Concept Inventories in Engineering Education

Teri Reed-Rhoads and P.K. Imbrie
School of Engineering Education
Purdue University

"What is the reason for the seasons?" or "What causes the phases of the Moon?" One would like to believe that if you asked these two seemingly innocuous questions to a group of "educated" people you would expect most, if not all, of individuals in the group to answer them correctly. However, would you believe that 21 of 23 randomly interviewed graduates, professors, and alumni from a prestigious university incorrectly responded to at least one of the astronomical questions (Harvard, 1977). If our college graduates have difficulty answering fundamental questions like these, what is their conceptual understanding of topics like Chemical Equilibrium, Thermal and Transport Sciences, Electromagnetics, or Nanotechnology? Faculty members would all like to say "my students understand, look at how well they do on my exams." But for each engineering subject taught, has the question ever really been asked, "What do our students know?" and "What do they think they understand and how do they come to these understandings?"

Concept inventories represent a relatively unique form of an assessment instrument with a multitude of possible uses that range from diagnostic and formative purposes to guide instructional planning, to summative purposes for evaluating overall learning and instructional effects at a student, classroom, and/or instructional program level. What makes them relatively unique compared to typical assessments of student academic achievement is that they tend to be highly focused on a small set of key constructs and understandings within a limited domain of academic content. Great care goes into conceptualizing the nature of the situations to be presented and in developing plausible distracters that represent a range of partially correct understandings to completely incorrect understandings and misconceptions. However, it should be stated that a specific definition of a concept inventory is not agreed upon. In fact, there are many published definitions that include the following succinct version "research-based distracter driven multiple-choice instruments" (Lindell, 2006: pg. 14). In addition, the method for developing concept inventories is not consistent or agreed upon. Again Lindell (2006) has a very nice summary of this where comparisons are made across Physics and Astronomy related concept inventories. This was also found to be the case in the creation of the Concept Inventory Central website (www.purdue.edu/SCI/Workshop) where developers were asked to post similar type summaries related to their instruments for others to consider in making a decision to use each instrument.

In recent years, the science, technology, engineering, and mathematics (STEM) disciplines have increased their use of Concept Inventories (CI) instruments to measure the value added to student learning by new ways of teaching important material (Evans, 2003). Utilizing a tool such as a CI can provide a learning opportunity for students and professors alike. Students can reveal their misconceptions for the first time, as well as open their minds to accepting scientific points of view. Professors "can form a basis for making instructional decisions, whether to validate students' correct yet unsure ideas, confront student misconceptions, reinforce ideas that are forming, or complement ideas that are accurate but only partial explanations." (http://www.learner.org, 2007)

Numerous studies have shown that many students lack correct conceptual understanding of science and engineering concepts, even after successful completion of courses in which these concepts are taught (Duit and Treagust, 2003, Hake, 1998, and McDermott, 1991). These studies point out the challenge of learning some concepts, as well as call for discussion on what level of conceptual understanding is needed throughout each aspect of a curriculum.

CI assessment instruments can provide valid, reliable data on conceptual understanding. Moreover, the creation, development, and use of these inventories foster constructive conversations among educators. The Force Concept Inventory (FCI) (Halloun and Hestenes, 1985; Hestenes, *et al.*, 1992) has proven instrumental in identifying effective pedagogy (Hake, 1998) and thus has served as an agent of change in Physics education (e.g., Crouch and Mazur, 2001).

A cited reference search of the FCI's major publications illustrates the potential influence of a well-developed instrument. Table 1, below, lists the number of citations (Web of Science *Cited Reference Search*; October 4, 2008).

Table 1: Cited references of major FCI publications

Article Year	Authors	Citations
1985	Halloun & Hestenes	145
1992	Hestenes, *Wells, & Swackhamer*	119
1998	Hake	160

Further analysis of the 1985 article highlights the influence of the FCI in fields outside physics and continuing intra-disciplinary interest (Figure 1 – data from September 6, 2006 search from Web of Science). The most-cited block is the five-year period ten years after initial publication.

Looking back at the FCI, it is apparent that significant development time was required to influence pedagogical practices. The 1985 publication was the result of three years' development, while the compelling results of Hake's 6000+ students was not presented until 13 years later. After the initial development and publication of the FCI, there have been at least 11 subsequent instruments developed in Physics alone (see Table No. 2). In the areas of science and mathematics, there are 9 general areas covered (see Table No. 3).

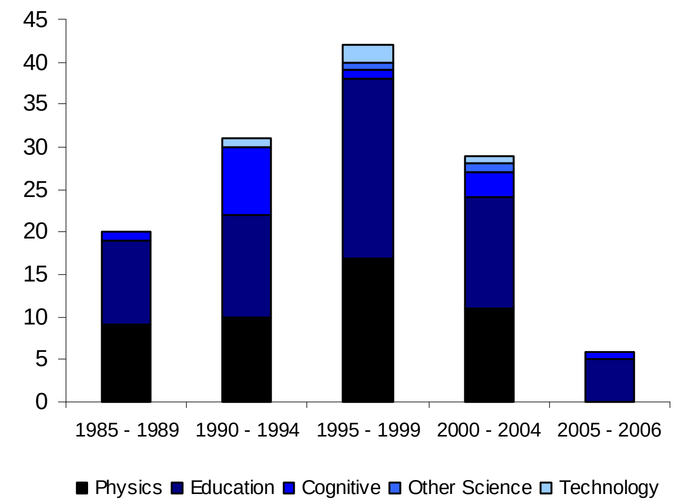

Figure 1: Cited references of 1985 FCI publication, by journal type

Table 2: Physics and Astronomy Specific Concept Inventories (Lindell, 2006)

