ASEE's 123rd Annual • Conference & Exposition • New Orleans, LA • June 26-29, 2016

New Orleans Jazzed about Engineering Education

Paper ID #14444

Building a Concept Inventory for Numerical Methods: A Chronology

Prof. Autar K. Kaw, University of South Florida

Autar Kaw is a professor of mechanical engineering at the University of South Florida. He is a recipient of the 2012 U.S. Professor of the Year Award from the Council for Advancement and Support of Education and Carnegie Foundation for Advancement of Teaching. The award is the only national program to recognize excellence in undergraduate education.

Professor Kaw received his BE Honors degree in Mechanical Engineering from Birla Institute of Technology and Science (BITS) India in 1981, and his degrees of Ph.D. in 1987 and M.S. in 1984, both in Engineering Mechanics from Clemson University, SC. He joined University of South Florida in 1987.

Professor Kaw's main scholarly interests are in engineering education research, open courseware development, bascule bridge design, fracture mechanics, composite materials, and the state and future of higher education.

Funded by National Science Foundation (2002-16), under Professor Kaw's leadership, he and his colleagues from around the nation have developed, implemented, refined and assessed online resources for an open courseware in Numerical Methods (http://nm.MathForCollege.com). This courseware annually receives 1,000,000+ page views, 1,000,000+ views of the YouTube lectures, and 120,000+ visitors to the ”numerical methods guy” blog.

Professor Kaw has written more than 85 refereed technical papers and his opinion editorials have appeared in the Tampa Bay Times, Tampa Tribune and Chronicle Vitae. His work has been covered/cited/quoted in many media outlets including Chronicle of Higher Education, Inside Higher Education, U.S. Congressional Record, Florida Senate Resolution, ASEE Prism, and Voice of America.

Dr. Yingyan Lou, Arizona State University

Dr. Yingyan Lou is an assistant professor in the Civil, Environmental, and Sustainable Engineering program in the School of Sustainable Engineering and The Built Environment Engineering at Arizona State University. She holds a B.S. and a B.A.Econ degree from Beijing University, and received her M.S. and Ph.D. degrees in Civil and Coastal Engineering from the University of Florida. Before ASU, she worked at the Department of Civil, Construction and Environmental Engineering at the University of Alabama.

Dr. Lou is very passionate about teaching and education research. In her teaching, she always emphasizes not just the ”how” but also the ”why” by providing background information on broader issues of the discipline and insights into theories and procedures. Dr. Lou has introduced active learning technologies (such as Clickers) to engage students more effectively during lectures and in-class examples. She also participated in a dissertation study about active learning in engineering disciplines when teaching at The University of Alabama.

Dr. Andrew Scott, Alabama A&M University

Andrew Scott has been a faculty member with the Department of Electrical Engineering and Computer Science at Alabama A&M University, Huntsville, since 2002. He has a strong background in high-performance scientific computing, including algorithms and numerical analyses on parallel and distributed systems. He has expertise in the following areas: Field Programmable Gate Arrays for reconfigurable computing applications, software development for heterogeneous computing environments, domain decomposition, process mapping and data structuring techniques for distributed platforms, and finite element analysis. He holds both BS and MS degrees in mechanical/aerospace engineering from the University of Missouri, Columbia, and PhD in computer science and engineering from the University of Missouri, Kansas City.

Dr. Ronald L. Miller, Colorado School of Mines

Ronald L. Miller is professor emeritus of chemical engineering at the Colorado School of Mines, where he taught chemical engineering and interdisciplinary courses and conducted engineering education research for 28 years. Miller received three university-wide teaching awards and held a Jenni teaching fellowship at CSM. He 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 published widely in engineering education literature. His research interests include measuring and repairing engineering student misconceptions using well-constructed concept inventories.

Building a Concept Inventory For Numerical Methods: A Chronology

1. Introduction

While 46% of college students take more than six years to graduate or simply drop out[1], and the nation seeks one million additional STEM (Science, Technology, Engineering and Mathematics) graduates[2], the competence of these STEM graduates is still paramount over quantity of graduates in the global competitive market. As much as traditional assessment tools of examinations and projects address the procedural and hopefully higher-order thinking in a particular course, we also need tools to assess the level of conceptual thinking of our students. One such tool is the concept inventory (CI) instrument that allows instructors to not only measure a student's conceptual understanding but also the misconceptions they may have developed. The instrument is typically a multiple-choice question test. The questions focus on critical thinking and logic with little need for memorization or calculation, with the goal of showing students' depth of understanding of the topic.

