Development of a Concept Inventory for the Nursing General, Organic and Biochemistry Course

Anne P. Vonderheide* Cijy Elizabeth Sunny Kathleen Koenig
University of Cincinnati
* Corresponding Author

Abstract

Many disciplines have used concept inventories (CI) to better understand the alternative ideas held by students. Administering a CI at both the beginning and the end of the semester can yield much information. For example, a comparison of both pre- and post-test responses can help determine if specific gains in student learning were met. In this paper, we describe the development of a concept inventory over four academic years for the one-semester general, organic, and biochemistry (GOB) course required for nursing students. Questions were designed around course topics and to identify common misconceptions, which were gathered from the literature, student interviews, and content experts. The instrument was refined over several iterations to ensure that all questions were appropriate and understandable while providing useful information. The final version demonstrated appropriate gains in the comparison of pre- and post-test results for an active learning classroom and results are discussed. In addition, item quality was ascertained via the discrimination and difficulty values, and Cronbach's alpha coefficient was computed for the entire test as a measure of its reliability.

Keywords

Concept inventory, General, Organic and Biochemistry, Nursing, Undergraduate, One-Semester

Introduction

College educators have made great strides in the delivery of effective instruction in the preceding decades, with general trends moving from lecture-based approaches to those involving active learning. Many studies have shown the latter to effectively increase student comprehension as well as support information retention. (Freeman et al., 2014; Ruder & Hunnicutt, 2008; Stockwell, Stockwell, Cennamo, & Jiang, 2015) For example, in the area of chemistry education, Wright and team conducted 30-minute exit interviews with students in an attempt to discern scientific competence and discovered that students from the interactive classroom were consistently ranked higher by independent examiners than those from the traditional lecture classroom.(Wright et al., 1998) For broad comparisons, however, it is important to be able to measure such differences, either temporally or across different teaching methods or student populations, using standardized tools. This can be done using Concept Inventories (CI), which are typically designed as multiple-choice tests to assess student knowledge of a specific set of concepts. CIs may be used as pre- and post-tests to measure gains over an entire course, and thereby show possible shifts in student understanding, or may be used to capture students' level of understanding at an instant of time.

One of the most widely employed concept inventories in the STEM field is the Force Concept Inventory (FCI).(Hestenes, Wells, & Swackhamer, 1992) Through the comparison of studies that used this tool in multiple introductory physics courses at different institutions, Hake was able to discern greater gains in students' understanding of related topics in courses with interactive engagement as opposed to traditionally taught lecture courses. (Hake, 1998) Although physics education researchers have developed a large number of CIs around a range of topics in introductory physics courses, fewer are available in other disciplines. In chemistry, a CI was developed for first year general chemistry in 2002(Mulford & Robinson, 2002) and was later validated by Barbera in 2013.(Barbera, 2013) In addition, Krause and colleagues created a CI in an effort to measure chemistry concepts essential to material engineering courses.(Krause, Birk, Bauer, Jenkins, & Pavelich, 2004) Others have published CIs of more narrow topics, including chemical bonding(Luxford & Bretz, 2014), acid/base chemistry(Undersander, Lund, Langdon, & Stains, 2017), precipitation/dissolution(Abell & Bretz, 2019), and redox reactions.(Jin, Rodriguez, Shah, & Rushton, 2020)

Besides the limited number of available CIs in chemistry, a comprehensive CI does not exist for certain courses. For example, introductory chemistry courses have become more major-specific, and nursing students are generally required to take a one semester course covering pertinent topics in general, organic, and biochemistry.(Brown, Henry, Barbera, & Hyslop, 2013) This makes for a rather fast-paced term covering a wide range of topics.(Ball, Hill, & Scott, 2011) While a few of these topics may be included on one of the general chemistry CIs previously mentioned, there are no instruments related to the organic and biochemistry portions of the course. Further, due to the fast-paced nature of the course, some topics are covered in greater depth in the general chemistry course than in the nursing GOB course, causing some of the existing CI questions to be inappropriate.

Given the lack of a suitable CI for a GOB course taken by nursing students, this paper presents ongoing work in the development of an assessment for a course taught by one of the authors. Although such instruments may serve a range of purposes, this CI was developed to serve as a benchmark by which to measure teaching effectiveness, i.e., whether learning outcomes were met and students were able to overcome apparent misconceptions. The instrument went through two iterations of change, and the questions initially found troublesome are discussed in this paper and subsequent modifications are presented. Finally, the results of two academic years (2018-19 and 2019-20) of administering the revised CI both at the beginning and end of the semester are presented as well as associated gains and psychometric properties obtained via classical test theory.

