2015

# Building a Database for the Historical Analysis of the General Chemistry Curriculum Using ACS General Chemistry Exams as Artifacts

Cynthia J. Luxford
*Iowa State University*

Kimberly J. Linenberger
*Kennesaw State University*

Jeffrey R. Raker
*University of South Florida*

John Y. Baluyut
*Iowa State University*

Jessica J. Reed
*Iowa State University*

***See next page for additional authors***

# Building a Database for the Historical Analysis of the General Chemistry Curriculum Using ACS General Chemistry Exams as Artifacts

**Abstract**
As a discipline, chemistry enjoys a unique position. While many academic areas prepared "cooperative examinations" in the 1930s, only chemistry maintained the activity within what has become the ACS Examinations Institute. As a result, the long-term existence of community-built, norm-referenced, standardized exams provides a historical artifact about the nature of content coverage in courses that stretches over decades. This work reports efforts to capture information and formulate it into a database about general chemistry content coverage over the past 20 years. Roughly 2000 items have been characterized in several ways, including (i) content, using an Anchoring Concepts Content Map; (ii) item construct, such as the presence of symbolic or visual information; and (iii) cognitive processing required, in terms of recall, algorithmic, or conceptual thinking.

**Keywords**
first-year undergraduate/general, curriculum, testing/assessment

**Disciplines**
Curriculum and Instruction | Higher Education | Other Chemistry | Science and Mathematics Education

**Comments**


**Authors**
Cynthia J. Luxford, Kimberly J. Linenberger, Jeffrey R. Raker, John Y. Baluyut, Jessica J. Reed, Chamila De Silva, and Thomas Holme

JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

# Building a Database for the Historical Analysis of the General Chemistry Curriculum Using ACS General Chemistry Exams as Artifacts

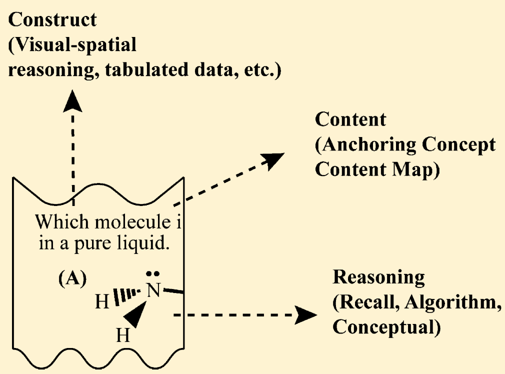
Cynthia J. Luxford,[†] Kimberly J. Linenberger,[‡] Jeffrey R. Raker,[§] John Y. Baluyut,[†] Jessica J. Reed,[†] Chamila De Silva,[†] and Thomas A. Holme*,[†]

[†]Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States

[‡]Department of Chemistry and Biochemistry, Kennesaw State University, Kennesaw, Georgia 30144, United States

[§]Department of Chemistry, University of South Florida, Tampa, Florida 33620, United States

**ABSTRACT:** As a discipline, chemistry enjoys a unique position. While many academic areas prepared "cooperative examinations" in the 1930s, only chemistry maintained the activity within what has become the ACS Examinations Institute. As a result, the long-term existence of community-built, norm-referenced, standardized exams provides a historical artifact about the nature of content coverage in courses that stretches over decades. This work reports efforts to capture information and formulate it into a database about general chemistry content coverage over the past 20 years. Roughly 2000 items have been characterized in several ways, including (i) content, using an Anchoring Concepts Content Map; (ii) item construct, such as the presence of symbolic or visual information; and (iii) cognitive processing required, in terms of recall, algorithmic, or conceptual thinking.




**KEYWORDS:** *First-Year Undergraduate/General, Curriculum, Testing/Assessment*

## ■ INTRODUCTION

The American Chemical Society Examinations Institute (ACS-EI), sponsored by the American Chemical Society's Division of Chemical Education (DivCHED), released the first standardized General Chemistry Exam in 1934.[1] Since then, the exams have been rewritten and released multiple times through a series of committees. Each committee developed their own version of the exam through an iterative process that consisted of deciding distribution of exam topics across the exam, writing items, trial testing the items, and finally collecting data on the final version in order to release standardized statistics that would help instructors determine where their students are compared to the norm.[2] It has recently been shown that the organic ACS exams can be used as historical markers indicating the state of the organic curriculum at the point of each release of the exam.[1] The organic database has been used to show fluctuations and stability of topics in organic chemistry across 60 years worth of ACS exams. It was determined that the curricula stabilized around 20−30 years ago. Because all ACS exams are developed using the same process, the various tests released in general chemistry provide similar historical artifacts of what the chemistry education community values in the general chemistry curriculum.