1.	Force CI	7.	Determining and Interpreting Resistive Electric Circuits Concept Test
2.	Mechanics Baseline Test	8.	Energy and Motion Conceptual Survey
3.	Astronomy Diagnostic Test	9.	Force and Motion Conceptual Evaluation
4.	Brief Electricity and Magnetism Assessment	10.	Lunar Phases CI
5.	Conceptual Survey in Electricity and Magnetism	11.	Test of Understanding Graphs in Kinematics
6.	Diagnostic Exam Electricity and Magnetism	12.	Light and Color

Unlike the FCI, engineering CI development is occurring in a different environment. First, there is very little known or published on the engineering concepts and subject matter misconceptions. Whereas the FCI generally placed prior educational research in a new format (i.e., a multiple choice test), engineering faces the challenge of using the inventories as a tool to identify misconceptions, being only guided in a general sense of instructors' perceptions of potential misconceptions (Streveler et al., 2008). Second, the engineering inventories are being developed in an era of evolving accreditation standards that focus on learning outcomes. This coincidence could infer the purpose of the CIs is for accreditation instead of instruments aimed at increasing student learning. The FCI, on the other hand, might be viewed as a more charitable contribution to existing research. This difference could infuse a tincture of skepticism regarding the researchers' motivation (i.e., *have-to* vs. *want-to*). Third, in part due to the FCI, evidence is beyond anecdotal that traditional teaching is ineffective providing a landscape ripe for encouraging faculty to improve their practices with conceptual assessment imperative to demonstrate the effectiveness of innovations.

Table 3: Science and Mathematics Concept Inventories

1.	Biology	6.	Geosciences
2.	Calculus– Nelson	7.	Natural Selection
3.	Calculus– Epstein	8.	Organic Chemistry
4.	Chemical Equilibrium	9.	Physics (See Table No. 2 for full list)
5.	Genetics		

Despite these challenges, engineering experienced a flurry of development, primarily associated with the Foundation Coalition, beginning from 2001 – 2003 (see Evans, *et al.*, 2003 for a summary). However in more recent years, it appears that continued development, refinement, deployment, and application of concept inventories has waned. A literature review (Allen, 2006) identified 16 engineering concept inventories which have been expanded to 21 since a national workshop on concept inventory development and use sponsored by the National Science Foundation was held in May of 2007 (see Table No. 4). Figure 2 shows the number of people who have taken some of the instruments that are the most highly published at this time. The first two

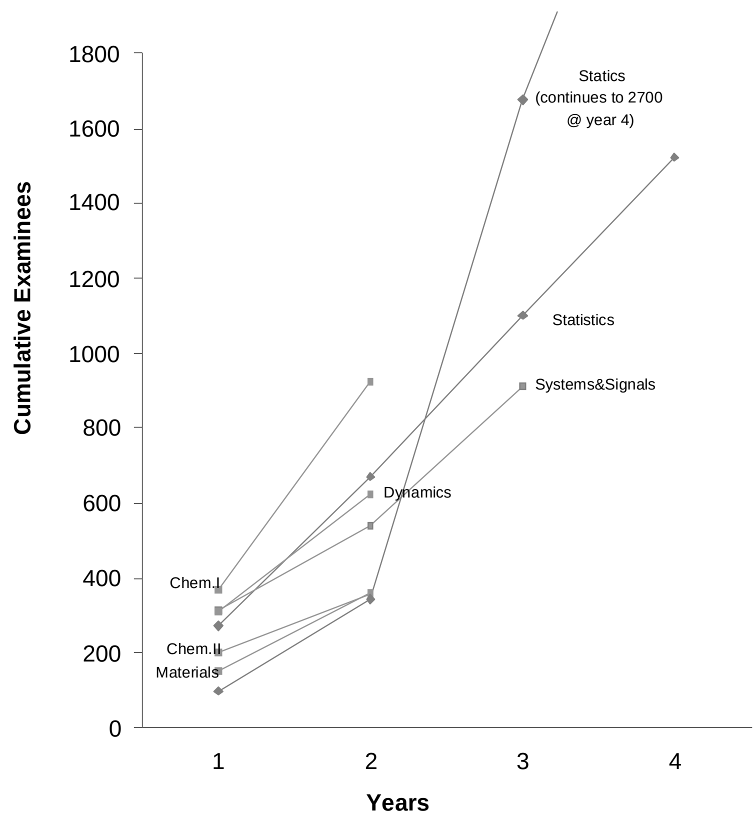

Figure 2: Cumulative examinees across years for engineering concept inventories

years' data for all inventories are similar, suggesting strong foundations with the capability to expand. It is important to note that Steif's Statics CI has achieved the high numbers through participation of upwards of 20 institutions through online access. This same trend is illustrated with the Statistics CI, which also went online early in its development. This highlights the broad applicability of these instruments and the potential broad applicability of other instruments if online access is facilitated.

Given the increasing intensity of the call to reform engineering education (NAE 2004, etc.), building a cohesive CI Community will provide a basis for constructive action toward a national collaborative effort to support continued development, refinement, analysis, and application of multiple instruments and to engage the engineering education community in productive conversations about assessing and improving conceptual understanding.