Developing concept inventories in STEM fields has its roots in the Mechanics Diagnostic Test[3,4] given to students in Introductory Physics courses at Arizona State University. Surprisingly, answers to questions that seemed to be trivial were answered incorrectly by a large number of students. The questions were initially posed as free-response questions and common incorrect answers were used later to develop distractors for its conversion into a multiple-choice test. This Mechanics Diagnostic Test finally evolved into what we now know as the Force Concept Inventory (FCI) [5,6]. FCI data established that students' prior beliefs play a prominent role in science education, and many of the beliefs are even drawn from casual observations. Hence, instruction that does not account for misconceptions would be deficient. This realization created a strong interest in developing concept inventories in other STEM fields. As of 2015, many concept inventories have been developed for engineering courses such as materials, statics, dynamics, fluid mechanics, design logic, thermodynamics, etc[7].

In this paper, we chronicle the development of a concept inventory for a course in "Numerical Methods for Engineers" as part of a current National Science Foundation Transforming Undergraduate Education in Science, Technology, Engineering and Mathematics (NSF TUES) grant[8]. As concept inventories in mathematics courses are limited[9-11], there is little guidance available in the current literature. Moreover, the Numerical Methods course is taught by instructors in various engineering majors who emphasize different topics and approaches in the course. For example, some courses discuss various numerical methods at length, while others may emphasize mathematical modeling of problems from their particular major and solve these models using mathematical packages such as MATLAB. At other universities, Numerical Methods is the course where for the first time a student is introduced to formal computer programming. These constraints make it more difficult to create a single assessment tool that is common to all offerings of Numerical Methods courses. Also, being generally an upper-level course, students are expected to have more complex understanding of physics and mathematics, hence creating an assessment based solely on conceptual understanding is challenging[12]. To address these challenges, we took a deliberate and formal route to develop the CI. This route is explained via a timeline so that readers can themselves follow the intricate process of developing a concept inventory. The process included attending a workshop on CI development; identifying top concepts through subject-matter experts using the Delphi technique[13], developing, assessing and refining individual questions; and testing for reliability and validity of the instrument.

2. Chronology

In September 2013, instructors at three universities received a grant[8] to compare the flipped and blended modes of instruction in a Numerical Methods course. As part of the grant, a CI for Numerical Methods[14] was to be developed for nationwide use to measure conceptual understanding of numerical methods.

February 2014 – March 2014: A Concept Inventory Workshop:
An engineering professor, who has thirty years of teaching experience and is a chief developer of concept inventories of three engineering topics[13,15], conducted a holistic workshop on CI development. Three instructors who are investigators of the grant and two external members of the evaluation team attended the workshop. The four-hour workshop was administered via two online sessions.

In the first online session, a timeline was reviewed to develop the CI. The purpose of the workshop was to:

1) identify key concepts and important misconceptions in the domain of numerical methods,
2) review steps required to develop a valid and reliable concept inventory,
3) write reliable and valid items for each concept, and
4) decide how to collect and analyze pilot data to measure effectiveness of inventory items (questions and distractors).

The discussion in the workshop involved the definition of a concept, why we should measure conceptual understanding and how it can be measured. A few sample numerical methods concepts were brainstormed. The framework of developing the CI using the assessment triangle[16] was discussed and this would form the basis of the steps needed to produce a high quality concept inventory.

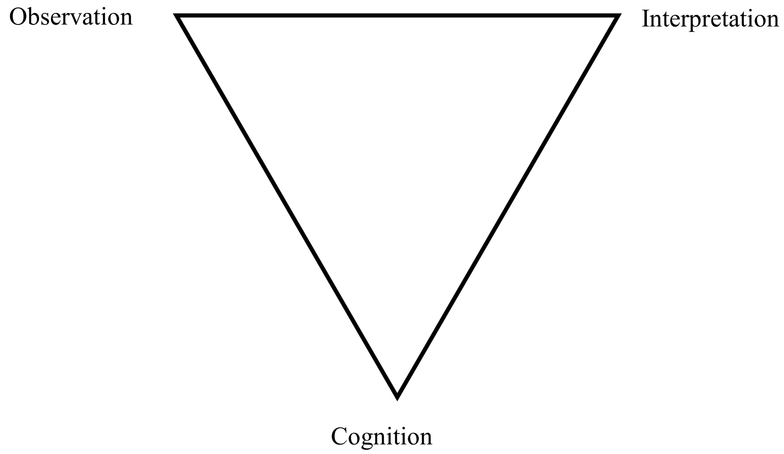

Figure 1. The assessment triangle of cognition, observation, and interpretation[16]

The assessment triangle consists of three interconnected elements—cognition, observation and interpretation (See Figure 1). The *cognition* corner accounts for how students learn about the course, and it is here that one would contemplate the misconceptions students develop about the subject matter. The *observation* corner represents the tasks that will make the assessment. The *interpretation* corner is how we interpret the results of the assessment tasks. These three corners also need to be aligned—how students are learning, the tasks making the assessment, and how the assessment results are analyzed[13,16]. As we present the chronology, these three corners related to Numerical Methods will be discussed accordingly.