The current study covers 28 exams from four different types of general chemistry exams covering a 20 year time span of general chemistry. For general chemistry courses, there are several different types of exams that can be given to students in order to test their content knowledge. The four most commonly used exams for introductory college chemistry courses are the Full-Year General Chemistry Exam (designated GC), the First-Semester General Chemistry Exam (designated GCF), the Second-Semester General Chemistry Exam (designated GCS), and the General Chemistry Conceptual Exam (designated GCC). The development of the database will be presented by first describing the use of ACS Exams as artifacts, followed by the frameworks that were incorporated into the development and general findings that emerged. Each exam was analyzed using four lenses:

1. Content covered
2. Algorithmic, conceptual, recall item typology
3. Visualization components incorporated into the exams
4. Student performance

These categories are similar to those utilized in previous work analyzing ACS Organic Exams[1] and add the student performance dimension.

## ■ ACS EXAMS AS ARTIFACTS

A historical look at general chemistry exams over the past 20 years is expected to reflect stability in this course. Textbooks play an important role in shaping a course curriculum, and the topics covered in standard general chemistry textbooks have shown substantial stability over the 20 year span being analyzed

**Published:** November 26, 2014




dx.doi.org/10.1021/ed500732q | J. Chem. Educ. 2015, 92, 230−236

here. The topics currently being taught are largely the same ones that were taught in the 1990s. Therefore, it was expected that any differences in content coverage of questions would tend to reflect content the members of the test development committees felt were important for assessment of students' understanding. Most textbooks include more content than can be covered in its entirety, and instructors often view modern textbooks as encyclopedic.[3] Thus, while textbooks influence curriculum significantly and are important artifacts of the teaching enterprise, they do not necessarily represent content coverage in courses. By contrast, because of the process used by ACS Examinations Institute,[2] general chemistry exams will reflect more closely the topics that instructors emphasize in their courses. Although ACS general chemistry exams are likely to cover more content than is taught in an individual course, they do allow a glimpse into content coverage essentially as an average across multiple universities and they include less content than general chemistry textbooks. An item with a focus on content not widely taught would be expected to be discovered via trial test psychometrics, which would cause it to be eliminated before becoming part of the released, final version of the exam.

All ACS exams in general chemistry within the time period analyzed here follow a multiple choice format that uses either four or five answer options. Incorrect answer options (distractors) are developed to represent answers committee members expect a student would select based on incorrect interpretation of prompt or through common algorithmic mistakes. Trial test statistics verify these expectations. Most ACS exams contain either 60 or 70 items; however, the number of items on ACS exams could range from 40 to 120 items.[2]

## ■ COMPONENTS OF THE HISTORICAL DATABASE

Four types of general chemistry exams (GC, GCF, GCS, GCC exams) were analyzed in order to perform a historical analysis of the general chemistry curriculum spanning roughly the past 20 years. All GC, GCF, GCS, and GCC exams released between 1989 and 2012 were analyzed. Eleven GC exams with a biannual publication spanning from 1991 to 2011 were first analyzed by five raters. Exam items were independently classified using the algorithmic, conceptual, recall framework, and assigned a place on the general chemistry Anchoring Concepts Content Map (ACCM-GC). For the GC exams, the raters also assigned format types (visual spatial, specific content knowledge, recall, and computation) to the exam items. In addition to the GC exams, 10 GCF exams spanning from 1989 to 2012, four GCS exams spanning from 1998 to 2010, and three GCC exams spanning from 1996 to 2008 were analyzed by three of the raters from the GC exam analysis and a fourth separate rater. These items were analyzed identically to the full-year exams with the exception of the format types, where only the visual spatial format was coded. For all four exams, raters met after analysis of each exam to discuss differences in ratings. Raters were then asked to re-evaluate the items on the exam based on the discussion. Due to the evolving nature of the ACCM, the items were later adjusted to fit with the revised version of the general chemistry ACCM.[4]

### Content

One way of organizing the content contained in the exam questions is to code individual exam questions to the Anchoring Concepts Content Map in General Chemistry which was designed by the ACS-EI as part of a large ongoing project to develop a tool that can be used to assess student progress across the undergraduate curriculum.[5,6] ACCMs are being developed for multiple areas of chemistry with the Organic map being released recently.[7] In order to span the entire undergraduate curriculum, each ACCM has levels 1 and 2 (the topmost levels) in common. The first level of the ACCM contains 10 anchoring concept statements. Each anchoring concept (big idea) statement was then articulated via statements of enduring understandings. The big idea statements and enduring understanding statements are consistent across each of the six ACCMs (general chemistry, organic, analytical, biochemistry, inorganic, and physical chemistry). Because different courses emphasize different aspects of chemistry, the enduring understanding statements are further elaborated via "subdisciplinary articulations". The statements comprise level 3 of the ACCM, and they are further described by fine-grained content details.