Table 4: Engineering Related Concept Inventories

1.	Chemistry	8.	Electromagnetics	15.	Statics
2.	Circuits	9.	Electronics	16.	Statistics
3.	Computer Engineering	10.	Fluid Mechanics	17.	Strength of Materials
4.	Computer Science	11.	Heat Transfer	18.	Thermal and Transport Sciences
5.	CS (Intro)	12.	Materials	19.	Thermodynamics
6.	Discrete Math (CS)	13.	Nanotechnology	20.	Waves
7.	Dynamics	14.	Signals and Systems	21.	Control Systems

To guide the inventory development process and/or to evaluate the measurement quality of an existing inventory, individual test items from an inventory can be analyzed using tools from classical test theory. For example, each question on an inventory (or on an inventory subscale) can be evaluated in terms of a set of indices such as an item's (a) difficulty level, (b) discriminability, (c) correlation with the total scale score, and (d) scale alpha if deleted. This information, along with the distribution of responses and information obtained from focus groups or protocol studies can be used to make judgments about the measurement quality of an item or set of items and help focus on needed revisions.

Item Difficulty: The *item difficulty* ranges from 0 to 1 and is simply the proportion of students who answered the item correctly. Questions with a low item difficulty are harder questions and those with a high item difficulty are easier. Item difficulty would be expected to change as a result of quality instruction and thus can be used as an index of instructional sensitivity.

Discrimination Index: The *discrimination index* is a measure of how well an item separates students who have a high score on the total test from those who have a low score. Questions with a large, positive discrimination index are generally useful because this provides evidence that the item is measuring the same construct as the whole inventory (or the subscale) and helps to contribute to the reliability of the test. On the other hand, questions with a low or negative discrimination index may need to be rewritten or reconsidered.

Correlation with the Total Score: As part of the item analysis, the correlation between an item's score and the total score for the remaining items is calculated. Such correlations will typically range from zero to 0.4, with values above 0.2 considered good (Nunnally 1967).

Questions that have higher correlations with the total score are more discriminating and contribute to a more reliable test.

Overall Alpha Rank: Coefficient alpha (α) is a commonly used estimate of the reliability of an instrument (or subscale). When analyzing the individual items of an instrument, it is possible to gain some sense of how each individual item contributes to the overall test reliability by looking at the *alpha-if-item-deleted* statistic. This is determined by omitting the item from the data set and calculating α for all of the remaining items of the test. This value can then be compared to the overall coefficient alpha for all items. If the alpha-if-item-deleted value is smaller, then removing the item would lower the overall test reliability. If the alpha-if-item-deleted value is larger, this indicates a poor question and removing the question causes the test reliability to go up. These questions should be examined to see if they can be improved (e.g. eliminating ambiguous wording or cues within the question, or reframing questions that involve too much guessing or that require recall only).

In addition to traditional classical test theory analysis, an item response theory (IRT) analysis can also be undertaken using both the two parameter logistic model and the nominal response model (Stone 2006). This analysis has been used in the development of the Statistics Concept Inventory to make decisions during the revision process, enabling sample independent comparisons of question versions to be made as well as giving insight as to how a question behaves over a range of abilities. The nominal response model provides item response curves for each distractor, not just the correct answer (Figure 3). The item response curve shows the probability of selecting that response across the ability distribution. These curves indicate which responses remain popular even as ability levels increase and may therefore indicate persistent misconceptions.

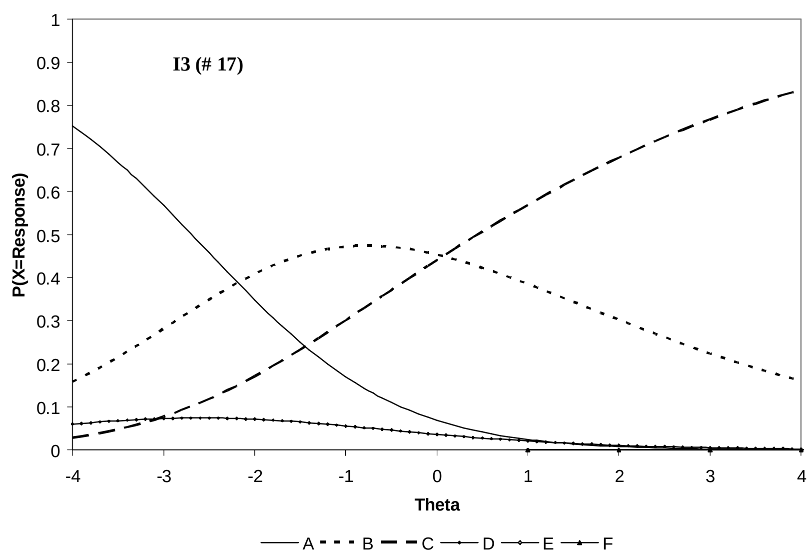

Figure 1: Item response curves for one question on the SCI from the nominal response model.

Work that has been done to date on the psychometrics of concept inventories has been somewhat limited in terms of the statistics applied to item and test/scale data, typically representing approaches derived from classical test theory and unidimensional IRT. While such approaches are extremely useful for estimating critical measurement properties of items and instruments, they are not terribly powerful for determining the validity of the instruments relative to the underlying constructs that such instruments purport to measure (see e.g., Pellegrino, Chudowsky, & Glaser, 2001).

Statics Concept Inventory

The Statics Concept Inventory is designed to "detect errors associated with incorrect concepts, not with other skills (e.g., mathematical) necessary for Statics" (Steif, 2004). Questions which contain numbers require only trivial calculations, and incorrect answers are based on incorrect assumptions rather than incorrect calculations. The questions on the Statics Concept Inventory were developed primarily through the experience of the author and two statics professors at different universities. Distracters were also based on students' written responses to questions which require multiple statics concepts. Based on these findings, the instrument is divided into five sub-topics, one of which contains four further specific situations. Several of the misconceptions are: forces are missing from the diagrams; extra forces are included; incorrect understanding of a couple; incorrect understanding of normal forces; inability to balance forces in equilibrium.

