2018 ASEE Annual Conference & Exposition
June 24-27, 2018, Salt Palace Convention Center, Salt Lake City, UT

Paper ID #22470

# Creating a Concept Inventory - Lessons Learned

**Dr. Nancy E. Study, Pennsylvania State University, Erie (Behrend College)**

Dr. Nancy E. Study is on the faculty of the School of Engineering at Penn State Behrend where she teaches courses in engineering graphics and rapid prototyping, and is the coordinator of the rapid prototyping lab. Her research interests include visualization, standardization of CAD practices, and haptics. Nancy is a former chair of the ASEE Engineering Design Graphics Division and is currently the Circulation Manager and Treasurer of the Engineering Design Graphics Journal. She received her B.S. from Missouri State University, and M.S. and Ph.D. from Purdue University.

**Dr. Steven Nozaki, Pennsylvania State University, Erie (Behrend College)**

Ph.D. Engineering Education - Pennsylvania State University, Erie (Behrend College)

**Dr. Sheryl A. Sorby, University of Cincinnati**

Dr. Sheryl Sorby is currently a Professor of STEM Education at the University of Cincinnati and was recently a Fulbright Scholar at the Dublin Institute of Technology in Dublin, Ireland. She is a professor emerita of Mechanical Engineering-Engineering Mechanics at Michigan Technological University and the P.I. or co-P.I. on more than $14M in grant funding, most for educational projects. She is the former Associate Dean for Academic Programs in the College of Engineering at Michigan Tech and she served at the National Science Foundation as a Program Director in the Division of Undergraduate Education from January 2007 through August 2009. Prior to her appointment as Associate Dean, Dr. Sorby served as chair of the Engineering Fundamentals Department at Michigan Tech. She received a B.S. in Civil Engineering, an M.S. in Engineering Mechanics, and a Ph.D. in Mechanical Engineering-Engineering Mechanics, all from Michigan Tech. Dr. Sorby has a well-established research program in spatial visualization and is actively involved in the development of various educational programs.

**Dr. Mary A. Sadowski, Purdue University, West Lafayette (Polytechnic Institute)**

Mary A. Sadowski has been at Purdue since 2003 and until September 1, 2011 served as the Associate Dean for Undergraduate Programs and Learning in the Purdue College of Technology where she provided leadership for strategic initiatives in undergraduate education. She was Dean of Purdue Extended Campus from 2011 to 2015 and professor of Computer Graphics from 2015 -2017 when she officially retired from Purdue.

As a professor of Computer Graphics, her research interests include enhancing visualization skills, creative thinking, and learning styles. She developed a Delphi instrument to gather data to create a concept inventory for engineering graphics and has worked with a team to develop and test the fully developed Engineering Graphics concept inventory. As a professor at both Purdue University and Arizona State University, Mary's specialty is computer and technical graphics. Dr. Sadowski received her B.S. from Bowling Green State University, her M.S. from The Ohio State University, and her Ph.D. from Purdue University.

**Dr. Heidi M. Steinhauer, Embry-Riddle Aeronautical Univ., Daytona Beach**

Heidi M. Steinhauer is a Professor of Engineering, Department Chair of the Engineering Fundamentals Department, co-advisor for the only all-women's Baja SAE Team, Founding Member of FIRST (Female Initiative Reaching Success Together), and former director for GEMS (Girls in Engineering, Math, and Science). Dr. Steinhauer's awards include the ABET Presidential Award of Diversity and a three time winner of the Women's Vision Award. She has presented papers at ASEE Annual Conference, the ASEE Global Colloquium, Research in Engineering Education Symposium, Engineering Design Graphics Division Mid-Year Conference, Additive Manufacturers Users Group, and Solid Free-Form Fabrication Symposium. Her research interests center around the development and assessment of students' spatial visualization skills, the effective integration of 3D modeling into engineering design, and the impact of

contextualized hands-on applications on student learning and success. She has taught Engineering Graphics, Introduction to Engineering Design, Automation and Rapid Prototyping, and has developed several advanced applications of 3D modeling courses. Dr. Steinhauer received her B.S. in Aircraft Engineering and her M.S. in Systems Engineering, and her Ph.D. in Engineering Education from Virginia Tech.