The distribution of items within the ACCM shows differences across the four types of ACS general chemistry exams based on the type of exam (Figure 1). The distribution

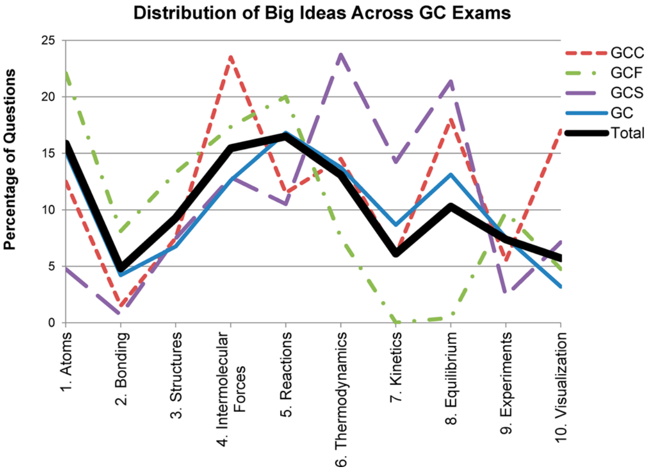



**Figure 1.** Average distribution of ACCM-GC level 1 (big ideas) across the four types of ACS general chemistry exams: full-year general chemistry (GC), conceptual general chemistry (GCC), first-term general chemistry (GCF), and second-term general chemistry (GCS). Note: The percentages do not add to 100% due to some items being coded to both a primary and secondary map statement.

of items in each big idea for the GC exam followed the percentage of items in each big idea for all 1995 items. The GCS exams show more emphasis on equilibrium, thermodynamics, and kinetics than the full-year exam with 60% of all GCS exams being classified as big ideas 6, 7, and 8. This observation is commensurate with traditional content coverage in the second semester of the full-year course. In a similar way, the GCF exam had a larger emphasis on atoms, bonding, and reactions than the other exams and almost no coverage of kinetics and equilibrium. The GCC exam had a larger emphasis on intermolecular forces with 23% of all GCC items fitting into big idea 4 and with another 18% of the items testing big idea 8. Only 2% of the GCC items focused on bonding. Further analysis of each exam by year and by examining the level 2 articulation statements can reveal further differences between the exams.

## Algorithmic, Conceptual, and Recall (ACR) Framework

Another framework that was used to classify ACS exam items was to label items as algorithmic (A), conceptual (C), or recall (R) questions. Nurrenbern and Pickering first described conceptual learning as being different from problem solving.[8] It was pointed out that students were mainly being asked to solve problems, but evidence was given to show that although students were able to successfully solve problems, they did not necessarily need to have a strong understanding of the concepts behind the algebraic approaches. The distinction between conceptual learning and algorithmic learning (problem solving) has been further expanded upon in the literature.[9−15] The algorithmic problems can be solved using algorithms or determined mathematic expressions. The patterns can be memorized limiting the measurement of students' conceptual understanding. Conceptual problems are higher order questions that can assess student understanding of the ideas behind the chemical phenomena. Conceptual problems tend to assess student understanding when applied in situations that are beyond what students may have memorized. Recall questions are written to assess knowledge that the student either knows or does not know based on memorization. An expanded framework was developed to help aid chemical educators in constructing exams using three primary levels: definition (recall), algorithmic, and conceptual.[16]

Each of the 1995 items across the four different types of exams were classified as algorithmic, conceptual, or recall using the expanded framework developed by Smith, Nahkleh, and Bretz.[16] This exercise is summarized by Figure 2. It was found

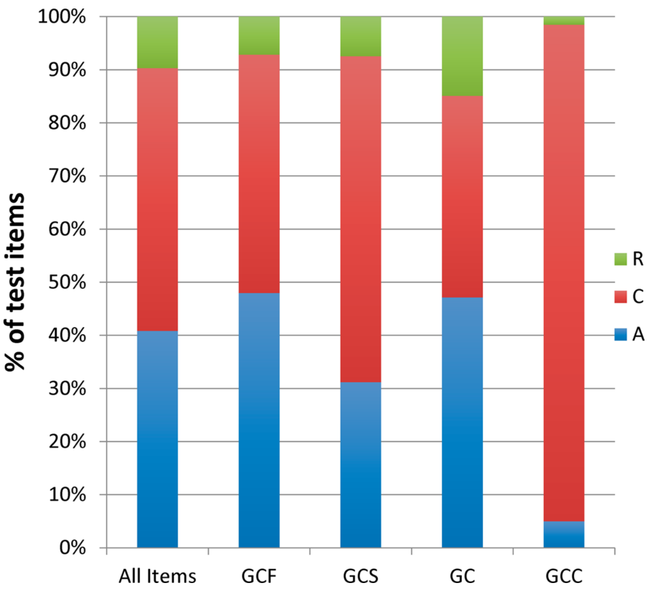



**Figure 2.** Distribution of general chemistry ACS exam items as classified using the algorithmic (A), conceptual (C), and recall (R) framework.