The instrument has 7 of 27 items with discriminatory indices below 0.20 on a pre-test, with 12 of 27 above 0.40. Three of the low discriminating items relate to friction and another three relate to static equivalence. The friction questions had very low percent correct, which limits the discriminatory index. The static equivalence questions were not quite as low, but other research has shown that misconceptions in this area persist even after a statics course.

The Statics Concept Inventory is reported to have an alpha of 0.712 as a pre-test; no post-test results are reported. The author considers the instrument reliable for "an initial version" but would like to attain a value above 0.80.

In a second study, Steif and Dantzler (2005) perform more psychometric analyses of the Statics Concept Inventory using data from 245 students at five universities. Total scores indicated no significant differences due to gender or ethnicity in a 2×2 ANOVA. The instrument was highly reliable with α of 0.89. At one university (n = 105), a strong correlation was found between course grade (coded A=1, B=2, C=3) and inventory score (Spearman's $\rho = -0.547$, $p < 0.001$; ρ is negative because higher grades are coded with lower numbers). At the same university, mean scores nearly doubled from pre-test (39.2%, n = 127) to post-test (75.3%, n = 105).

Analysis of data from 100 students in fall 2004 found strong correlations between scores on the Statics Concept Inventory and average course examination scores or final examination scores. Comparisons were made between course exam scores and inventory scores by quartile. These tables show a clear trend between higher post test inventory scores and higher exam scores. An example of this is reproduced in Table 5. Also, a finer analysis of student course exams showed that students who committed certain errors in their course examinations were found to have lower related sub scores on the inventory.

Table 5: A comparison of the percentage of students in each quartile on the Statics Concept Inventory and the final exam score received at one site.

	Quartile 1	*Quartile 2*	*Quartile 3*	*Quartile 4*
A (n=9)	0%	11%	44%	44%
B (n=8)	0%	25%	25%	50%
C (n=10)	50%	20%	0%	30%
D (n=2)	50%	50%	0%	0%
F (n=9)	33%	44%	22%	0%

The authors also define the Inventory-Exam Discrepancy (IED) for each student by:

$$\text{IED} = \frac{\text{Inventory Score} * \text{Class Exam Mean}}{\text{Class Inventory Mean}} - \text{Exam Score}$$

IED has a mean value of 0 and is positive for students who did well on the inventory compared to their exam score (relative to the class average on each measure). Highly correlated exam and inventory scores result in a narrow range of IED. The range of IED for the three classes varied widely along with the correlations between scores.

Items performed very well on the discriminatory index, with values ranging from 0.26 to 0.84, with only the former below 0.30. Difficulties were also in a preferred range, from 31% to 85%. A confirmatory factor analysis (CFA) model was fit to the item scores with each item assumed to measure one of eight hypothesized constructs. The fit indices suggest the model is "acceptable" (e.g., goodness-of-fit 0.90), although there is some room for improvement.

Further results (Steif and Hansen, 2006) are tempered, although still encouraging. For a survey of 1331 students at seven universities, the reliability was 0.82 and four items had a discrimination index below 0.30. For five courses, the correlation (Pearson's *r*) between inventory score and course final exam score ranged between 0.24 and 0.62. Two additional courses provided data from two mid-term examinations in addition to the final. The inventory correlated most-highly with the first exam (r = 0.65 and 0.46), while the correlations were mixed across second-exam/inventory and final-exam/inventory. Further, six classes provided data which showed that correlations across test scores *within each class* were of similar magnitude to the various inventory-exam correlations (read: there is no pattern of course exams being better or worse predictors of future exams than the inventory as a predictor of exam scores). These data were collected from an online test [http://engineering-education.com/CATS/intro.htm], and more recent results are available therein.

Statistics Concept Inventory

Topic selection for the Statistics Concept Inventory (SCI) began with an instructor survey to identify critical topics (Stone et al., 2003). In addition, the Advanced Placement Exam syllabus for statistics, widely used textbooks, and research from statistics education literature identifying student misconceptions were utilized. This information was used to draft question and response sets, incorporating known student misconceptions when possible. The target audience for the SCI was the engineering student population, however due to the homogeneity of content within introductory statistics courses and intentionally limiting the engineering contexts and jargon, the SCI should be able to be widely used.

The SCI consists of 38 items categorized into 4 sub-areas based on the content: probability, descriptive statistics, inferential statistics, and graphical representations. In addition to the item classification, a taxonomy of errors and misconceptions with their associated responses has been compiled.

More than 1200 students have taken the SCI in a variety of statistics courses from engineering, mathematics, psychology, and communication departments. The majority of students have been engineering majors taking an engineering statistics course. Revisions were made to the instrument after each administration based on item analysis including item discrimination, item difficulty, the item response distribution, comments from student focus groups, evaluation for test-wiseness cues, and reliability analysis. Eighteen of the items have a discrimination index higher than 0.4 (considered good), 14 items have moderate values between 0.2 and 0.4, while six items have poorer values of less than .2.

Post-test scores have been consistently low ranging from 45-50% each semester since the pilot version in fall 2002. Scores vary more by course, with courses serving non-engineering student populations tending to score lower. These courses generally have younger students with less mathematics and science backgrounds. Where pre-test scores are available, gains have been minimal (normalized gains range from 1-25%), consistent with the range found with other concept inventories.