**Dr. Ronald L. Miller, Colorado School of Mines**

Ronald L. Miller is a professor of chemical engineering and Director of the Center for Engineering Education at the Colorado School of Mines, where he has taught chemical engineering and interdisciplinary courses and conducted engineering education research for the past 26 years. Miller has received three university-wide teaching awards and has held a Jenni teaching fellowship at CSM. He has received grant awards for education research from the National Science Foundation, the U.S. Department of Education FIPSE program, the National Endowment for the Humanities, and the Colorado Commission on Higher Education and has published widely in engineering education literature. His research interests include measuring and repairing engineering student misconceptions in thermal and transport science.

**Mr. Kaloki Nabutola, Embry-Riddle Aeronautical Univ., Daytona Beach**

I am currently pursuing a Ph. D. in Mechanical Engineering at Embry-Riddle Aeronautical University focusing on heat transfer. I am also a graduate research assistant for the Engineering Fundamentals department. I have earned a Master of Science in Mechanical Engineering (CFD and Heat Transfer) and a Bachelor of Science in Aerospace Engineering (Propulsion) from Embry-Riddle Aeronautical University, Florida.

# Creating a Concept Inventory - Lessons Learned

## Abstract

A concept inventory (CI) is an instrument in educational assessment that can help identify conceptual understanding and identify common misconceptions among students. Since the development of the first concept inventory (the Force Concept Inventory in Physics education), a number of other instruments have been created in a variety STEM topics. For example, there are currently concept inventories in thermodynamics, heat transfer, and statics. The use of concept inventories has spurred educational reform in a wide range of settings and have also been used in course development to identify potential topics for inclusion and to aid in assessment of course outcomes. This paper describes the lessons learned through the development of a concept inventory for engineering graphics.

## Introduction

At the start of this project, there was no nationally normed and validated instrument for engineering graphics. A group of graphics professionals set out to create one with support from the National Science Foundation. The instrument is referred to as the Engineering Graphics Concept Inventory (EGCI) and is currently on its gamma (third) version of the "final" document. Concept inventories have been used in engineering education research for over 20 years, there are numerous publications describing the methods used in creating the instruments. For this project, a previously established procedure called the "Assessment Triangle" was utilized. This paper will discuss many of the valuable lessons learned through the development process of the Engineering Graphics Concept Inventory and will also describe the challenges encountered for anyone considering developing a concept inventory to measure student learning or misconceptions.

Measurement can be thought of as the systems of rules and related instrumentation used to quantify properties of various entities [1], [2], [3]. In STEM and other natural sciences, it is a relatively straightforward process to conceptualize, as many of the dimensions being measured are physical properties that can readily be observed. For example, measuring distance in meters or miles. In social sciences however, measurement can be more problematic, as the properties being measured are latent. Consider attempting to measure aspects such as behavior or learning; the very nature of these attributes makes measurement and instrumentation a difficult task, as there are no readily available "rulers" to measure them. Currently there is no singular decisive methodology for CI instrument development, though there are several agreed upon steps that should be included. Methodologies in implementing these steps also vary, though all have common traits that quality instruments share [4], [5], [1], [6], [7]. The procedure outlined by Netemeyer [2] was the basis for several other established instruments, and served as a driving influence for this project.

Netemeyer suggests a linear model of instrument development motivated with empirical evidence. The critical first step, is to clearly define the traits or abilities being measured. Accurate definitions are necessary as they will inform item creation and the overall character of the instrument. The definitions should be informed by theory, and accurately reflect the content domain being measured. Literature and other appropriate sources should be thoroughly reviewed to best inform the theory that is framing the instrument. Once the construct has been defined, items can be generated whose aim is to measure understanding of the instrument. Systematic iterations of item testing and revisions to the instrument will improve the reliability and accuracy of the instrument. Depending on the nature of the trait and domain being measured, the breadth and depth of statistical measures will vary as to which are most appropriate for the items. Once the instrument has met determined standards for reliability and validity, the final steps are to try to explain sources of possible variance to more consistently interpret results**.**

**The Assessment Triangle**

The EGCI used the assessment triangle model proposed by Pellegrino [8] as the theoretical framework for CI development. This model has been recommended by the National Research Council as a framework for creating state-of-the-art assessment instruments and consists of three elements – cognition, observation, and interpretation. It is important to note the non-linearity of this development model and keep in mind interrelatedness of each of the elements. Each one of the three elements is informed by, and influences, the other two. At any stage in instrument development, the process can benefit from considering how one corner may influence the other two elements in the framework. More information about the assessment triangle and its use in concept inventory development has been reported by Streveler, et al. [9]. A visual representation of the triangle is shown in Figure 1.