that, as expected, the GCC exams held a much higher percentage of conceptual items. There were, however, still a few items that could be solved using algorithms as well as a few items that required memory recall in order to select the correct answer. The GCF and GC exam contained the largest percentage of algorithm questions, with roughly half of the items requiring students to either use formulas or memorized algorithms in order to answer the test item. The GC exams also contained the largest percentage of recall items, with 15% of all GC exam items requiring specific memorized content knowledge in order to solve the question.

While the ACR framework gives an indicator of the types of questions being asked, pairing the ACR framework with the ACCM-GC mapping gave a better picture of how each big idea is being tested. The 1995 items can be mapped to show the percentage of the items that were classified into each big idea that tested using A, C, and R questions. Figure 3 shows the

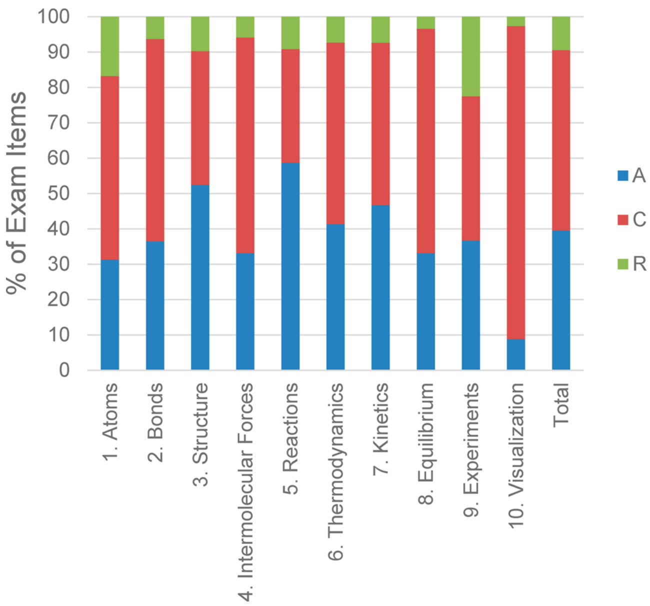



**Figure 3.** Distribution of general chemistry ACS exam items to the algorithmic (A), conceptual (C), and recall (R) framework across the 10 big ideas from the ACCM for general chemistry.

distribution of ACR items across the different big ideas. Topics that typically contain large number of calculations such as stoichiometry (in reactions) tend to have a higher percentage of items that are algorithmic. Indeed, 58% of all items that align with the reaction ideas anchoring concept require algorithmic thinking. Topics that are traditionally taught as more conceptual such as intermolecular forces (61%) and visual understanding of graphs and particulate nature of matter (visualization, 89%) tend to be tested with a higher percentage of conceptual questions. Equilibrium, which is traditionally taught through a high usage of algebraic examples of "ICE tables" and $K_{eq}$ problems, had a high amount of conceptual problems. Sixty-three percent of all equilibrium questions were labeled as conceptual questions, with only 33% of the equilibrium questions classified as algorithmic questions. This distinction is important. The ranking of items in the ACR framework does not imply that all mathematical items are algorithmic. Items relating to experiments had the highest percentage of recall type questions at 22% of the 33 items relying on straight memorization.

## Item Format

Each item was also coded based on the format. In this category, the construct of the items was analyzed to determine if specific elements were required such as visual spatial components, specific content knowledge, reasoning, and computation.[17] Specific content knowledge (SCK) was defined as facts that students simply "need to know" in order to answer the


dx.doi.org/10.1021/ed500732q | J. Chem. Educ. 2015, 92, 230−236

**Table 1. Agreement between the Algorithmic, Conceptual, Recall, and the Specific Content Knowledge, Reasoning, Computation Frameworks for ACS Full-Year General Chemistry Exam Items**

| | Agreement of Question Type with SCK/RS/Comp Framework Item Format, % ($N$ = 785) | | | | | |
|---|---|---|---|---|---|---|
| ACR Framework Question Type | Computation | No Computation | Reasoning | No Reasoning | Specific Content Knowledge | No Specific Content Knowledge |
| Algorithmic | 98.91[a] | 1.44 | 3.82 | 76.01 | 3.88 | 55.64 |
| Conceptual | 1.09 | 70.50 | 93.95[a] | 0.64 | 6.20 | 44.21 |
| Recall | 0.00 | 28.06 | 2.23 | 23.35 | 89.92[a] | 0.15 |