The SCI moved to online administration during the fall of 2005. No systematic differences have been found between the paper and online administrations of the instrument. As part of the online version, participants were asked to rank their confidence in their answers on a scale from 1 (not confident at all) to 4 (very confident) (Allen, Rhoads, and Terry, 2006; Allen, 2006). This confidence ranking was then compared to the percent correct for each question using rank order. There was a significant positive correlation between the two rankings (r = 0.334, p = 0.020). Items were categorized as having over-confidence or under-confidence when the rank order of their confidence rating differed from the rank order of the fraction correct by greater than 10 (Table 6). Probability is an area where many known misconceptions have been identified. Of the 11 questions covering probability topics, 5 fell into the over-confident category with none in the under-confident category. This type of analysis may help identify not only legitimate misconceptions, but also areas of guessing and mastery.

Table 6: Comparison of items by confidence and subcategory on the SCI.

Subcategory	*Over-confident*	*Under-confident*	**Neither**
Probability	5	0	6
Descriptive	1	4	6
Inferential	3	2	4
Graphical	2	3	2
Total	11	9	18

Reliability of the instrument has been assessed from a classical test theory perspective and an IRT perspective. Coefficient alpha (a sample dependent measure) has varied from semester to semester, but has generally been near 0.75, with the fall 2005 administration having a value of 0.77. The item response theory reliability estimate obtained from all the data (based on the concept of test information) was found to be 0.78 (Allen et al., 2004; Stone, 2006; Allen, 2006). This initial work is now being extended through the development of “clicker” or personal response system questions. These questions are again driven by student misconceptions where each distractor has a meaning to the instructor and is able to alert him/her to specific and/or common misconceptions within their class (Murphy et al., 2008).

Gaps in research

Gaps in research exist in several areas related to engineering concept inventories. Essentially, only the basic engineering science courses have been studied to date. Therefore, the upper division courses and subject areas are sparsely represented. These include the lack of instruments in specific areas such as engineering design and economics, as well as courses like quality engineering to just name a few.

Few of the engineering concept inventories have related attitudinal instruments that allow for the measurement of both the affective and cognitive domain. Gal, I. and Ginsburg, L. (1994) suggest that attitudes and beliefs of participants towards a certain field of study may affect the recorded instrument results. Thus, potential dependencies between instrument performance and attitudinal data should be investigated. The general area of statistics has three such instruments; the Statistics Attitude Survey (Roberts and Bilderback, 1980), the Attitude Toward Statistics (Wise, 1985) and Survey of Attitude Toward Statistics (Schau et al., 1995). However, this is not the norm in other areas where CI have been developed within engineering and these attitudinal instruments in statistics were not written specifically for engineering though some research exists in the application of these instruments specifically to engineering populations (Reed Rhoads and Hubele, 2000). A recent application of an attitudinal instrument is being developed with the Computer Science Attitude Survey by Moskal (2004).

The effects of demographic data such as gender, discipline, and class level with respect to both, instrument performance and displayed attitude is needed for most of the currently available CI's. The study of various types of bias that could be manifested within a CI has only been briefly approached. Again, the most extensive but not the only research has been done with the FCI where authors have created animated versions (Dancy, 2000) and "gendered" versions (McCullough and Meltzer, 2001 and McCullough, 2004) and tested each with samples of students. Relative to engineering CI, gender bias is mentioned more often than racial or ethnic bias, while "English as a second language" bias is rarely mentioned (Allen, et. al., 2004, Olds, et.al., 2004, Richardson, et. al., 2003 and Wage, et. al., 2002). It should be noted that language is not the only bias created by geographic differences. The FCI found cultural bias in that one of the original questions had one person pushing another person with their feet – an offensive act in other parts of the world. The lack of study of bias is most often related to the low numbers of students of various populations having utilized the instruments so that statistical testing can be appropriately performed. Though women are still the most under-represented population in engineering, their numbers tend to be high enough to allow for some statistical testing. Finally, a few of the engineering CI's have been/are being translated in other languages, namely the Statics and Signal and Systems CI's. Additional work in this area needs to be increased as the majority of these instruments are very American-centric to date.

Faculty Valuing/Changing Behavior

Though a Hake-type event has not occurred with the engineering concept inventories, there are indications of faculty use and changing behavior. These include the increasing requests for access/use of concept inventories to each of their authors. In addition, as assessment is a required portion on all education related grants, the concept inventories are often sought as an accepted means of using a somewhat standardized instrument to assess educational innovation in the cognitive domain. Their use and discussion of their use is appearing in more and more conference proceedings and lagging behind is their appearance in journal articles. As discussed earlier, another driver for use, though this is not necessarily related to faculty behavior, is the need for outcomes assessment as related to ABET.

ciHUB: Developing a Community of Concept Inventory Developers and Users

Accessibility to valid and reliable Ci's is but one of the issues faculty encounter when trying to use CI's to assess students' understanding of fundamental concepts. They also need access to a larger community of faculty who have developed novel approaches to help students overcome conceptual miss-understandings. One approach to facilitating accessibility to CI resources is to create a virtual CI community (VC) where faculty can collaborate with others. A cybercommunity of researchers, developers, and faculty has the potential to eliminate existing barriers for those seeking to use CI's and enhance the educational experience for students and their learning. We propose the VC to be called ciHUB and modeled after the highly successful nanoHUB.org VC. In addition, this infrastructure would host a national database to allow the CI researcher community to look at longitudinal data and cross-instrument correlations.