Five weeks later, in the second online session, the participants reviewed the key concepts and important misconceptions in the domain of numerical methods. Steps were reviewed to develop a valid reliable concept inventory. The development of CI items for each concept using the Delphi process[17] was discussed. Methods to collect and analyze pilot data to measure effectiveness of inventory items were decided. Based on the assigned readings[13,18] from the first session, the three corners in the assessment triangle were reviewed. The second session focused mainly on the observation vertex of the assessment triangle. The writing of the CI items was discussed, that is, how to write the stem part of the question, the correct answers, and the distractors. As an exercise, one concept was picked for each instructor participant to draft a couple of CI questions. A timeline was discussed for the development of the CI.

March 2014 – June 2014: The Cognition Corner: The Delphi Process to Elicit Misconceptions in Numerical Methods

In the cognition corner, the authors needed to investigate the misconceptions students may have and why these misconceptions may exist in Numerical Methods. There is almost no literature on the latter. As far as what misconceptions students may have, the authors themselves could have come up with a list[14,19], but such an approach has its drawbacks. Streveler et al[13] enumerated these drawbacks: 1) trusting a single expert will be biased, 2) calculating a group average to choose misconceptions does not include an attentive and deliberate input, and 3) a round-table discussion may be unduly influenced by a few of the discussers. For these reasons, we adopted the Delphi technique to elicit the concepts from subject matter experts as follows.

The PI invited numerical methods instructors around the nation from different engineering majors and with varied experience to join a team that would participate in a Delphi methodology to identify the 5 to 10 most important concepts in Numerical Methods. This process has been used since the 1960s as a rational and structured method to develop a consensus of ranking a list of items (in our case, student misconceptions of Numerical Methods concepts). In addition to the three PI instructors, ten additional highly qualified instructors accepted the invitation.

During the process, the 13 participants were asked for input in four rounds. The data were collected independently by the CI expert.

- Round 0 – Participants were requested to generate a list of important concepts and common student misconceptions in numerical methods. The list had 135 items but it was because one participant delineated 80 misconceptions on every topic and every method in a Numerical Methods course as opposed to universal concepts. We were able to cluster the lists from expert participants into nine concepts given below.

1. *Demonstrate the deep relationship of Taylor series to numerical methods*, such as derivation of methods, error analysis, and order of accuracy.

2. *Depict, interpret, and transform numerical methods to and from various forms*, such as graphical, pseudo code, and mathematical equation representations.

3. *Selection of relevant data points when solving discrete data problems*, such as numerical differentiation, interpolation, and integration.

4. *Ability to monitor, establish and interpret convergence of numerical methods*, such as understanding pre-specified tolerance, iterations, and step sizes.

5. *Ability to convert a numerical methods problem from a traditional mathematical model into a format suitable for use in an algorithm* in problems such as coupled ODEs, matrix representation of equations such as ordinary differential equations, simultaneous linear equations, and nonlinear equations.

6. *Interpret the concept of solution accuracy* and its relationship to computer architecture, significant digits, and error propagation.

7. *Identify all possible solutions or the lack thereof for numerical models*. This includes questions such as what methods to use, which mathematical procedure it falls under, if the problem can be solved at all with what we know as an undergraduate.

8. *Knowing when to invoke a numerical method and which technique is most appropriate*. For example, do we need a numerical method and which one? Would an analytical solution suffice?

9. *Ability to identify and specify the correct form of the desired solution*, such as discrete function table, plots, continuous functions, or just a discrete value.

In what we called as Round 0.5, we sought feedback again from the 13 participants to make sure that we were posing the concepts properly as well as not having left anything out. However, for a test that would be given in a typical 50 minute session, one would not be able to test all nine concepts. Assuming a need for a minimum of 3 to 4 questions per concept as is the case for several successful national CIs[19], we settled on having a maximum of six concepts in the CI.

- Round 1 – Participants were asked to rate each of the nine concepts using two rating criteria on a 0-10 scale:
    - Importance – how important is it that students understand this concept?
    - Difficulty – how many of your students do not understand this concept?
- Round 2 – Using anonymous results from Round 1 (mean, median, standard deviation, and range for each rated concept by middle 50% of participant, also known as the 50% interquartile range), participants re-ranked the concepts on the same two criteria while

writing short justifications for any ratings that deviated by more than one standard deviation from Round 1 results.

- Round 3 – Using anonymous results from Round 2 (mean, median, standard deviation, 50% interquartile range, and justification statements), the participants re-ranked each concept again. Again, they provided a justification for any ranking that deviated by more than one standard deviation from the Round 2 results.