[a]$p$ < 0.001.

question. An example of an item measuring SCK would be a question asking the name of a compound. Reasoning (RS) items are expected to require students to reason through the different concepts or skills in order to answer the question. It is important to note that items with RS are rated as such from the perspective of students who are answering the item using methods commonly associated with general chemistry. This is designed to contrast with items that tend to allow students to solve problems with set heuristics or mathematical expressions. Such items do not necessarily require them to reason through to the answer. Computation (Comp) questions involve one or more calculations that students must do in order to solve the question. Computation also includes questions that can be solved using a set algorithm or memorized set of steps such as writing out the electron configuration.

The GC exams were all coded using the SCK/RS/Comp framework.[17] It was found that the computation items were strongly correlated to items labeled as algorithmic using the ACR framework (Table 1). The reasoning questions were also correlated with the conceptual questions from the ACR framework, and the specific content knowledge correlated with the recall questions. Fisher Exact tests showed that all three comparisons were significant with $p$ < 0.001. Therefore, for the coding of the GCC exams, GCF exams, and the GCS exams, only the ACR framework was used rather than largely duplicating this information with two components in the emerging historical database.

The items were also analyzed based on visual−spatial components used to convey information. This categorization includes a range of constructs that share the characteristic of not being only text. Thus, items were coded as containing visuals if chemical equations, chemical structures, particulate nature of matter diagrams, tables, graphs, pictures, or mathematic expressions were present. This level of description allows two forms of analysis. The first is a sum of any visual−spatial component. The second level of analysis includes categorizing the specific nature of the nontext information.

Therefore, each exam item on the four general chemistry ACS exams was classified first based on whether there were visual representations in the item, then classified a second time based on the types of visuals represented in the items (Figure 4). From the graphic, it is not obvious which differences are significant. An ANOVA showed a significant difference in the ratio of visual items to nonvisual items at the $p$ < 0.05 level for items from the four different types of exams [$F(3)$ = 10.15, $p$ < 0.001]. Post hoc comparisons using independent $t$ tests indicated that the ratio of visuals to nonvisuals for the GCC exams ($M$ = 0.63, SD = 0.03) was significantly different from the GC exam ($M$ = 0.47, SD = 0.02) and the GCF exams ($M$ = 0.44, SD = 0.02). The GCC exam was not significantly different from the GCS exam ($M$ = 0.57, SD = 0.03). The GCS exam was significantly different from the GC exam and the GCF exam. The GC and GCF exams were not significantly different

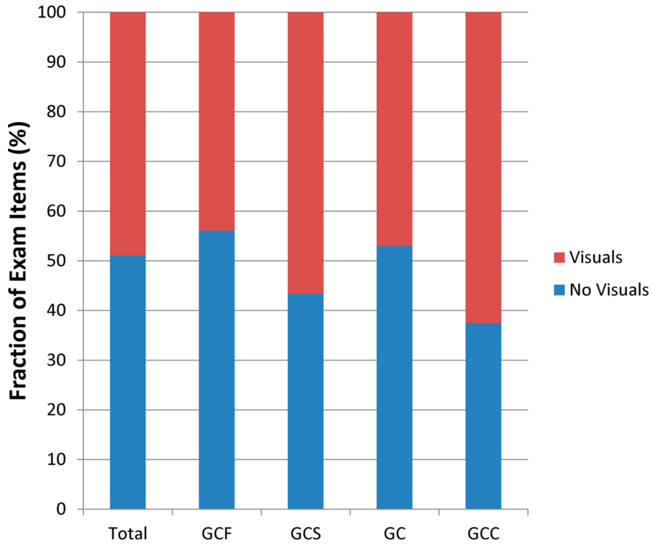



**Figure 4.** Percentage of items that contain visual elements on the full-year general chemistry (GC), general chemistry conceptual (GCC), first-term general chemistry (GCF), and second-term general chemistry (GCS) ACS exams.

from each other. This means that the GCC and GCS exams had similar ratios of visuals to each other but significantly different ratios from the GC and GCF exams. The GC and GCF exams are not significantly different from one another in terms of the use of nontext constructs in the items.