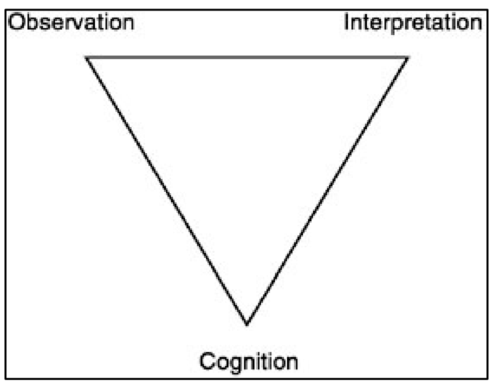


Figure 1: Assessment Triangle as suggested by Pellegrino [8]

- The cognition corner includes the identification of the concept inventory domain. Specifically, "what concepts will be measured by the instrument?" The activities in this corner typically involve conducting a Delphi Study (or similar consensus-building technique) and a literature review to identify concepts cited by others as being problematic.
- In the observation corner, the items that make up the instrument are developed through a series of steps. A typical procedure for creating the items includes the initial development of open-ended items followed by conversion to multiple-choice items. It should be noted that the items on a concept inventory have typically gone through a number of iterations before achieving their final form.

- Rigorous statistical analyses are applied in the interpretation corner. Typical statistical tests include those that assess reliability and validity or methods such as Item Response Theory to determine the relative difficulty of items.

The Cognition Corner

To address the cognition corner of the assessment triangle, in the first phase of developing the EGCI, a Delphi Study was conducted to identify core topics in engineering graphics [10]. In the initial round of the Delphi, a group of graphics professionals met at a workshop and used a variety of brainstorming techniques to identify a total of 120 unique topics in graphics education for potential inclusion on the instrument. An expert panel of industry representatives, high school teachers, and community college and university faculty was convened for the three rounds of the Delphi. Because the intended future uses of the EGCI are expected to extend beyond just the university classroom, it was important to have this variety of representatives on the expert panel. The final Delphi rounds were completed online. Through this process, the initial 120 topics were reduced to 37, and these topics were organized by the researchers into 10 concepts that served as the constructs for the development of the EGCI.

The Observation Corner

Identifying relevant topics and concepts in a domain such as engineering graphics can be difficult, as many of the topics were deemed to be interrelated. A framework that was in place from the onset of the Delphi provided a suitable means to categorize the concepts. When initially defining concepts, and later during testing the EGCI, those developing it needed to keep in mind that curricula vary and concepts that are taught and emphasized at one university may not be addressed at all, or addressed with a lower priority at another university. To use a specific example from the testing of the EGCI, students at one university may perform quite well on an auxiliary view question because those students have received instruction on auxiliary views. Students at another university may not perform well on the same question because auxiliary views have not been taught at the time of administering the EGCI, or may not be taught at all. Suggested methodologies for instrument development may require a strict definition for performance interpretation – such may not be the case with EGCI and the potential variation in curricula. This does not compromise the instrument's usefulness or generalizability. Rather, it requires that special attention be paid to score reporting and how scores are used.

Trying to identify student misconceptions was a time-consuming but very necessary process to complete the cognition corner. Item distractors were guided by student responses to pilot items. To do this, items with an open-ended format, typically requiring a sketched solution, were used in order to better observe the conceptions held by the students. Using the 10 concepts identified in the Delphi Study as a basis, pilot items were created by individual members of the research team. These items were intended to address a single identified concepts whenever possible. Participants completing pilot items were all engineering and engineering technology students enrolled in introductory engineering graphics courses at three different universities.

Administering open-ended items generated formative feedback in two areas. The first was that subjects' answers could provide a wealth of information from which to create potential future distractors, as patterns in the responses would reveal trends in student understanding. The second area was related to item structure. Participants were encouraged to comment on the format of the questions, making note of interpretations, possible errors, and any ambiguity in the items. This would help with revisions by revealing clarity issues with item presentation. To help identify major errors before the pilot study administration, twelve students in a senior level graphics and modeling course at one of the participating universities reviewed the items. These upper level students gave feedback on clarity of the questions and noted typos or mistakes in the graphics. These students did not participate in the open-ended response phase.