All rounds were conducted anonymously by the CI expert and it took four rounds of ranking and discussion to come up with the top six concepts. The results of the Numerical Methods Delphi study are given in Table 1.

Table 1. Results of numerical methods Delphi study. (*The italicized concepts are the ones that were selected after three rounds of voting*)

		Understanding: Median and 50% interquartile range			**Importance: Median and 50% interquartile range**		
	Concept	**Round 1**	**Round 2**	**Round 3**	**Round 1**	**Round 2**	**Round 3**
1	*Taylor series*	4.0 (4.0-5.5)	4.0 (4.0-4.5)	4.0 (4.0-4.0)	8.0 (7.0-9.3)	8.0 (8.0-9.0)	8.5 (8.0-9.0)
2	*Numerical method forms*	5.5 (5.0-6.0)	5.0 (5.0-6.0)	5.0 (5.0-5.0)	8.0 (6.8-9.0)	8.0 (7.0-8.5)	8.0 (7.0-9.0)
3	*Use of relevant data*	6.0 (6.0-7.0)	6.0 (6.0-7.0)	7.0 (6.0-7.0)	7.5 (6.0-8.0)	8.0 (7.0-8.0)	7.5 (7.0-8.0)
4	*Convergence of methods*	5.5 (4.8-6.3)	5.5 (5.0-6.0)	5.0 (5.0-6.0)	8.5 (7.0-9.3)	9.0 (8.0-9.0)	8.5 (8.0-9.0)
5	*Mathematical model to numerical model*	5.0 (4.8-7.0)	5.0 (5.0-5.5)	5.0 (5.0-5.0)	8.0 (8.0-9.0)	8.0 (8.0-9.0)	8.5 (8.0-9.0)
6	*Solution accuracy*	4.0 (3.8-6.0)	5.0 (4.0-6.0)	4.0 (4.0-5.0)	7.5 (7.0-8.3)	7.0 (7.0-8.0)	7.0 (7.0-8.0)
7	*All possible solutions*	4.5 (4.0-5.0)	4.0 (4.0-5.0)	4.0 (4.0-5.0)	8.0 (7.8-9.0)	8.0 (8.0-9.0)	8.0 (8.0-9.0)
8	*When to invoke numerical method*	5.0 (4.8-5.3)	5.0 (5.0-5.0)	5.0 (5.0-5.0)	9.0 (8.0-10)	9.0 (8.0-9.5)	9.0 (8.0-9.0)
9	*Correct solution form*	5.5 (4.0-6.0)	5.0 (4.0-6.0)	5.0 (4.0-6.0)	8.0 (7.0-10)	8.0 (7.0-9.5)	8.0 (7.0-9.0)

Six concepts were chosen on the basis of being both poorly understood by students but also highly important concepts as indicated by the subject-matter experts. The final six concepts are given below and are called Concepts A through F for reference in this paper.

Concept A: *Demonstrate the deep relationship of Taylor series to numerical methods* such as derivation of methods, error analysis, and order of accuracy.

Concept B: *Depict, interpret, and transform numerical methods to and from various forms* such as graphical, pseudo code, and mathematical equation representations.

Concept C: *Ability to monitor, establish and interpret convergence of numerical methods* such as understanding pre-specified tolerance, iterations, and step sizes.

Concept D: *Ability to convert a numerical methods problem from a traditional mathematical model into a format* suitable for use in an algorithm in problems such as coupled ODEs, matrix representation of equations such as ordinary differential equations, simultaneous linear equations, and nonlinear equations.

Concept E: *Identify all possible solutions or lack thereof for numerical models*. What methods to use, which mathematical procedure does it fall under, can the problem be solved at all with what we know as an undergraduate?

Concept F: *Knowing when to invoke a numerical method and which technique is most appropriate*. Do we need a numerical method and which one? Would an analytical solution suffice? Does an analytical solution exist?

June 2014 – November 2014: The Observation Corner

The three instructor PIs from 3 different universities (University X – a large urban university in the Southeast USA, University Y – a large urban university in the Southwest USA, and University Z – a historically black university in the Southeast USA) developed stems of questions for each of the six concepts. Coupled with questions that aligned with the six concepts from a concept test developed earlier by the PI at University X[14], we drafted 32 questions – with at least five questions for each of the six concepts.

Table 2. Distribution of questions amongst concept categories and open-ended and multiple-choice types.

Concept	**A**	**B**	**C**	**D**	**E**	**F**	**Total**
Number of questions	5	6	6	5	5	5	32
Open-ended questions for University X	3	3	5	2	0	1	14
Multiple choice questions for University X	2	3	1	3	5	4	18

The questions developed were read and answered in a talk-aloud format separately by four students—two undergraduate teaching assistants for the course and two students who had recently completed a Numerical Methods course at University X. Changes were made on the wording of some of the questions based on their feedback.