The items with visuals were also coded based on whether the items had chemical equations, tables, graphs, chemical structures, particulate nature of matter (PNOM) diagrams, pictures, or mathematical expressions, and this analysis is presented in Figure 5. Note that several items contained multiple different types of visuals. Unsurprisingly, the two most common forms of representations are chemical equations (reactions) and tabulated data. Of all the exam items that had visual components, 52.3% of those items contained chemical equations while 32.0% of the items with visuals contained tables. A closer look at individual exams revealed that GCF has relatively more tables than the other three exam types. In addition, GCS and GC both have relatively more chemical equations than the GCF and GCC exams. However, GCC has more PNOM and more pictures than the other exams. GCF has more structures because it tests students understanding of Lewis structures. It is also important to note that exam items seldom explicitly include mathematical expressions; however, many ACS exams include data sheets with some common equations provided that they are not item specific. It can also be noted that, while chemical equations have different levels of influence on the item, it is not adjudicated in the current analysis.

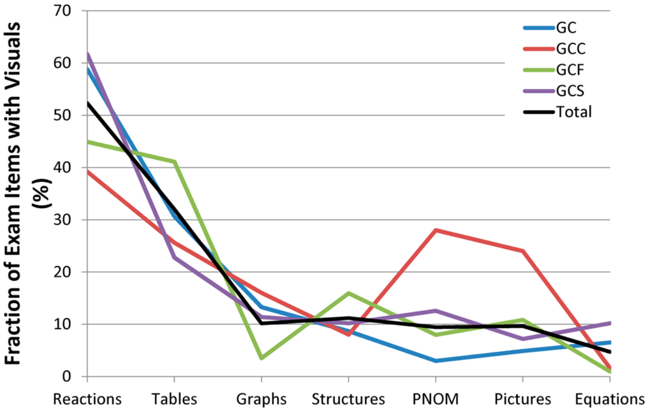



**Figure 5.** Distribution of items that contain visual components across the four types of ACS general chemistry exams: full-year general chemistry (GC), conceptual general chemistry (GCC), first-term general chemistry (GCF), and second-term general chemistry (GCS). Note: The percentages do not add to 100% due to some items containing more than one visual component.

## Student Performance Norm Data

As part of the ACS-EI exam standardization process, the ACS-EI gathers norm data for each of the exams. The item difficulty and item discrimination were included in the database for the 1690 items where student level data were available. Item difficulty is the fraction of students answering the item correctly, while item discrimination indicates how well the item is able to distinguish between top-scoring students and lower-scoring students.[18] The average item difficulties and discrimination values are reported in Table 2. Overall, the item difficulties and the item discriminations on the various exams had a large range of values across all four exam types. Because ACS exams are designed as norm-referenced tools and their development includes a trial-testing round, it is not surprising that the average discrimination for each of the four exams is not significantly different from one another. Since the overall average of different tests is not substantially different, differences in difficulty related to content as described by the ACCM can help reveal the relative success that students have learning different topics. This is best accomplished by considering difficulties within the 10 anchoring concepts.

Considering all items analyzed, the average difficulties for items in each of the 10 big ideas ranged from 0.494 to 0.592. Overall, equilibrium and thermodynamics questions tend to have lower average difficulty values (fewer students answer them correctly, so these topics are "harder" for students), while bonding, atoms, experiments, and visualization questions tend to have higher difficulty values, as seen in Figure 6. It is important to note that the iterative process associated with trial testing of items may tend to eliminate items that are particularly easy or difficult. In addition to looking at the overall average difficulties per big idea on the ACCM, a comparison of difficulties by ACR and big ideas was also carried out and is presented in Figure 7. Perhaps the key observation is that there is no single category of item that is always easier or always more difficult. For example, items that are coded as recall tend to be less difficult (higher $p$) when the content is atoms, kinetics, or equilibrium, but not so for other areas. In "bonding and reactions", conceptual items test easier than recall, essentially the opposite of the trend in "atoms". Algorithmic items show low $p$ scores (are relatively hard) in both "equilibrium" and "visualization", but this observation is potentially influenced by

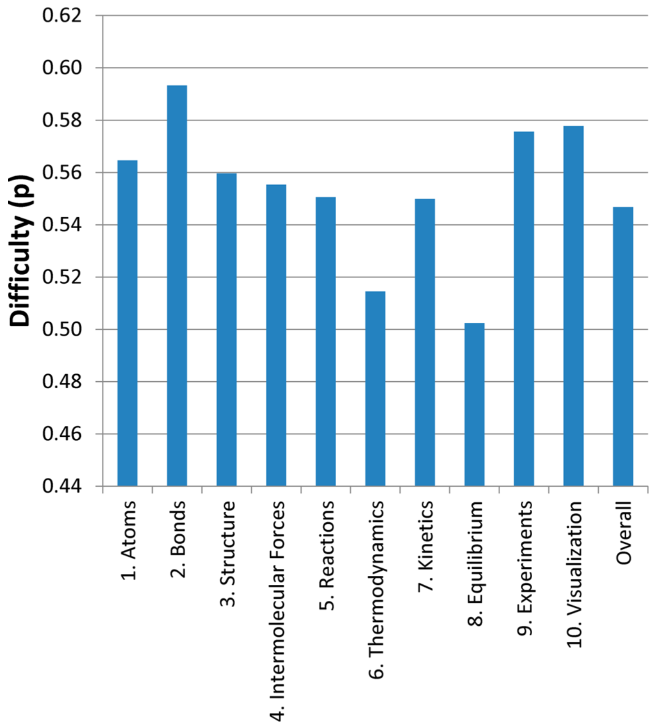



**Figure 6.** Average difficulty values ($p$) across the 10 big ideas for all general chemistry ACS exam items with psychometric data available ($N$ = 1690).