Development of the GOB CI Questions

The development of the questions to include on the GOB CI began with scrutiny of a list of course topics.(Ball et al., 2011) About half of the chapters are general chemistry, allowing one quarter of the time to be given to each of organic and biochemistry, with the organic topics setting the stage for the biochemistry portion (Table 1). That is, understanding the intermolecular forces and reactions of studied functional groups is translated to the behavior of biological macromolecules. Therefore, the course places heavy emphasis on bond formation and subsequent intermolecular forces. From these, a list of essential concepts was generated by one of the authors of this paper who has extensive experience both in the discipline and practice of chemistry and has taught this course for 7 years.(Treagust, 1988)

Chapter 1 – Chemistry, Matter and Measurement
Chapter 2 – Elements, Atoms and the Periodic Table
Chapter 3 – Ionic Bonding and Simple Ionic Compounds
Chapter 4 – Covalent Bonding and Simple Molecular Compounds
Chapter 5 – Introduction to Chemical Reactions
Chapter 6 – Quantities in Chemical Reactions
Chapter 7 – Energy and Chemical Processes
Chapter 8 – Solids, Liquids and Gases
Chapter 9 – Solutions
Chapter 10 – Acids and Bases
Chapter 11 – Nuclear Chemistry
Chapter 12 – Organic Chemistry: Alkanes and Halogenated Hydrocarbons
Chapter 13 – Unsaturated and Aromatic Hydrocarbons
Chapter 14 – Organic Compounds of Oxygen
Chapter 15 – Organic Acids and Bases and Some of Their Derivatives
Chapter 16 – Carbohydrates
Chapter 17 – Lipids
Chapter 18 – Amino Acids, Proteins and Enzymes
Chapter 19 – Nucleic Acids
Chapter 20 – Energy Metabolism

Table 1. Representative GOB Chapters

Questions to include on the CI were then written based on the expected learning outcomes of each concept. Table 2 presents the concepts targeted in each question. The questions were written as multiple choice for large-scale use, with distractors including common misconceptions obtained from the literature(Abell & Bretz, 2019; Luxford & Bretz, 2014; Taskin, Bernholt, & Parchmann, 2015; Usta & Ayas, 2010), content experts, or former students. Decisions were also made about the level of student understanding to be assessed in each question. For example, no questions were included that asked for functional group identification (organic chemistry section) as such questions would rely on memorization with little reasoning. Rather, questions were designed to carry concepts such as bond formation learned in the general and organic portion of the course and allow that knowledge to serve as a basis for reasoning through biochemical phenomena. Further, although it is impossible to eliminate every instance of science rhetoric when asking science-based questions, great effort was made to keep conceptual settings ordinary in order to ensure the questions were not vocabulary limited.(Smith & Tanner, 2010) In other words, questions were designed using scenarios such as salt on a driveway or the miscibility of water and vinegar rather than more classic textbook examples. The final set of questions were reviewed by content experts for accuracy as well as relevancy to the identified concepts that each question was designed to assess.

Question	Concept
1.	Mass balance, balancing chemical reactions
2.	Nuclear reactions
3.	Acid/base chemistry
4.	Protein denaturation
5.	Solution chemistry
6.	Density
7.	Boiling point
8.	Thermodynamics
9.	Avogadro's number
10.	Intermolecular forces
11.	Ionic bonds
12.	Ionization energy
13.	Bond polarity
14.	RNA/DNA
15.	Lipid Metabolism
16.	Proteins
17.	Colligative properties
18.	Transcription/translation
19.	Intermolecular forces
20.	Dimensional analysis
21.	Ionic and covalent bonds
22.	Dilution
23.	Gas laws

Table 2. The GOB concepts covered in this concept inventory

Experimental Overview

The CI was administered over four academic years: 2016-17, 2017-18, 2018-19 and 2019-20. Due to revisions of some of the questions, the CI version used in each of the first two years differed from the version used in the last two years. Each year the CI was administered to approximately 300 nursing and pre-nursing students, who were enrolled in one of two sections each taught by the same instructor (one of the authors). The CI was administered on paper and students recorded their answers on a scantron sheet. Although unnecessary, a periodic table was on the classroom wall. Further, students were allowed the use of a scientific calculator, although it was not necessary as the questions were conceptual. Students were given 20 minutes (at the end of the class) to complete the CI, with the same version given on the first and the last day of the course. In all cases, all students finished the test in the time given, suggesting that adequate time was provided. Students were encouraged to take the test seriously, with the instructor indicating that results would inform future instruction. Additionally, students were awarded 10 extra credit points at the end of the semester (1000 points in course) for completion of both the pre- and post-test, regardless of the correctness of answers.

Revision of Test Questions

The instrument went through two iterations after initial development and this section discusses the changes made and the reasons behind the revisions. For the 2016-17 academic year, 235 students had matching pre- and post-tests, which were used in the analysis. The CI consisted of 23 multiple choice questions with three possible answers each. Results showed an average pre-test score of 57% and a post-test score of 73%, with a normalized gain of 37% (Table 3). The number of correct responses for in-

Academic Year	Number of matched students	Pre-test score	Post-test score	Normalized Gain
2016-17	235	57%	73%	37%
2017-18	228	54%	72%	39%
2018-19	254	53%	70%	36%
2019-20	254	51%	70%	38%

Table 3. Summary of student performance over four iterations.

dividual students ranged from 7 to 21 on the pre-test and 8 to 22 on the post-test of 23 questions. The individual test scores and average score distributions for this cohort suggested that this CI was a reasonable measure of student understanding with 94% of students earning a score between 12 and 22.

A few outcomes stood out from the use of this initial version of the CI. First, questions 14, 18, and 20 all demonstrated a ceiling effect on both the pre- and post-test, which indicated students had this knowledge when entering the course, and these questions would not be able to provide a measure of student learning.