A total of 60 pilot items were drafted for the pilot study that covered all 10 concepts. Ideally, items would be randomly distributed and equally administered throughout each participating institution. Due to time constraints in classes at the participating institutions, not all 60 items could be tested equally among each population. A practice that proved to be beneficial in the acquisition of data was the flexibility in administration. Depending on the allowable time available at each institution, items were compiled into different sized packets based on the time available for testing. A testing protocol was established by the group so all packets in the pilot study were administered in a uniform fashion. Items were distributed so that each concept would receive equal exposure at each setting. This flexibility allowed the acquisition of a sufficient amount of responses.

Pilot responses from the participating institutions resulted in over 2000 unique responses to open-ended items. Open-ended responses were aggregated, printed, bound, and labeled for reference. At this point, the research team held a face-to-face meeting where the responses were coded by the researchers to look for trends in responses. Before extensive coding, inter-rater reliability was established by having the experts on the team, along with outside raters, look at the student responses and determine specifically what would make a response correct or incorrect based on what concept the question was testing. Having consistency established between raters allowed for independent coding to be conducted in a reliable manner for consideration by the group.

To provide a measure of consistent qualitive feedback on the open-ended items, students were also asked to rate each question using a Likert scale indicating whether or not they understood what the question was asking, and where they considered the question’s difficulty to lie on a scale from very easy to very difficult. While not directly used with any statistical measure, it was useful to have any additional insights when trying to infer student understanding of an item. For example, an item that was indicated as “difficult to understand” might have had an especially broad range of responses, and suggest a problem with the stem.

Having the hard copies of student responses was invaluable to the research team because it allowed the team to take notes, highlight items, and discuss specifically, for example, item number 54 on page 123. With members of the research team based at different institutions throughout the country, a majority of the work was done remotely, often over the phone and email. Regular progress was made possible in part by having indexed artifacts available for ready reference by each member.

With the data from the pilot study coded, work began on drafting potential distractors for the alpha version of the EGCI. Researchers reviewed the data from the pilot study to begin the work on drafting distractors. The group collectively wrote potential distractors for each item using the open-ended incorrect student responses as a guide. Aligning with suggested methodology, the final version of items would have three distractors in addition to the correct response. Often, the distractors were similar to the three most common incorrect coded responses from the pilot study. If there were less than three alternatives to use as distractors, it was often due to having multiple types of errors in a particular response. To address this, the multiple errors were divided up so that any potential distractor would have a single variation from the correct response. The images for the distractors were created using the software packages of Solidworks and Adobe Illustrator.

For the observation corner, after creating the distractors for the instrument, the alpha version of the EGCI was compiled. The alpha EGCI had three different versions with 20 questions each, and was tested with approximately 900 students at the participating universities. Soon after the alpha version was administered, the research group held a face-to-face meeting with their advisory board. This board was comprised of experts in engineering graphics, testing, concept inventories, and statistics. One of the major outcomes of that meeting was a request by the board to create a theory of action and a concept map, which should help ensure not only that the EGCI was comprehensive and covered the concepts from the Delphi study, but also that the members of the advisory board (who were not all graphics experts) had a better understanding of how the concepts are related. The concept map would be used to help identify and describe the relationships of the topics and help frame the structure of the instrument using "if, then" statements that demonstrate a causal relationship between the action and the intended outcome. The concept map that was developed by the research team and used to identify items for the final version of the instrument is shown in Figure 2.