**Table 2. Summary of Item Difficulties and Item Discrimination for Four Types of General Chemistry ACS Exams**

| Measures of Item Difficulty ($p$) and Item Discrimination ($D$) | Average Item Difficulty and Discrimination Values[a] by Exam Type (Item Number): Full-Year General Chemistry Exam ($N$ = 785) | General Chemistry Conceptual Exam ($N$ = 200) | First-Term General Chemistry Exam ($N$ = 490) | Second-Term General Chemistry Exam ($N$ = 295) | Total ($N$ = 1690) |
|---|---|---|---|---|---|
| Minimum $p$ | 0.160 | 0.053 | 0.160 | 0.130 | 0.053 |
| Maximum $p$ | 0.900 | 0.960 | 0.930 | 0.927 | 0.96 |
| Average $p$ | 0.532 ± 0.154 | 0.545 ± 0.181 | 0.578 ± 0.157 | 0.535 ± 0.160 | 0.547 ± 0.160 |
| Minimum $D$ | −0.067 | −0.134 | −0.067 | −0.080 | −0.134 |
| Maximum $D$ | 0.800 | 0.672 | 0.770 | 0.760 | 0.800 |
| Average $D$ | 0.404 ± 0.140 | 0.407 ± 0.160 | 0.416 ± 0.129 | 0.379 ± 0.151 | 0.403 ± 0.141 |

[a]Item difficulty is defined as the fraction of students answering the item correctly (values above 0.8 suggest an easy item, while difficulties below 0.3 are considered difficult questions). Item discrimination indicates how well the item is able to distinguish between top-scoring students and lower-scoring students (values above 0.3 typically are considered to have good item discrimination).[18]

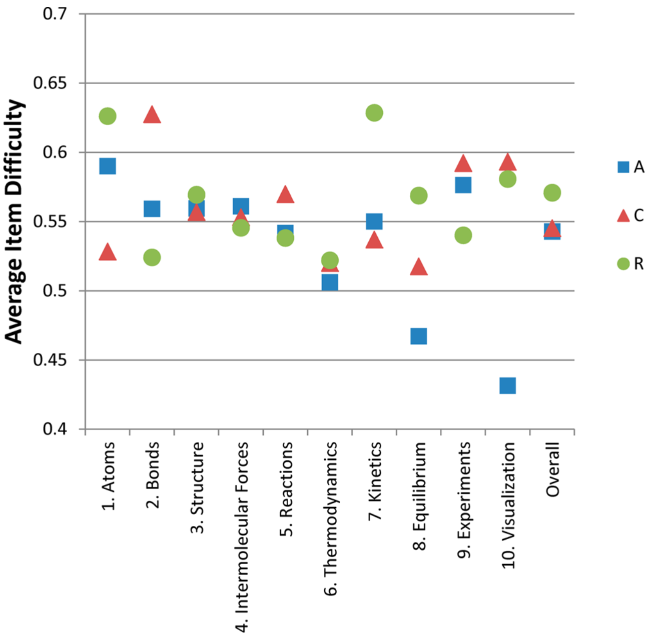



**Figure 7.** Average difficulty values for ACS general chemistry exam items as classified as algorithmic (A), conceptual (C), and recall (R) questions for each of the 10 big ideas on the ACCM for general chemistry.

the relatively small number of items in these specific categories. Because items are chosen for released exams based on statistics from trial testing,[2] it is dangerous to overgeneralize these findings. Nonetheless, given the long time frame and large numbers of items analyzed here, they may point to interesting topics for further research.

## CONCLUSIONS AND IMPLICATIONS

The database being developed by ACS-EI contains a wide variety of different measures that can be used to explore the changes in the general chemistry curriculum over the past 20 years as encapsulated in the development stages of the various ACS general chemistry exams. The database has so far shown that certain topics such as atoms, bonds, reactions, and intermolecular forces have been tested more frequently than others such as kinetics and experiments. Roughly half of the items on ACS exams over the past 20 years test students' conceptual understanding of general chemistry; however, there are variations across the different exams and the different content areas within each exam. Since the ACS exams are standardized and norm-referenced by design, the average item difficulties across the different content areas have a narrow range with the majority of the ACCM big ideas having approximately 0.5 average difficulties.