---

14. A genetic disease is a direct result of
   a. Insufficient protein intake
   b. Exposure to radiation
   c. An incorrect nucleic acid sequence
   d. Aging*

18. DNA contains the code for
   a. Protein primary structure
   b. Organ function
   c. Infection control
   d. Brain regulation*

20. A patient is to receive an infusion of sodium bicarbonate. How long will the 100 mL infusion last if it is running at 12.5 mL per hour?
   a. 2 hours
   b. 8 hours
   c. 24 hours
   d. 125 hours*

---

* Distractor added in second iteration, as discussed below.

Second, question 7 had an unexpected negative shift in that students received a higher score on the pre-test than the post-test (33% correct on the pre-test and 28% correct on the post-test), with "oxygen" the most common answer chosen. Interestingly, prior to the 2016-17 academic year, it was decided to condense the coverage of thermodynamics. Although boiling points and exo/endothermic reactions remain covered in their entirety, the topic of energy for phase transitions and changes in temperature was eliminated from this course. This "floor effect" result on both the pre- and post-test most likely reflects the fact that our coverage of thermodynamics is insufficient to provide students a true understanding of the physical process of boiling.

---

7. When heating water on a stove, one can observe the formation of bubbles at the bottom of the pot as the solution comes to a boil. These bubbles contain
   a. Oxygen
   b. Oxygen and Nitrogen
   c. Water vapor
   d. Nitrogen*

---

* Distractor added in second iteration, as discussed below.

At this time, it was decided that for the next test version the answer choices for all questions would be increased to four, rather than three, to reduce the chance of students guessing the correct answer. This change would also allow for deeper analysis of student understanding and possibly address the ceiling effect observed for questions 14, 18, and 20. A fourth answer choice was subsequently added to each question by one of the authors, and again this distractor included a common misconception obtained from the literature, content experts, or former students.

The second iteration of the CI was given as a pre- and post-test for the 2017-18 academic year to approximately 300 nursing students. A total of 228 students had matched pre- and post-tests and were included in the analyses. Pre- to post-test mean scores of 54% to 72% were observed, yielding a 39% normalized gain. The range of correct answer choices was 5 to 21 on the pre-test and 8 to 23 on the post-test. Like the previous test administration, these total scores and distribution of scores suggest the GOB CI to be a reasonable measure.

Student performance on questions 7, 14, 18, and 20 was like that of the previous academic year, and the addition of a fourth distractor did nothing to address issues of ceiling effect for the latter three questions. Rather than remove these three questions, they were retained to ensure that students did not become bored or disinterested as they took the test. However, there remains the potential to eliminate one or more to shorten the test in the future.

Unfortunately, the addition of a fourth distractor gave issue to two other questions. The first was question 2, which previously had not included answer choice b:

---

2. A nuclear reaction could entail:
   a. Loss of a valence electron
   b. Separation of all protons from neutrons in the nucleus
   c. Changing element identity
   d. A highly endothermic process

When the question had just three answer choices, correct responses on the pre- and post-test increased from 56 to 77%, respectively. However, with the addition of a fourth distractor, the scores from pre- to post-test increased from 24% to 53%, with the most chosen distractor now 'b'. A follow-up of student interviews revealed that some had fixated on the word "nucleus" in this distractor. A few others commented that some protons would be separated from some neutrons in a fission reaction. It was decided that this distractor was vague, and it was replaced with "A shift in equilibrium towards the reactants" for the third iteration of the CI.

A second question that became troublesome with the addition of a fourth distractor was question 12, which had previously not included choice d:

---

12. If cesium has 2 valence electrons and iodine has 7 valence electrons, which element takes more energy to form a positive ion?
   a. Cesium
   b. Iodine
   c. They take equivalent amounts of energy to move from the atomic to the ionic state.
   d. Neither will exist in the ionic state

Although the initial version with three answer choices showed negligible gain, using the revised question, a negative gain resulted with students' pre- to post-test scores on this question dropping from 40% to 31%. Of note, 44% of students selected "cesium" on the pre-test and 63% selected this answer on the post-test. In subsequent student interviews, a large majority indicated that they were not accounting for the "energy" part of the question and were simply answering the question based on which element would be more likely to form a positive ion. Some also mentioned that they felt the question was asking two questions, one about valence electrons/ion formation and one about ionization energy. The question stem was revised for the third iteration to read "If cesium has 2 valence electrons and iodine has 7 valence electrons, which element would be more likely to form a positive ion?".

The slightly revised CI, which included only the described changes to questions 2 and 12, was administered in the 2018-19 academic year. The average correct responses on the pre- and post-test were 53% and 70%, respectively, representing a 36% normalized gain. The range of correct answer choices was 6 to 21 on the pre-test and 8 to 23 on the post-test. These results are similar to past administrations of the CI and no additional anomalies were noted in this iteration. Student performance was similar for question 7 and questions 14, 18, and 20 continued to show a ceiling effect. For the revised question 2, students' pre- to post-test scores improved from 44% to 65%. Likewise, students' scores improved from 51% to 74% for the revised question 12.