Out of the 32 questions, 14 were written as open-ended questions to gather student responses for distractors and the other 18 were written as multiple-choice questions since these would have otherwise been deemed as ambiguous (Table 2). Because of the concern that suitable distractors may not develop because of the open-ended nature of the 14 questions, all questions at University Y were posed as multiple-choice based on correct answers and best distractors chosen through the expertise and experience of the three instructor PIs, and previous concept test questions developed by the first author[14].

November 2014 – March 2015: The Observation Corner

The 32 questions drafted would have been too long a test to give in a 50-minute period, hence, two tests of 16 questions each at University X (a large urban university in the Southeast) and two tests of 15 questions[a] each at University Y (a large urban university in the Southwest) were

---

[a] Two questions were not asked at University Y because the instructor believed that they were out of scope with the current syllabus.

made. Data were gathered in Fall 2014 at University X and University Y. The course was not taught in Fall 2014 at University Z.

Student responses were collated, and the point-biserial correlation coefficients (PBCC) values (a measure of item reliability calculated by the correlation between a student's score on the item being right or wrong and student's total score on the test as a whole)[20] and the difficulty index (DI) values (percentage of test takers who answer a question correctly) were calculated for each question. A PBCC value of 0.2 or greater and a DI between 30 and 90 were used as criteria to accept a question for further consideration[21]. Questions where more than one distractor was chosen by less than 5% of the students were also noted. A summary of the results is given in Table 3.

Table 3. Results from first draft of concept inventory

University	Number of questions	Number of open-ended questions	Number of multiple-choice questions	Number of questions meeting DI limits	Number of questions meeting PBCC limits	Number of questions meeting both DI and PBCC limits	Cronbach alpha	Multiple choice questions with more than one distractor chosen by less than 5%
X	16	8	8	12	9	9	negative	3
X	16	6	10	13	10	9	0.34	2
Y	15	0	15	14	14	12	0.50	1
Y	15	0	15	13	11	12	0.54	2

This analysis helped us to refine the inventory by identifying questions that were acceptable as is, those that needed revision of stem and/or distractors, and those that were outright inadequate. The negative Cronbach alpha[20] for one of the two tests given at University X was noted, and this is due to several reasons - eight questions were open-ended, seven questions did not meet either or both of the criteria, and for three multiple-choice questions, two distractors were chosen by less than 5% of the students. The concept inventory along with the statistical data was then reviewed by the CI expert to make sure that we were following the correct process. The CI expert also gave us feedback on individual questions, such as to toss, revise, or accept as is.

A summary of how each question progressed to the next stage is given in Table 4 for Concept A.

Table 4. Action take to revise the first draft of the CI for Concept A

Question	**Univ X Fall 2014**		**Univ Y Fall 2014**		**Open-ended at X**	**Action taken**
	DI	PBCC	DI	PBCC		
1	37	0.45	43	0.51	Yes	The question was kept.
2	19	-0.20	23	0.23	No	The stem of the question was revised to remove ambiguity.
3	70	-0.22	75	0.14	No	The question was tossed because of low PBCC.
4	59	0.24	68	0.58	Yes	One distractor was changed and the question was kept.
5	62	0.37	57	0.39	Yes	The question was kept.

We describe the progression in Table 4 to illustrate the granular thought process that went into consideration of each question. The questions themselves are not shown for purposes of maintaining integrity of the test.

The actions taken (keep/minor edit, revise, and toss) for questions for each concept are summarized in Table 5. A *revise* means that a question did not meet the DI and/or PBCC criteria at one of the institutions, and that a revision of the stem and/or distractors was worth another try. Based on the answers of the 14 open-ended questions at University X, distractors were revised if they were used more often by students than some other distractors at University Y.

Table 5. Distribution of decision to keep, revise or toss questions

Concept category	**A**	**B**	**C**	**D**	**E**	**F**	**Total**
Questions	5	6	6	5	5	5	32
Keep/minor edit	3	6	3	2	4	2	20
Revise	1	0	3	2	1	2	9
Toss	1	0	0	1	0	1	3

April 2015 – August 2015: The observation corner continued

Two second draft versions of the CI were given to students in University X in Spring 2015. Two tests were made to keep the number of questions limited to 18. University Y did not offer the course in this semester.

The main goal of giving the CI test was to collect data on questions that were reworded, had new distractors, or were new. The tests were not balanced between concepts as the goal was to test reworded and new questions. However, University Z was requested to make their own version of the CI by choosing three questions per concept from the keep/edit and the revise categories of Table 5. With only 14 students taking the course at University Z, we skipped analyzing any data collected there at the item level.