In today's technology rich environments one can easily imagine a VC that is more than simply an online resource where users can access or download content, or take CI's, but also as shared infrastructure where others can upload their own contributions, collaborate with colleagues, and develop partnerships. Such an environment could allow students, educators, and researchers in academia access to a variety of different CI's that provide real-time feedback, provide tutorials, learning modules, specific course materials, and even seminars by faculty; all for the common purpose of enhancing learning, Add the most relevant education content and collaborative software, and create a virtual community dedicated to strengthening CI development, use, and dissemination. In addition, this type of community would facilitate research not just within a specific topic area relative to a specific CI, but it would also allow research between CI. For example, the FCI might be the correct pre-test to Steif's Statics CI. This type of research could have larger implications as CI development progresses into upper division course content.

Acknowledgements

There are many persons involved and interested in this area of research and its future. Many have contributed to this field through continuing to write grants, papers, summaries, etc. The National Science Foundation has supported this research through many grants, such as DUE-0731232 and DUE-0206977**.** In addition, thanks to other collaborators who continue to share the CI passion are in order. These include, but are not limited to, Kirk Allen, Andrea Stone, TJ Murphy, Robert Terry, Paul Steif, Jim Pellegrino, Jeff Froyd, Barbara Moskal, Ron Miller, Jay Martin, and Louis Dibello.

Comprehensive Bibliography
(review paper by Allen, Stone, Reed-Rhoads, and Murphy and additional resources)

Allen, Kirk. 2006. Statistics Concept Inventory. Doctoral Dissertation, Industrial Engineering, University of Oklahoma.

Allen, Kirk, Teri Reed Rhoads, and Robert Terry. 2006. Work in Progress: Assessing Student Confidence of Introductory Statistics Concepts. Paper presented at the 36th ASEE/IEEE Frontiers in Education Conference, October 28-31, at San Diego, CA.

Allen, Kirk, Andrea Stone, Teri Reed Rhoads, and Teri J. Murphy. 2004. The Statistics Concept Inventory: Developing a Valid and Reliable Instrument. Paper presented at the American Society for Engineering Education Annual Conference and Exposition, at Salt Lake City, UT.

Anderson, D.L., K.M. Fisher, and G.J. Norman. 2002. Development and Evaluation of the Conceptual Inventory of Natural Selection. *Journal of Research in Science Teaching* 39 (10):952-978.

Beichner, Robert J. 1994. Testing student interpretation of kinematics graphs. *American Journal of Physics* 62 (8):750-762.

Crouch, Catherine, and Eric Mazur. 2001. Peer Instruction: Ten years of experience and results. *American Journal of Physics* 69 (9):970-977.

Dancy, M.H. 2000. Investigating Animations for Assessment with an Animated Version of the Force Concept Inventory. Doctoral Dissertation, Physics, North Carolina State University.

Duit, R. and D.F. Treagust. 2003. Conceptual Change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*. 25 (6): 671-688.

Elrod, Susan. 2007. http://bioliteracy.net/Readings/papersSubmittedPDF/Elrod.pdf, 3/18/07.

Engelhardt, Paula Vetter, and Robert J. Beichner. 2004. Students' understanding of direct current resistive electrical circuits. *American Journal of Physics* 72 (1):98-115.

Evans, D. L., Gary L. Gray, Stephen Krause, Jay Martin, Clark Midkiff, Branisla M. Notaros, Michael Pavelich, David Rancour, Teri Reed Rhoads, Paul Steif, Ruth Streveler, and

Kathleen Wage. 2003. Progress on Concept Inventory Assessment Tools. Paper presented at the 33rd ASEE/IEEE Frontiers in Education Conference, November 5-8, at Boulder, CO.
Foundation Coalition. 2005. *Foundation Coalition Key Components: Concept Inventories* [cited February 22 2005]. Available from http://www.foundationcoalition.org/home/keycomponents/concept/index.html.
Gal, I. and Ginsburg, L. (1994), "The role of beliefs and Attitudes in Learning Statistics: Towards and Assessment Framework," *Journal of Statistics Education,* 2 (2).
Gary, G. L., D. L. Evans, Phillip Cornwell, Francesco Costanzo, and Brian Self. 2003. Toward a Nationwide Dynamics Concept Inventory Assessment Test. Paper presented at the 2003 American Society for Engineering Education Annual Conference and Exposition.
Gray, Gary L., Francesco Costanzo, Don Evans, Phillip Cornwell, Brian Self, and Jill Lane. 2005. The dynamics concept inventory assessment test: a progress report and some results. Paper presented at the 2005 American Society for Engineering Education Annual Conference & Exposition.
Hake, R. R. 1987. Promoting Student Crossover to the Newtonian World. *American Journal of Physics* 55 (10):878-883.
———. 1992. Socratic Pedagogy in the Introductory Physics Laboratory. *The Physics Teacher*.
Hake, Richard R. 1998. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 66 (1):64-74.
Halloun, Ibrahim, and David Hestenes. 1985. Common sense concepts about motion. *American Journal of Physics* 53 (11):1056-1065.
———. 1985. The initial knowledge state of college physics students. *American Journal of Physics* 53 (11):1043-1055.
Heller, Patricia, and Douglas Huffman. 1995. Interpreting the Force Concept Inventory: A Reply to Hestenes and Halloun. *The Physics Teacher* 33:503-511.
Hestenes, David. 1987. Toward a modeling theory of physics instruction. *American Journal of Physics* 55 (5):440-454.
Hestenes, David, and Ibrahim Halloun. 1995. Interpreting the Force Concept Inventory: A Response to March 1995 Critique by Huffman and Heller. *Physics Teacher* 33 (8):502,504-06.
Hestenes, David, and Malcolm Wells. 1992. A Mechanics Baseline Test. *The Physics Teacher* 30:159-166.
Hestenes, David, Malcolm Wells, and Gregg Swackhamer. 1992. Force Concept Inventory. *The Physics Teacher* 30 (3):141-58.
Huffman, Douglas, and Patricia Heller. 1995. What Does the Force Concept Inventory Actually Measure? *Physics Teacher* 33 (3):138-43.
Jacobi, Anthony, Jay Martin, John Mitchell, and Ty Newell. 2003. A Concept Inventory for Heat Transfer. In *33rd ASEE/IEEE Frontiers in Education Conference*. Boulder, CO.
———. 2004. Work in Progress: A Concept Inventory for Heat Transfer. In *34th ASEE/IEEE Frontiers in Education Conference*. Savannah, GA.
Jordan, William, Henry Cardenas, and Chad B. O'Neal. 2005. Using a Materials Concept Inventory to Assess and Introductory Materials Class: Potentials and Problems. Paper presented at the American Society for Engineering Education Annual Conference and Exposition.
Kim, E., and S.-J. Pak. 2000. Students do not overcome conceptual difficulties after solving 1000 problems. *American Journal of Physics* 70 (7):759-765.
Knight, Randall D. 2002. *Five Easy Lessons: Strategies for Successful Physics Teaching*. San Francisco: Pearson Education.