Based on this success, the concept inventory from 2018-19 was not altered in any way and was sent to three content experts to establish content validity. The experts noted that the wording of the questions was appropriate and clear and that the distractors represented common student misconceptions. Based on this feedback, an identical version of the CI was administered during the 2019-20 academic year. Average pre- and post-test scores along with gain are provided in Table 3. The range of correct answers was 4 to 21 on the pre-test and 6 to 22 on the post-test. All questions had similar outcomes to those found in 2018-19, with questions 14, 18, and 20 again demonstrating a ceiling effect and question 7 showing no change in student learning (29% on pre-test and 27% on post-test), as expected. For questions 2 and 12 examined in the previous iteration, the 2019-2020 academic year showed mean shifts of 40% to 56% and 45% to 67%, respectively.

Results and Discussion

In considering the overall pre-test scores across the four years, as well as the stability of student responses, it is apparent that the students enter the course with base knowledge and are not just guessing on the pre-test. This is expected given that many students have taken chemistry in high school. Score distributions for the post-test are shown in bar graphs for the four academic years in Figure 1. The averages and standard deviations for each of the academic years are as follows: 2016-17, average 16.8, standard deviation 4.4; 2017-18, average 16.4, standard deviation 4.6; 2018-19, average 16.1, standard deviation 4.9; 2019-20, average 16.1, standard deviation 4.2. The post-test distributions for the final version (administered in the 2018-19 and 2019-20 academic years) demonstrated a normal distribution as the skewness and kurtosis values are within +/- 2 SE. As a first step in analyzing the complete sets of data, the post-test scores were plotted

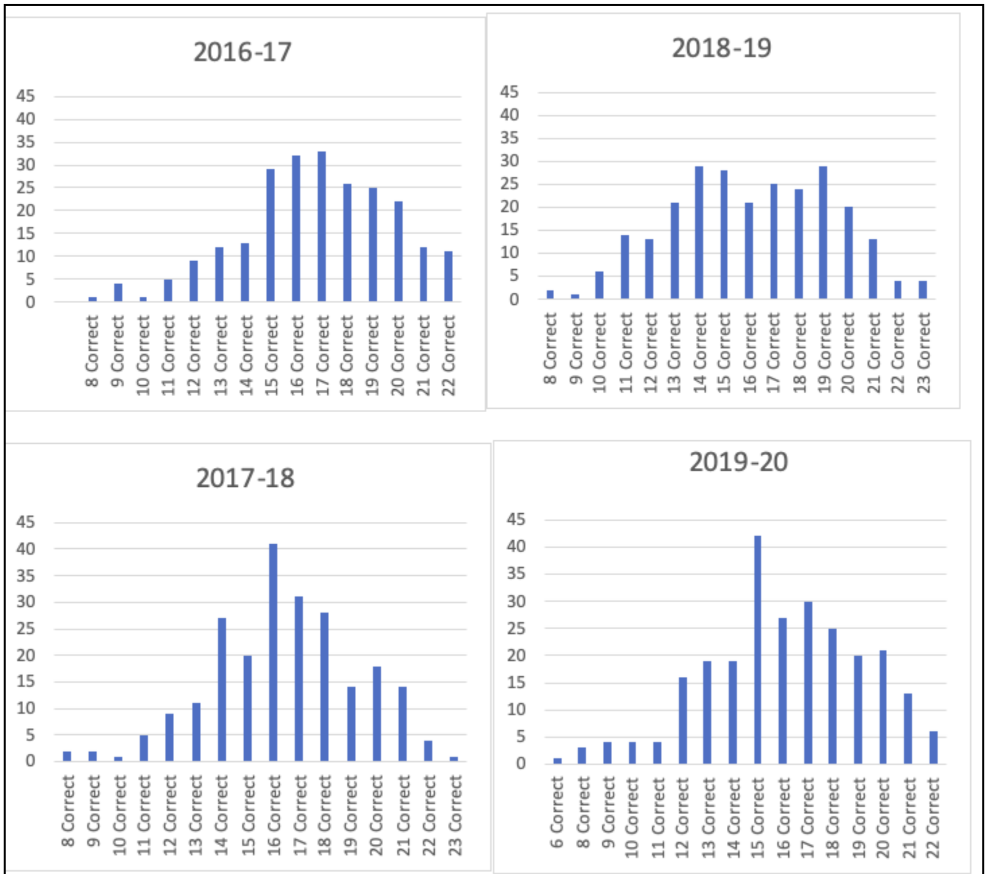


Figure 1. Score Distributions for Post Test Over Four Academic Years. Y-axis represents number of students.

Academic year	N	CI score vs. course grad, R	R^2	Final grade in course	Average Post-test score (standard deviation)
2016-17	235	0.47	0.22		
	34			A	19.3 (1.9)
	80			B	17.2 (3.0)
	78			C	16.7 (2.4)
	25			D	15.1 (2.4)
	18			F	13.8 (3.3)
2017-18	228	0.52	0.27		
	34			A	18.6 (2.3)
	56			B	17.4 (2.6)
	52			C	16.1 (2.2)
	64			D	15.8 (2.4)
	22			F	13.0 (2.7)
2018-19	254	0.53	0.28		
	50			A	18.6 (2.7)
	110			B	16.6 (2.6)
	60			C	14.5 (3.2)
	31			D	14.1 (2.8)
	3			F	11.0 (1.0)
2019-20	254	0.52	0.27		
	56			A	18.5 (2.2)
	88			B	16.2 (2.9)
	75			C	15.2 (2.6)
	28			D	14.2 (2.8)
	7			F	11.3 (3.7)