We show Concept D as an example to illustrate the process (Table 6). For Concept D, out of the five questions from Fall 2014, two were to be kept, two could use revision, and one was tossed. In Spring 2015, we introduced two new questions to possibly have at least 5 questions to choose from for the potential final draft of the CI in Fall 2015.

Table 6. Results for Concept D in second draft after having taken action on first draft of CI

Question	**Univ X Fall 2014**		**Open-ended Fall 2014**	**Univ X Spring 2015**		**Action taken**
	DI	PBCC		DI	PBCC	
1	69	0.35	No	-	-	This question was not asked again as it was acceptable in Fall 2014.
2	44	0.27	Yes	76	0.37	Distractors were revised based on the open-ended responses of Fall 2014 at University X and it meets the DI and PBCC criteria.
3	-	-	-	-	-	This question was limited to the syllabus at University Z and has since been tossed. It did not meet the PBCC criterion anyway at University Z.
4	7	0.00	Yes	49	0.34	This question created confusion as an open-ended type in Fall 2014. It was now asked with multiple choices.
5	79	0.09	No	-	-	This question was tossed in Fall 2014 because of low PBCC.
6	-	-	-	39	0.37	A new multiple-choice question was introduced in both versions and it meets the DI and PBCC criteria.
7	-	-	-	8	0.34	A new multiple-choice question was introduced in both versions and it did not meet the DI criterion. This question was tried again with revisions in Fall 2015.

November 2015 – January 2016: The interpretation corner

Based on the statistical results of the first and second drafts of the CI tested in Fall 2014 and Spring 2015, a common 24-multiple-choice-question CI was tested at University X and University Y. The course was not offered in Fall 2015 at University Z. For each of the six concepts, four questions that met the DI and PBCC criteria (except 1 question as that was the last best one available in that concept category) in the most recent semester were used. If more than four questions were available for a concept, ones with DI closest to 62.5 and a high PBCC were chosen.

Having a common test now provided a larger sample size of a total of 130 students. The PBCC and DI for each question and the Cronbach alpha for measuring the reliability of the CI were calculated, and a summary is shown in Table 7.

Table 7. Results from third draft of CI

University	Number of Questions	Questions meeting DI criterion	Questions meeting PBCC criterion	Questions meeting DI and PBCC criteria	Cronbach alpha	Questions with more than one distractor chosen by less than 5%
X & Y	24	23	20	20	0.4725	0

None of the questions had more than one distractor which was chosen by less than 5% of the students. Figure 2 shows the difficulty index and PBCC value of each of the 24 questions in a scatter plot.

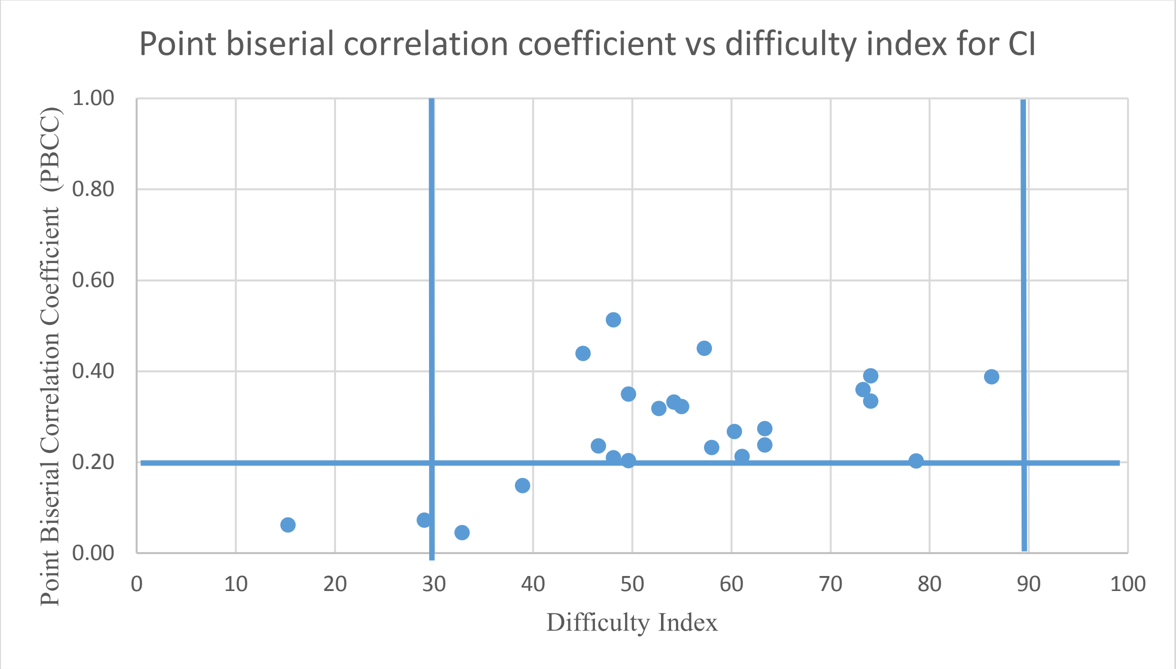

Figure 2. Scatter plot showing point-biserial correlation coefficient and difficulty index for third draft of CI