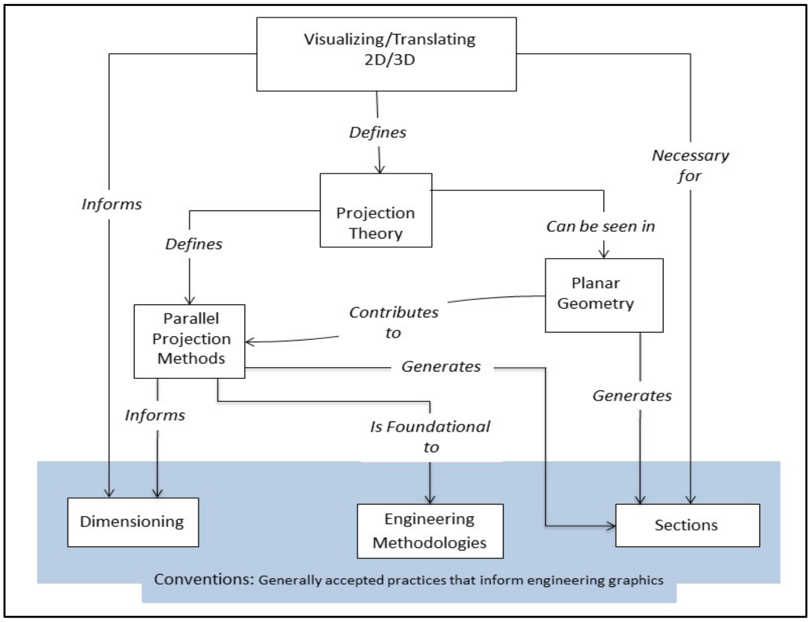


Figure 2: Concept Map used for final version of instrument

Having the relationships clearly defined through a concept map helped in two specific ways. First, it showed the complex inter-relationships between the concepts included on the instrument. It was evident that within the scope of the instrument, it was unlikely that concepts would be completely standalone. Understanding relationships between concepts would help for later interpretation of instrument scores and abilities. Second, it helped with the alignment of items into appropriate concepts. It is important to include several items that cover the breadth of each concept. With the complex relationships of each concept identified, any re-classification of items could be completed with a higher degree of confidence, through consulting the concept map.

Based on the analysis of the alpha version, items that did not meet the criteria for difficulty and discrimination were eliminated. Other questions were shifted due to the concept map's consolidation of topics. From this work, the beta version of the EGCI was developed. It covered six concepts and had three versions with 25 questions each. Some questions were on multiple versions of the EGCI. The beta version was tested with around 850 subjects at the participating universities. The results from the beta version were analyzed and the current gamma EGCI, which has only one version with 30 questions, was developed. There are questions of easy, medium, and hard difficulties for each concept, and all questions meet the criteria for discrimination. The testing of this gamma version is ongoing and currently nearly 2000 subjects have been tested in a variety of majors including aeronautical and mechanical engineering, plastics engineering technology, and mechanical engineering technology using the instrument.

The Interpretation Corner

The interpretation corner consisted of making sense of the observed performance of students on the instrument. When developing a new instrument, each step required interpretation to be able to continue with the instrument development. A major consideration for the project was that members were located at various institutions. As a result, communication and coordination was important to ensure members were able to work effectively. For many steps, members of the group were tasked with assignments that aligned with areas of individual expertise. It was essential that individual assignments be carried out thoroughly, and reported to the group in a manner that was practical and useful.

One example of management that was necessary for project progress during the interpretation corner phase was the need to track iterations of the items through the four versions. As one could imagine, managing such a large amount of information can be an arduous task. One member was placed in charge of data management and communication in order to streamline our efforts. For each delegated task, reported data would be collected and placed into a central location and format. This helped to facilitate consistent sharing of findings among group members. Many of the quantitative measures used were not predicated on exact numbers; for example, a *range* of difficulties were looked for in items across a certain concept, or the *most common* responses for open ended items were considered for distractors. Visual cues were used in spreadsheets to identify such patterns in data. When looking for ranges of difficulties in items, values above and below the acceptable ranges would be colored red and green, respectively. When looking for common responses in open-ended items, bars were generated within columns of the spreadsheet to identify trends quickly. Incorporating visual elements helped to efficiently recognize tendencies in performance and indicate any areas that could be looked into further.

**Practical Considerations throughout the Development Process**

Having a dedicated team with a variety of competencies is paramount to developing a concept inventory. There must be a member who understands test construction and implementation, someone with considerable expertise in statistics, and multiple domain experts. An advisory board is desirable, and like the team creating the EGCI, the advisory board must be familiar with test construction, statistics, and the domain being assessed. That said, not every member of the advisory board needs to be a domain expert. Having input from those familiar with, but not an expert in, the domain of the CI is invaluable in providing outside opinions because of how easy it is for the development team to get tunnel vision. Some of the most insightful feedback that led to major advances in developing the EGCI was from individuals on the advisory board who were not experts in engineering graphics and asked very pointed questions that challenged the assumptions of the researchers. In this project, there were three face-to-face meetings of the advisory board and the research team.

An awareness of how long the process will take is also important. A fully developed, normed, and validated instrument can take years to create. Currently the team has spent over five years in developing the EGCI, beginning with the initial meeting of engineering graphics experts on the Delphi Study, to the currently ongoing testing of the gamma version. All team members must be willing to commit necessary time to the project including: writing and editing test items; scheduling class time for their students to take the CI, or finding volunteers willing to use the CI in their classes; collecting, analyzing, and interpreting data; scheduling time for conference calls; and meeting face-to-face. Weekly phone meetings were found to be the best way to keep everyone on track. When the meetings were further apart, it became too easy to put off the assigned tasks. The willingness to arrange personal and work schedules to accommodate occasional travel is required too.