Table 4. Comparison of performance on CI vs. final grade in course

against the final course grade. It should be noted that exams account for 90% of the final course grade, which include the ACS Final Exam (General-Organic-Biochemistry Exam (Form 2007) of American Chemical Society). Results are shown in Table 4 and suggest a moderate positive correlation between final grade and the CI (Final grade plotted numerically using a 4.0 scale). Generally, weak to no correlation is expected as a concept inventory and

2018-19 Academic year			
Item #	**High Performers**	**Low Performers**	**Item Discrimination**
1	0.76	0.34	0.42
2	0.55	0.36	0.19
3	0.73	0.28	0.46
4	0.53	0.25	0.28
5	0.88	0.45	0.43
6	0.83	0.46	0.37
7	0.47	0.24	0.23
8	0.54	0.06	0.48
9	0.82	0.43	0.39
10	0.47	0.24	0.23
11	0.75	0.42	0.33
12	0.80	0.29	0.51
13	0.43	0.11	0.33
14	1.00	0.84	0.16
15	0.89	0.58	0.31
16	0.76	0.35	0.41
17	0.83	0.22	0.61
18	0.94	0.71	0.23
19	0.49	0.36	0.13
20	0.95	0.73	0.22
21	0.80	0.12	0.67
22	0.72	0.41	0.31
23	0.54	0.17	0.37

2019-20 Academic year			
Item #	**High Performers**	**Low Performers**	**Item Discrimination**
1	0.55	0.28	0.28
2	0.52	0.25	0.27
3	0.47	0.18	0.29
4	0.55	0.19	0.36
5	0.80	0.33	0.47
6	0.73	0.54	0.19
7	0.40	0.27	0.13
8	0.40	0.05	0.35
9	0.86	0.34	0.52
10	0.57	0.23	0.34
11	0.75	0.46	0.29
12	0.67	0.29	0.39
13	0.34	0.13	0.20
14	0.99	0.86	0.13
15	0.93	0.65	0.28
16	0.76	0.41	0.35
17	0.72	0.23	0.49
18	0.96	0.71	0.25
19	0.53	0.25	0.28
20	0.96	0.78	0.18
21	0.63	0.11	0.52
22	0.83	0.35	0.48
23	0.57	0.17	0.40

Table 5. CI Item Discrimination for the 2018-19 and 2019-20 Academic Years

a course exam are very different tests. That is, an exam is announced beforehand, assesses declarative knowledge, and students prepare for it. Compared to a CI, exams are high stakes. A CI may be a better measure of learning gain as it measures understanding versus static knowledge.(Sands, Parker, Hedgeland, Jordan, & Gallowy, 2018)

To state with confidence that the CI measures understanding, analyses must be extended to further statistical tests. As stated previously, content validity was established by content experts. The latter two years of data (2018-19 and 2019-20) were used for statistical analysis as they were both collected using the final instrument. A t-test found a significant difference between the 2018-19 and 2019-20 groups of students for the pre-test data so the two could not be combined for validation. As a result, the psychometric evaluation of the CI using classical test theory was performed separately for each of these cohorts.

The Cronbach alpha value for the 2018-19 data was 0.59 and for the 2019-20 data was 0.55. An alpha value of 0.7 or above generally indicates that the test is reasonably reliable. (Nunnally, 1978; Peterson, 1994). Our values are lower but may not be the best measure of the test's reliability because the CI covers a broad range of concepts, as shown in Table 2, with each generally represented by 1 question(Jin et al., 2020) due to the nature of the course. Without interrelated items that test the same concept, the reliability of the test will be lowered.

Tables of Item Discrimination of the pre-test for the two academic years are given in Table 5 along with a bar graph for both sets of data in Figure 2. Item discrimination is a measure of how well an item is able to distinguish between examinees who are knowledgeable and those who are not. Items that have a discrimination index of more than 0.2 should be kept on the test as these are marginal to reasonably good items.(Önder, 2016; Ozcelik, 1998; Taib & Yusoff, 2014) Items with a discrimination index of less than 0.2 should be reviewed and negative items should be removed. In the 2018-19 data set, questions 2, 14 and 19 showed poor discrimination between the "higher-performing" and "lower-performing" students, however, no negative values were calculated, demonstrating that lower-performing students are not doing better on questions than higher-performers. In the 2019-20 academic year, questions 6, 7, 14, and 20 showed poor discrimination. Again, no negative values were calculated. Of these, question 7 demonstrated floor effect, and questions 14 and 20 had ceiling effect but were retained (discussed previously).

