negative item discrimination and it was eliminated from the TTCI. Interviews with students indicated that this item was confusing and poorly worded and since we already had an adequate number of questions for that concept, it was decided we should eliminate the question altogether.

Reliability. (Do repeated administrations of the test yield the same results?) A test cannot be valid if it is not reliable. Using Cronbach's alpha to check reliability for each of the three parts of the TTCI, we found each alpha exceeded .6 even with the limited data we now have. Elimination of the worst items in each part increased the Cronbach alphas and result in the lowest alpha of .66.

Summary and Conclusions

Concept inventories can be used for multiple purposes. The TTCI was designed, not for summative assessment, but primarily for formative assessment. Our objective was to provide an instrument that would assist instructors in identifying the common misconceptions which prevent students from acquiring a deep understanding of the important concepts the underlie heat transfer, fluids and thermodynamics.

We began our process by having experts identify those concepts which are crucially important but poorly understood by students. With ten concepts identified, we developed open-ended questions to help us determine if the concepts truly were poorly understood and to aid us in writing distractors for the multiple choice versions of the questions. We beta tested the multiple choice instrument, made appropriate changes to wording and eliminated poorly performing questions where necessary. The present version of the TTCI will be further tested in the coming year, but preliminary analyses demonstrate that the instrument is able to identify critical micro-level misconceptions that hinder students learning.

Our hope for future research is to begin developing simulations which will help students to confront their misconceptions and enable them to understand these crucial ten concepts that are fundamental to understanding the thermal sciences.

References

Adler, M. and Ziglio, E., eds. (1996). *Gazing into the Oracle: The Delphi Method and its Application to Social Policy and Public Health*. London: Jessica Kingsley Publishers.

Allen, K., Stone, A., Reed-Rhodes, T., & Murphy, T.J. (2004). *The statistics concepts inventory: Developing a valid and reliable instrument*. Paper presented at the American Society for Engineering Education Annual Conference, Salt Lake City.

Chi, M.T.H & Roscoe, R.D. (2002). The processes and challenges of conceptual change. In M. Limon & L. Mason (Eds.). *Reconsidering conceptual change: Issues in theory and Practice.* The Netherlands: Kluwer Academic Publishers.

Chi, M.T.H., Slotta, J.D., de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, 4, 27-43.

Chinn, C. A., and Brewster, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research,* 63, 1-49.

Dalkey, N. and Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9, 458-467.

Darmofal, D.L., Soderholm, D.H., & Brodeur, D.R. (2002). *Enhancing conceptual understanding with concept maps and concept questions*. Paper presented at the meeting of the American Society for Engineering Education, Montreal, Quebec.

Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-study survey of mechanics test data for introductory Physics courses. *American Journal of Physics* 66(1), 64-75.

Hestenes, D., Wells, M. and Swackhamer, G., Force Concept Inventory, *The Physics Teacher,* 30, 159-166, 1992.

Linstone, H. A. and Turoff, M., eds. (1975). *The Delphi Method: Techniques and Applications*. Reading MA: Addison-Wesley Publishing Company. Also available at http://www.is.njit.edu/pubs/delphibook/ch1.html

Martin, J.K., Mitchell, J., & Newell T. (2003). *Development of a concept inventory for fluid mechanics.* Paper presented at the 33rd annual ASEE/IEEE Frontiers in Education Conference, Boulder, CO.

Olds, B., Steveler, R., Miller, R. & Nelson, M.A. (2004). *Preliminary results from the development of a concept inventory in thermal and transport sciences*. Paper presented at the American Society for Engineering Education Annual Conference, Salt Lake City.

Sadler, R. (1989). Formative assessment and the design of instructional systems, *Instructional Science*, 18, 119-144.

Shepard, L. (2001). The role of classroom assessment in teaching and learning. *CSE Technical Report*. U.S.: California; 2000-02-00.

van Someren, M.W., Barnard, Y.F. & Sandberg, J.A.C. (1994). *The Think Aloud Method*. Academic Press.