Krause, Stephen, James Birk, Richard Bauer, Brooke Jenkins, and Michael Pavelich. 2004. Development, Testing, and Application of a Chemistry Concept Inventory. Paper presented at the ASEE/IEEE Frontiers in Education Conference, at Savannah, GA.
Krause, Stephen, J. Chris Decker, and Richard Griffin. 2003. Using a Materials Concept Inventory to assess conceptual gain in introductory materials engineering courses. Paper presented at the 33rd ASEE/IEEE Frontiers in Education Conference, November 5-8, at Boulder, CO.
Krause, Stephen, A. Tasooji, and Richard Griffin. 2004. Origins of Misconceptions in a Materials Concept Inventory From Student Focus Groups. Paper presented at the Ameican Society for Enfineering Education Annual Conference & Exposition.
Libarkin, J.C., and S.W. Anderson. 2005. Assessment of Learning in Entry-Level Geoscience Courses: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education* 53:394-401.
Lindell, Rebecca S., Elizabeth Peak and Thomas M. Foster, "Are They All Created Equal? A Comparison of Different Concept Inventory Development Methodologies", 2006 Physics Education Research Conference Proceedings, American Institute of Physics.
Maloney, David P., Thomas L. O'Kuma, Curtis J. Hieggelke, and Van Heuvelen Alan. 2001. Surveying students' conceptual knowledge of electricity and magnetism. *Physics Education Research, American Journal of Physics Supplement* 69 (7-S1):S12-S23.
Martin, Jay, John Mitchell, and Ty Newell. 2003. Development of a concept inventory for fluid mechanics. Paper presented at the 33rd ASEE/IEEE Frontiers in Education Conference, November 5-8, at Boulder, CO.
———. 2004. Work in Progress: Analysis of Reliability of the Fluid Mechanics Concept Inventory. Paper presented at the 34th ASEE/IEEE Frontiers in Education Conference, at Savannah, GA.
Mazur, Eric. 1997. *Peer Instruction*. Upper Saddle River, NJ: Prentice Hall.
McCullough, L. and David Meltzer. 2001. Differences in Male/Female Response Patterns on Alternative-format Versions of the Force Concept Inventory. Paper presented at the 2001 Physics Education Research Conference, July 25 -26, Rochester, NY.
McCullough, L. 2004. Gender, context, and physics assessment. *Journal of Women's Studies* 5(4), 20-30.
McDermott, L. 1991. Millikan Lecture 1990: What we teach and what is learned: Closing the gap, *American Journal of Physics* **59**, 301–315.
Michel, H., J. Jackson, P. Fortier, and H. Liu. 2006. *Computer Engineering Concept Inventory* [cited April 17 2006]. Available from http://www.foundationcoalition.org/home/keycomponents/concept/computer.html.
Midkiff, Clark, L. Litzinger, and D. L. Evans. 2001. Development of Engineering Thermodynamics Concept Inventory Instruments. Paper presented at the 31st ASEE/IEEE Frontiers in Education Conference.
Morgan, J., and J. Richardson. 2005. *Strength of Materials (SOM) Concept Inventory* [pdf or ppt] [cited April 8 2005]. Available from http://www.foundationcoalition.org/home/keycomponents/concept/strength.html.
Moskal, Barbara, Debora Lurie, and Stephen Cooper. 2004. Evaluating the Effectiveness of a New Instructional Approach, Proceedings of the 35th SIGSE Technical Symposium on Computer Science Education, Norfolk, VA, 75-79.
Murphy, T. J., McKnight, C. C., Richman, M. B., & Terry, R. (submitted). Clickers in introductory statistics courses.
Murphy, T. J. , 2007-08-03 "Clicker Questions for Conceptual Understanding" Paper presented at the annual meeting of the Mathematical Association of America, The Fairmont Hotel, San Jose, CA <Not Available>. 2008-09-12 from http://www.allacademic.com/meta/p205363_index.html

National Academy of Engineering. (2004). *The Engineer of 2020.* The National Academies Press, Washington, D.C.