Item difficulty index is a measure of the percentage of correct answers given by participants and may range from 0–1. The closer the value is to 0, the more difficult the item; the closer it gets to 1, the easier the item. Items found to have a difficulty index of less than 0.20 should be examined and revised(Ozcelik, 1998). These values for the pre-test are displayed in tabular and graphic form in Table 6 and Figure 3. The average item difficulty for the 2018-19 data set was 0.53 and ranged from 0.24 to 0.92, while the average item difficulty for the 2019-20 data set was 0.50 and ranged from 0.21 to 0.94. As noted by others in the field of chemical education, values below 0.3 are considered difficult and those above 0.8 suggest an easy item.(Luxford & Bretz, 2014) In both data sets, question 7 was found to have a high difficulty value, while questions 14, 18 and 20 were found to have low difficulty values. This was expected, as discussed earlier. Additionally, questions 8 and 13 were found to have a high difficulty level. Question 8 necessitates understanding that heat is energy and can be transferred, while "cold" is simply the absence of heat. Question 13 revolves around the idea of a polar bond and where the shared pair of electrons lie. Both are historically challenging concepts for students.

Discrimination of an item depends on its difficulty and Figure 4 shows a scatter plot of Item Difficulty vs. Item Discrimination for the 2018-19 and 2019-20 academic years. The majority of the 23 items on the CI performed within the acceptable range of difficulty and discrimination. The plot of the 2018-19 academic year shows that question 7 (floor effect), and questions 14, 18, and 20 (ceiling effect), fail both indicators. The plot for the 2019-20 academic year shows that in addition to these 4 questions, questions 13 and 15 have difficulty and discrimination values outside of the recommended range for a CI. Question 13 demonstrates low discrimination coupled with high difficulty. Question 15 demonstrates both low discrimination and low difficulty. Both will be more closely examined and revised to reflect student misconceptions more accurately in the next iteration of the CI.

Summary

The ongoing efforts, which have been presented in this paper around the development of a CI for the GOB one-semester nursing chemistry course, show promising results. The motivation was to construct an instrument that

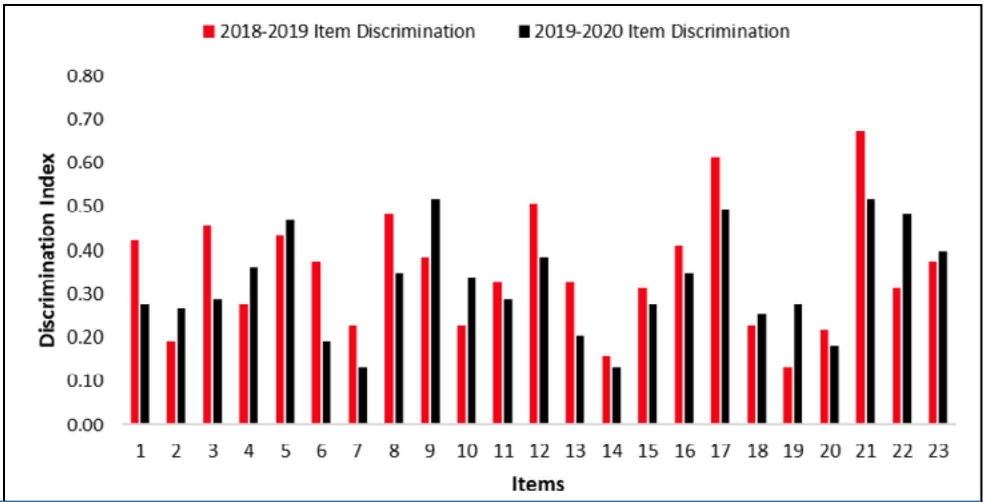


Figure 2. CI Item Discrimination for the 2018-19 and 2019-20 Academic Years

Item	2018-19 Item Difficulty	2019-20 Item Difficulty
1	0.50	0.44
2	0.45	0.37
3	0.43	0.33
4	0.41	0.33
5	0.70	0.60
6	0.63	0.59
7	0.29	0.29
8	0.28	0.21
9	0.62	0.60
10	0.34	0.35
11	0.58	0.61
12	0.48	0.43
13	0.24	0.21
14	0.92	0.94
15	0.77	0.82
16	0.59	0.57
17	0.53	0.53
18	0.86	0.87
19	0.41	0.37
20	0.87	0.89
21	0.43	0.30
22	0.56	0.61
23	0.32	0.32