Further study with the data from the general chemistry exams is currently underway. This work is expected to show patterns over time in the development of items at the enduring understanding level of the ACCM. Another key component of categorizing items from ACS exams in this way is that it will provide a mechanism by which instructors can compare student performance on items they construct themselves with patterns of student performance on these nationally normed exams. The development of different forms of analysis reported here is being done, in part, to facilitate this development of comparative tools. The tools themselves are being created and once complete will help instructors characterize their own exam questions by matching them to a location on the ACCM. Importantly, having the other characteristics elaborated should also provide tools to help instructors make comparisons, even though ACS exam items themselves cannot be directly accessed due to the secure nature of these tests.[2]

## AUTHOR INFORMATION

### Corresponding Author

*E-mail: taholme@iastate.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Partial funding for the development of the ACCM was provided by the National Science Foundation (NSF-DUE 071779, 0943783). Alignment work and continued refinements to the ACCM were carried out under the auspices of NSF-DUE 1323288. Any opinions, findings, and conclusions communicated in this material are those of the authors and do not necessarily reflect the views of NSF.

## REFERENCES

(1) Raker, J. R.; Holme, T. A. A Historical Analysis of the Curriculum of Organic Chemistry Using ACS Exams as Artifacts. *J. Chem. Educ.* **2013**, *90*, 1437.
(2) Holme, T. A. Assessment and Quality Control in Chemistry Education. *J. Chem. Educ.* **2003**, *80*, 594.
(3) Kerber, R. C. Elephantiasis of the Textbook. *J. Chem. Educ.* **1988**, *65*, 719.
(4) Holme, T.; Luxford, C.; Murphy, K. Updating the General Chemistry Anchoring Concepts Content Map. *J. Chem. Educ.* Submitted.
(5) Murphy, K.; Holme, T.; Zenisky, A.; Caruthers, H.; Knaus, K. Building the ACS Exams Anchoring Concept Content Map for Undergraduate Chemistry. *J. Chem. Educ.* **2012**, *89*, 715.
(6) Holme, T.; Murphy, K. The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map I: General Chemistry. *J. Chem. Educ.* **2012**, *89*, 721.
(7) Raker, J.; Holme, T.; Murphy, K. The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map II: Organic Chemistry. *J. Chem. Educ.* **2013**, *90*, 1443.
(8) Nurrenbern, S. C.; Pickering, M. Concept Learning versus Problem Solving: Is There a Difference? *J. Chem. Educ.* **1987**, *64*, 508.
(9) Sawrey, B. A. Concept Learning versus Problem Solving: Revisited. *J. Chem. Educ.* **1990**, *67*, 253.
(10) Pickering, M. Further Studies on Concept Learning versus Problem Solving. *J. Chem. Educ.* **1990**, *67*, 254.
(11) Nakhleh, M. B.; Mitchell, R. C. Concept Learning versus Problem Solving: There Is a Difference. *J. Chem. Educ.* **1993**, *70*, 190.
(12) Gabel, D. L.; Bunce, D. M. Research on Problem Solving: Chemistry. In *Handbook of Research on Science Teaching and Learning*; Gabel, D. L., Ed.; MacMillan: New York, 1994; p 301.
(13) Phelps, A. J. Teaching To Enhance Problem Solving: It's More than the Numbers. *J. Chem. Educ.* **1996**, *73*, 758.
(14) Nurrenbern, S. C.; Robinson, W. R. Conceptual Questions and Challenge Problems. *J. Chem. Educ.* **1998**, *75*, 1502.
(15) Robinson, W. R.; Nurrenbern, S. C. Conceptual Questions (CQs): What Are Conceptual Questions? http://www.chemedx.org/JCEDLib/QBank/collection/CQandChP/CQs/WhatAreCQs.html (accessed November 2013).
(16) Smith, K. C.; Nakhleh, M. B.; Bretz, S. L. An Expanded Framework for Analyzing General Chemistry Exams. *Chem. Educ. Res. Pract.* **2010**, *11*, 147.
(17) Kendhammer, L.; Holme, T.; Murphy, K. Identifying Differential Performance in General Chemistry: Differential Item Function-

ing Analysis of ACS General Chemistry Trial Tests. *J. Chem. Educ.* **2013**, *90*, 846.
(18) Ding, L.; Beichner, R. Approaches to Data Analysis of Multiple-Choice Questions. *Phys. Rev. Spec. Top.–Phys. Educ. Res.* **2009**, *5*, 1−17.