Notaros, Branisla M. 2006. *Electromagnetics Concept Inventory* [cited April 17 2006]. Available from http://www.foundationcoalition.org/home/keycomponents/concept/electromagnetics.html.

Notaros, Branislav M. 2002. Concept Inventory Assessment Instruments in Electromagnetics Education. Paper presented at the IEEE Antennas and Propagation Society International Symposium, at San Antonio, TX.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Olds, Barbara M., Ruth Streveler, Ronald L. Miller, and Mary A. Nelson. 2004. Preliminary Results from the development of a concept inventory in thermal and transport science. Paper presented at the American Society for Engineering Education Annual Conference and Exposition, June, at Salt Lake City, UT.

Rebello, N. Sanjay, and Dean A. Zollman. 2004. The effect of distractors on student performance on the Force Concept Inventory. *American Journal of Physics* 72 (1):116-125.

Rhoads, Teri Reed and Norma Faris Hubele. 2000. "Student Attitudes Toward Statistics Before and After a Computer-Integrated Introductory Statistics Course". *IEEE Transactions on Education*.43 (2): 182 - 187.

Rhoads, Teri Reed, and R. J. Roedel. 1999. The Wave Concept Inventory - a cognitive instrument based on Bloom's Taxonomy. Paper presented at the 29th ASEE/IEEE Frontiers in Education Conference.

Richardson, J., Paul Steif, J. Morgan, and Johan A. Dantzler. 2003. Development of a Concept Inventory for Strength of Materials. Paper presented at the 33rd ASEE/IEEE Frontiers in Education Annual Conference and Exposition.

Roberts, D. M.,&Bilderback, E.W. (1980). Reliability and validity of a statistics attitude survey. *Educational and Psychological Measurement*, *40*, 235-238.

Roedel, R. J., S. El-Ghazaly, Teri Reed Rhoads, and E. El-Sharawy. 1998. The Wave Concepts Inventory - An Assessment Tool for Courses in Electromagnetic Engineering. Paper presented at the Frontiers in Education Conference, November, at Tempe, AZ.

Saul, Jeffery M. 1998. Beyond problem solving: evaluating introductory physics courses through the hidden curriculum. Dissertation, Department of Physics, University of Maryland.

Schau, C., Stevens, J., Dauphinee, T. L.,&DelVecchio, A. (1995), The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement*, *55*, 868-875.

Simoni, M.F., M.E. Herniter, and B.A. Ferguson. 2004. Concepts to Questions: Creating an Electronics Concept Inventory Exam. Paper presented at the American Society for Engineering Education Annual Conference and Exposition.

Skromme, Brian J. 2005. *Spreadsheets to Promote Interactive Engagement in Semiconductor Device Courses* 2005 [cited September 20 2005]. Available from http://cresmet.asu.edu/proj_res/spreadsheets/Index.htm.

Steif, Paul. 2004. Initial data from a statics concept inventory. Paper presented at the 2004 American Society for Engineering Education Annual Conference and Exposition.

Steif, Paul, and Johan A. Dantzler. 2005. A Statics Concept Inventory: Development and Psychometric Analysis. *Journal of Engineering Education* 33:363-371.

Steif, Paul, Anna Dollar, and Johan A. Dantzler. 2005. Results from a statics concept inventory and their relationship to other measures of performance in statics. Paper presented at the 35th ASEE/IEEE Frontiers in Education Conference, at Indianapolis, IN.

Steif, Paul, and M. Hansen. 2006. *Comparisons between perfomance in a statics concept inventory and course examinations* (in press) 2006 [cited February 8 2006]. Available from http://www.me.cmu.edu/people/faculty/steif/educationalresearch.htm.
Stone, Andrea. 2006. A Psychometric Analysis of the Statistics Concept Inventory. Doctoral Dissertation, Mathematics, University of Oklahoma, Norman, OK.
Stone, Andrea, Kirk Allen, Teri Reed Rhoads, Teri J. Murphy, Randa L. Shehab, and Chaitanya Saha. 2003. The Statistics Concept Inventory: A Pilot Study. Paper presented at the ASEE/IEEE Frontiers in Education Conference, November 5-8, at Boulder, CO.
Streveler, Ruth A., Thomas A. Litzinger, Ronald L. Miller, and Paul S. Steif. 2008. Learning Conceptual Knowledge in the Engineering Sciences: Overview and Future Research Directions. *Journal of Engineering Education* 97 (3): 279-294.
Thornton, Ronald K., and David R. Sokoloff. 1998. Assessing student learning of newton's laws: the force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics* 66 (4):338-352.
Tobias, Sheila, and R. R. Hake. 1988. Professors as physics students: what can they teach us. *American Journal of Physics* 56 (9):786-794.
Voska, K.W., and H.W. Heikkinen. 2000. Identification and Analysis of Student Conceptions Used to Solve Chemical Equilibrium Problems. *Journal of Research in Science Teaching* 37:160-176.
Wage, Kathleen, John R. Buck, Thad B. Welch, and Cameron H.G. Wright. 2002. Testing and Validation of the Signals and Systems Concept Inventory. Paper presented at the 2nd IEEE Signal Processing Education Workshop, Proceedings of the 2002 IEEE 10th.
Wage, Kathleen, John R. Buck, Cameron H.G. Wright, and Thad B. Welch. 2005. The Signals and Systems Concept Inventory. *IEEE Transactions on Education* 48 (3):448-461.
Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement*, *45*, 401-405.