Table 6 – CI Item Difficulty for the 2018-19 and 2019-20 Academic Years

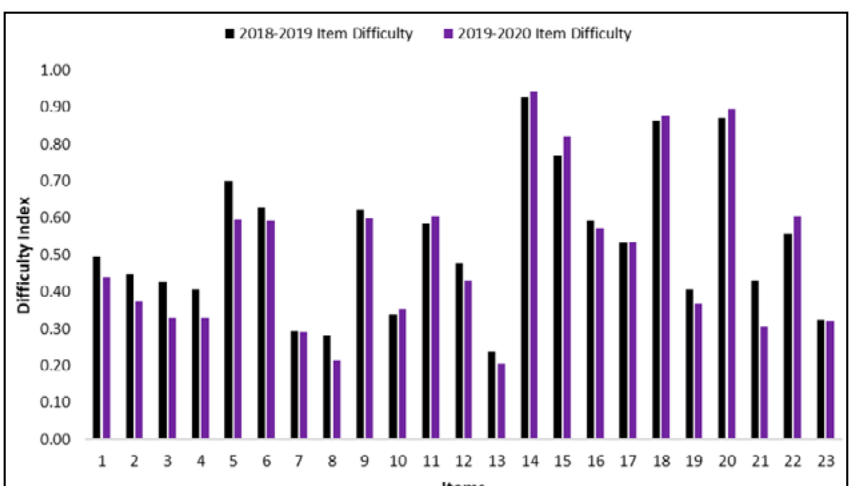
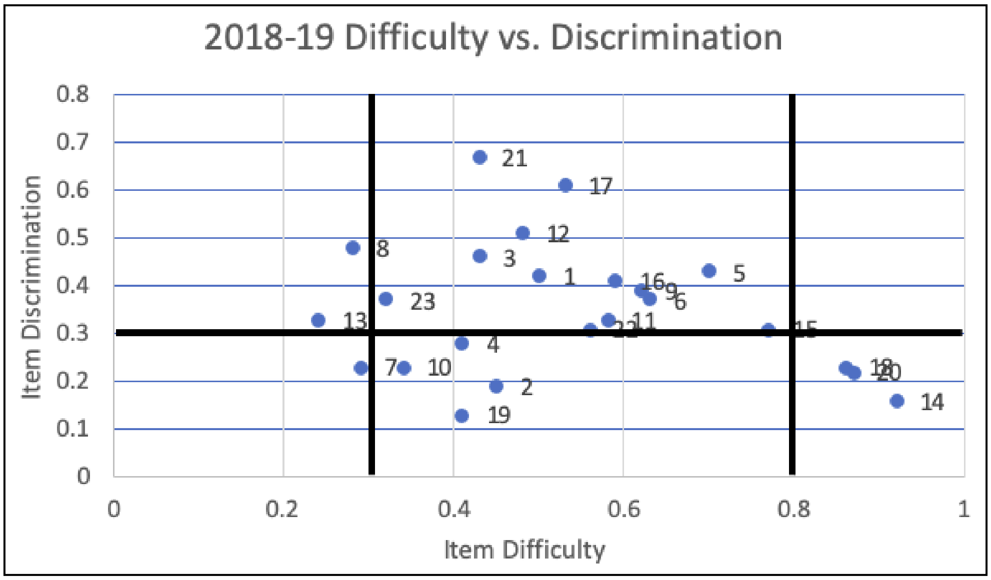


Figure 3. CI Item Difficulty for the 2018-19 and 2019-20 Academic Years

could reliably measure gains in students' understanding as well as teaching effectiveness. Normalized gains over four years of implementation of the CI were reasonable and demonstrated the accomplishment of these goals. In addition, the collected data will serve as a baseline for comparison with future course offerings. Question 7 (floor effect) and questions 14, 18, and 20 (ceiling effect) may be removed, depending on outcomes at dissemination sites. Forthcoming work will focus on modification of specific questions deemed to have low discrimination coupled with outlying difficulty values as determined by classical test theory. Specifically, distractors for questions 13 and 15 need further research to investigate their item functionality prior to the next iteration.

Acknowledgments

This work was supported by the University of Cincinnati, Departments of Chemistry, Physics, and Engineering Education and NSF grant DUE 1431350. We would also like to thank Professors Jill Shirokawa, Daniel Waddell and Peter Padolik for serving as content experts.



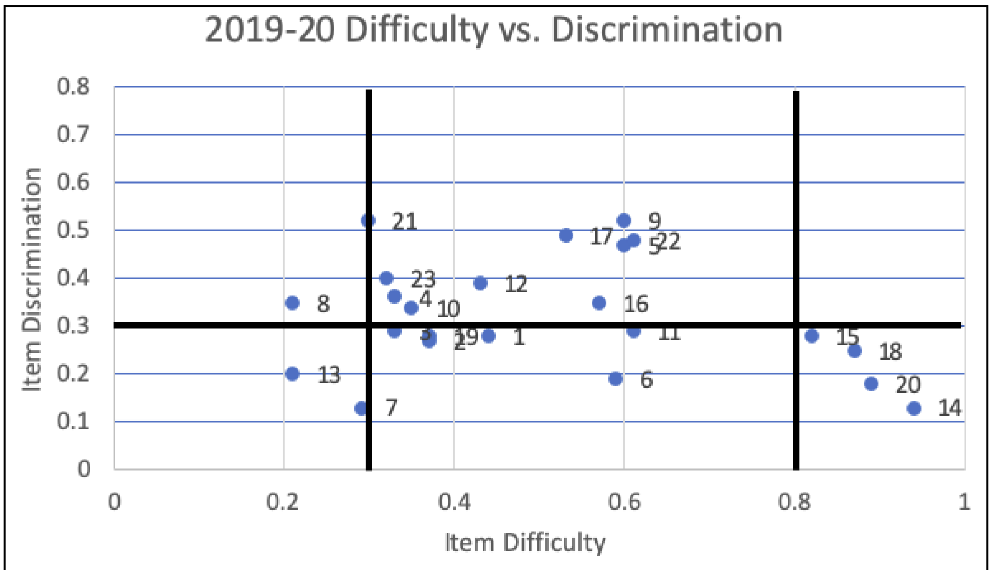


Figure 4. Scatter Plot of Item Discrimination and Item Difficulty for the 2018-19 and 2019-20 Academic Years

Literature Cited

Abell, T. N., & Bretz, S. L. (2019). Development of the Enthalpy and Entropy in Dissolution and Precipitation Inventory. *J. Chem. Educ., 96*, 1804-1812.