It was hoped that the 24-question CI draft would be the final version, but that did not turn out to be the case. Four of the questions (2A, 16D, 21F, 22F) highlighted in Table 8 did not meet one or both of the DI and the PBCC criteria (Figure 2), although three of them had met the criteria at least at University X or University Y in the Fall 2014 and Spring 2015 semesters.

We had expected that at least 3 out of 4 questions in each concept would meet the two criteria and hence, still create a possible final version of the CI. The two criteria were met for 3 questions/concept for five of the six concepts (Concept F had only two acceptable questions).

The Cronbach alpha for the overall test is 0.4725 and is considered to be a poor value while 0.7-0.9 is desirable[22]. The low PBCC (Figure 2) of several questions (although meeting the lower threshold of 0.2) may be the reason for the low value. To confirm this, the Cronbach alphas were calculated with each item score deleted; these alphas are called alpha-with-item-deleted. Seven questions (2A, 7B, 12C, 16D, 21F, 22F, 23F) had a higher alpha-with-item-deleted than the overall Cronbach alpha. It is deemed that if the alpha-with-item-deleted is equal to or greater than overall Cronbach alpha, then that item may be measuring a concept different from the other items[21]. The seven questions included the four that did not meet the two criteria of DI and PBCC. The other three questions (7B, 12C, 23F) that did not meet the alpha-with-item-deleted criterion had a PBCC in the 0.20-0.21 range, which is just above the lower threshold used of 0.20. To ensure a Cronbach alpha for the whole CI that is at least average, we will increase the lower threshold of PBCC to 0.3 as a criterion for acceptance for all questions.

The work on the CI will continue in Spring 2016 and beyond with the emphasis on writing new questions that will each yield a PBCC>0.3 in the needed concept categories.

Table 8. Classic test theory results of third draft CI (Cronbach alpha of whole CI =0.4725; *highlighted entries of the table show the questions and reasons for concern*)

Question No	**Concept**	**Difficulty Index**	**PBCC**	**Alpha with item deleted**
1	A	50	0.35	0.4486
2	A	33	0.05	0.5019
3	A	57	0.45	0.4269
4	A	60	0.27	0.4646
5	B	45	0.44	0.4293
6	B	73	0.36	0.4453
7	B	48	0.21	0.4766
8	B	74	0.39	0.4396
9	C	55	0.32	0.4543
10	C	63	0.27	0.4629
11	C	63	0.24	0.4697
12	C	50	0.20	0.4778
13	D	74	0.33	0.4498
14	D	54	0.33	0.4522
15	D	53	0.32	0.4551
16	D	15	0.06	0.4867
17	E	79	0.20	0.4707
18	E	58	0.23	0.4717
19	E	86	0.39	0.4418
20	E	47	0.24	0.4715
21	F	29	0.07	0.4955
22	F	39	0.15	0.4866
23	F	61	0.21	0.4749
24	F	48	0.51	0.4124

Conclusions

A concept inventory for a course in Numerical Methods is being developed. The Delphi process was followed in generating six concepts in which students have the most misconceptions but are most important as well. Questions were generated for each of the six concepts. Going through three drafts of the CI at three universities over three semesters, and performing a rigorous analysis is bringing us closer to finalizing the concept inventory. Our quest will continue in Spring 2016 where current questions that may need revisions, and more importantly, new questions will be tested to see if they meet the core requirements of the DI and PBCC. The PBCC threshold will be increased to 0.3 for all questions to warrant an acceptable reliability.

The authors will continue to follow the rigorous analytical framework enumerated by Jorion et al[21] to be able to make the claim of a valid concept inventory. In addition to the classical test theory (DI and PBCC criteria, Cronbach alpha of the entire CI, and alpha-with-item-deleted) used in this paper, the framework will also include predictive validity by correlating CI scores to final examination scores, item response theory analysis, structural analysis, and confirmatory factor analysis[21].

Acknowledgment

This material is based upon work supported partially by the National Science Foundation under Grant Number 1322586, and the Research for Undergraduates Program in the College of Engineering at University of South Florida (USF). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the 10 instructors who diligently participated in the Delphi study and returned their responses promptly. We wish to thank USF undergraduate assistants Benjamin Rigsby and Humberto Isaza for compiling the concept inventory data.