The questions for the EGCI were primarily developed by individual research team members. During the development and testing process, many of these questions were discarded or significantly revised. Team members must not become too personally attached to a given question, either one they wrote as a completely new item, or a “favorite” that is based on what they teach and use for assessment in their classes, because if it does not meet the criteria for difficulty and discrimination, it must be removed from the final version.

One of the facets of item creation that took a large amount of time was getting all of the graphics correct according to standards in the graphics community. Again, the expert blind spot often had an impact. For example, if a drawing view was to have hidden lines, the experts would assume all the hidden lines were there, and/or that they were the correct line type, and several revisions of the item would occur before someone would point out an error in the hidden lines that had been there all along. It is also important that those who are helping create any graphics have an understanding of what is correct and what is not according to standard, so they do not accidentally introduce errors into the process.

Efficient sharing of data amongst the team members is especially important. The team on this project used cloud storage so everyone could access the data from nearly any location. The shared data should also have consistent labeling and be stored according to a data management plan, that encompasses both the internal data, and data collected externally, i.e. data from testing the CI on students at different universities. The numbering of test items on the CI must remain consistent from the beginning so that, for example, item 37 is always item 37, even if items 36 and 38 have been removed from the set, so the statistical analyses are comparable over time.

Maintaining artifacts from the collection of data, both on the cloud and in printed form is necessary. In both face-to-face and phone meetings, the team found it useful to have printed booklets with the EGCI items, statistics, and other relevant data in front of them to mark up and discuss when making decisions about item development.

A plan for housing and dissemination of the final CI needs to be confirmed early in the development process, along with alternate plans as necessary. How to protect the integrity and privacy of the collected data also needs to be considered if the CI will be made available on the web or on the cloud. Questions such as who can use the CI, how it is to be accessed, and how student data is to be protected need to be answered.

**Conclusions**

Creating a valid and reliable concept inventory is not easy. Development, testing, and revisions will take multiple iterations. The team responsible for the development and testing of the EGCI learned valuable lessons along the way. Of paramount importance is staying organized and staying on task. Because there are only a few times during the year when a draft instrument can be tested (end of semester) it is important to stay on task so that a window of opportunity is not missed. The development of the EGCI is currently undergoing further statistical testing and will hopefully be available for widespread use in the near future. It has been a long, sometime frustrating, but worthwhile journey.

The authors would like to acknowledge the support of NSF grants 1432280 and 1432288.

**References**


[1] R. F. DeVellis, *Scale development: Theory and applications*. Newbury Park, Calif.: Sage Publications, 1991.

[2] R. G. Netemeyer, W. O. Bearden. & S. Sharma, *Scaling procedures: Issues and applications.* Thousand Oaks, CA: Sage Publications, 2003.

[3] J. C. Nunnally and I. H. Bernstein. *Psychometric theory*. New York: McGraw-Hill, 1994.

[4] G. A. Churchill. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64-73, 1979.

[5] L. A. Clark, & D. Watson, Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-19, 1995.

[6] P. C. Kendall, J. N. Butcher, and G. N. Holmbeck, *Handbook of research methods in clinical psychology*. New York: Wiley, 1999.

[7] P. E. Spector, *Summated rating scale construction: An introduction*. Thousand Oaks, CA: Sage Publications, 1992.

[8] J. Pellegrino, N. Chudowsky, and R. Glaser, *Knowing What Students Know: The Science and Design of Educational Assessment,* National Academy Press, Washington, DC, 2001.

[9] R.A. Streveler, R.L. Miller, A.I. Santiago Román, M.A. Nelson, M.R. Geist, and B.M. Olds, A Rigorous Methodology for Concept Inventory Development: Using the 'Assessment Triangle' to Develop and Test the Thermal and Transport Concept Inventory (TTCI), *International Journal of Engineering Education*, 27, 5, 968-984, 2011.

[10] M. A. Sadowski, and S. A. Sorby, *Defining Concepts for an Engineering Concept Inventory: A Delphi Study,* Proceedings of the 69th Midyear Meeting of the Engineering Design Graphics Division of ASEE, Normal, IL, pp. 67-72, 2014.