Ball, D. W., Hill, J. W., & Scott, R. J. (2011). *The Basics of General, Organic, and Biological Chemistry*: Saylor Academy.

Barbera, J. (2013). A Psychometric Analysis of the Chemical Concepts Inventory. *J. Chem. Educ., 90*, 546–553.

Brown, C. E., Henry, M. L. M., Barbera, J., & Hyslop, R. M. (2013). A Bridge between Two Cultures: Uncovering the Chemistry Concepts Relevant to the Nursing Clinical Practice. *J. Chem. Educ., 89*, 1114–1121.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Wenderoth, M. P., N, O., & Jordt, H. (2014). Active learning increases student performance in science, engineering. *Proc Natl Acad Sci, 111*, 8410–8415.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*, 64-74.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141-151.

Jin, Y., Rodriguez, C. A., Shah, L., & Rushton, G. T. (2020). Examining the Psychometric Properties of the Redox Concept Inventory: A Rasch Approach. *J. Chem. Educ., 97*, 4235-4244.

Krause, S., Birk, J., Bauer, R., Jenkins, B., & Pavelich, M. J. (2004). Development, testing, and application of a chemistry concept inventory. *34th Annual Frontiers in Education, FIE 2004*, T1G-1.

Luxford, C. J., & Bretz, S. L. (2014). Development of the Bonding Representations Inventory To Identify Student Misconceptions about Covalent and Ionic Bonding Representations. *J. Chem. Educ., 91*, 312–320.

Mulford, D. R., & Robinson, W. R. (2002). An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *J. Chem. Educ., 79*, 739-744.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw Hill.

Önder, F. (2016). Development and validation of the photoelectric effect concept inventory. *European Journal of Physics, 37*, 055709.

Ozcelik, D. A. (1998). Ölçme ve değerlendirme. *Ankara, ÖSYM Yayınları.*

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research, 21*, 381-391.

Ruder, S. M., & Hunnicutt, S. S. (2008). POGIL in Chemistry Courses at a Large Urban University: A Case Study. In R. S. Moog & J. N. Spencer (Eds.), *Process oriented guided inquiry learning (POGIL)*. Washington D.C.: American Chemical Society.

Sands, D., Parker, M., Hedgeland, H., Jordan, S., & Gallowy, R. (2018). Using Concept Inventories to Measure Understanding. *Higher Education Pedagogies, 3*, 173-182.

Smith, J. I., & Tanner, K. (2010). The Problem of Revealing How Students Think: Concept Inventories and Beyond. *CBE Life Sci Educ, 9*, 1-5.

Stockwell, B. R., Stockwell, M. S., Cennamo, M., & Jiang, E. (2015). Blended Learning Improves Science Education. *Cell, 162*, P933-936.

Taib, F., & Yusoff, M. S. B. (2014). Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences, 9*, 110-114.

Taskin, V., Bernholt, S., & Parchmann, I. (2015). An inventory for measuring student teachers' knowledge of chemical representations: design, validation, and psychometric analysis. *Chemistry Education Research and Practice, 16*, 460-477.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education, 2*, 159-169.

Undersander, M. A., Lund, T. J., Langdon, L. S., & Stains, M. (2017). Probing the question order effect while developing a chemistry concept inventory. *Chemistry Education Research and Practice, 18*, 45-54.

Usta, N. D., & Ayas, A. (2010). Common misconceptions in nuclear chemistry unit. *Procedia Social and Behavioral Sciences, 2*, 1432–1436.

Wright, J. C., Millar, S. B., Kosciuk, S. A., Penberthy, D. L., Williams, P. H., & Wampold, B. E. (1998). A novel strategy for assessing the effects of curriculum reform on student competence. *J. Chem. Educ., 75*, 986–992.

Anne P. Vonderheide, Ph.D., is an Associate Professor Educator in the Department of Chemistry at the University of Cincinnati. Her teaching responsibilities include general, organic and biochemistry (GOB) classes for nursing and allied health students and upper level analytical courses. Her research encompasses chemistry education with a focus on the GOB courses and undergraduate analytical laboratory courses. Additionally, she publishes in the area of analytical chemistry with a primary focus on mass spectrometry and inductively-coupled plasma mass spectrometry.

Cijy Elizabeth Sunny, Ph.D., is a research methodologist and psychometrician who has applied her skills in quantitative, mixed methods, and qualitative research methodologies in the substantive areas of STEM education research, medical education, engineering education, and more recently human trafficking, online abuse/support, and its impact on IT identity. Additionally, she has been an educator and has taught primarily physics and research methodology and has also conducted workshops in using concept mapping methodology for scale development, mixed methods research methodology for standardized patient educators, and standard setting for physician educators. She continues to invest her skills in discipline-specific STEM education research.

Kathleen Koenig, Ph.D, is a Professor of Physics at the University of Cincinnati, with her research in physics education. She earned a PhD in Physics Education from the University of Cincinnati and a Masters in Physics along with a Masters in Teaching from Miami University in Oxford, OH. She has extensive experience in the development and evaluation of pedagogies and curriculum that support student success in college-level science and math courses. She is a member of UC's Academy of Teaching and Learning and in 2021 she was named a Fellow of the American Association of Physics Teachers (AAPT).