References

1. O'Shaughnessy, L., "How Many College Students Graduate?" Moneywatch, December 17, 2013, http://www.cbsnews.com/news/how-many-college-students-graduate/, accessed March 15, 2016.

2. "Report to President: Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics," https://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf, accessed March 15, 2016.

3. Halloun, I., and Hestenes, D., "The Initial Knowledge State of College Physics Students", American Journal of Physics, 53, 1043–1055, 1985.

4. Halloun, I., and Hestenes, D., "Common Sense Concepts about Motion", American Journal of Physics, 53, 1056–1065, 1985.

5. Hestenes, D., Wells, M., and Swackhamer, G., "Force Concept Inventory", The Physics Teacher, 30, 141–158, 1992.

6. Hestenes, D., and Halloun, I., "Interpreting the FCI", The Physics Teacher. 33, 502–506, 1995.

7. Concept Inventory Assessment Instruments for Engineering Science, http://www.foundationcoalition.org/home/keycomponents/assessment_evaluation.html, accessed March 15, 2016.

8. Kaw, A., Besterfield-Sacre, M., Scott, A., Pendyala, R., Lou, Y., "Improving and Assessing Student Learning in an Inverted STEM Classroom Setting" NSF TUES-Type 2 Project, https://www.nsf.gov/awardsearch/showAward?AWD_ID=1322586, accessed March 15, 2016.

9. Epstein, J., "The Calculus Concept Inventory," in National STEM Assessment, Washington, D.C., 60–67, 2006, accessed January 4, 2013, from http://www.openwatermedia.com/downloads/STEM(for-posting).pdf#page=64.

10. Epstein, J., "Development and Validation of the Calculus Concept Inventory," Proceedings of the Ninth International Conference on Mathematics Education in a Global Community, Palermo, Italy, 2007.

11. Allen, K., The Statistics Concept Inventory: Development and analysis of a cognitive assessment instrument in statistics (Doctoral dissertation), SSRN Electronic Journal, 2006, doi:10.2139/ssrn.2130143

12. Wilcox, B., Caballero, M., Baily, C., Sadaghiani, H., Chasteen, S., Ryan, Q., and Pollock, S., “Development and uses of upper-division conceptual assessments”, Phys. Rev. ST Phys. Educ. Res. 11, 020115 – Published 23 September 2015, http://journals.aps.org/prstper/pdf/10.1103/PhysRevSTPER.11.020115

13. Streveler, R., Miller, R., Santiago-Roman, A., Nelson, M., Geist, M., and Olds, B., “Rigorous Methodology for Concept Inventory Development: Using the 'Assessment Triangle' to Develop and Test the Thermal and Transport Science Concept Inventory (TTCI)”, International Journal of Engineering Education 27(5), 968-984, 2011.

14. Kaw, A., and Yalcin, A., “Measuring Student Learning Using Initial and Final Concept Test in a STEM Course”, International Journal of Mathematical Education in Science and Technology, 43(4), 435-448, 2012.

15. Miller, R., Streveler, R., Yang, D., and Santiago-Román, A., “Identifying and Repairing Student Misconceptions in Thermal and Transport Science: Concept Inventories and Schema Training Studies”, Chemical Engineering Education, 45(3), 203-210, 2011.

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

17. Streveler, R. A., Olds, B., Miller, R. L., and Nelson, M. A. (2003). “Using A Delphi Study To Identify The Most Difficult Concepts For Students To Master In Thermal And Transport Science”, In Proceedings of the 2003 ASEE Annual Conference and Exposition. Retrieved from https://peer.asee.org/using-a-delphi-study-to-identify-the-most-difficult-concepts-for-studentsto-master-in-thermal-and-transport-sciences.

18. Richardson, J., “Concept Inventories: Tools for Uncovering STEM Students’ Misconceptions”, Assessment and Education Research, 19-25, 2004, http://www.aaas.org/sites/default/files/02_AER_Richardson.pdf, accessed March 15, 2016.

19. Steif, P.S., and Dantzler, J.A. “A Statics Concept Inventory: Development and Psychometric Analysis”, Journal of Engineering Education, 96(4), 363-371, 2005.

20. Brown, J. D., Understanding Research in Second Language Learning: A Teacher's Guide to Statistics and Research Design. Cambridge: Cambridge University Press, 1988.

21. Jorion, N., Gane, B., James, K., Schroeder, L., DiBello, L., and Pellegrino, J., “An Analytic Framework for Evaluating the Validity of Concept Inventory Claims”, Journal of Engineering Education, 104, 454–496, 2015, doi:10.1002/jee.20104.

22. Wells, C. and Wollack, J., (2003), “An Instructor’s Guide to Understanding Test Reliability”, University of Wisconsin, WI, https://testing.wisc.edu/Reliability.pdf, accessed March 15, 2016.