1-10-2017

Analyzing the Role of Science Practices in ACS Exam Items

Jessica J. Reed  
University of Wisconsin - Milwaukee

Alexandra R. Brandriet  
University of Iowa

Thomas A. Holme  
Iowa State University, taholme@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/chem_pubs

Part of the Chemistry Commons, Curriculum and Instruction Commons, Elementary Education Commons, Science and Mathematics Education Commons, and the Secondary Education Commons

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/chem_pubs/1057. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

This Article is brought to you for free and open access by the Chemistry at Iowa State University Digital Repository. It has been accepted for inclusion in Chemistry Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Analyzing the Role of Science Practices in ACS Exam Items

Abstract
Recent efforts to reform K–12 science curricula, embedded within the NRC Framework for K–12 Science Education and the Next Generation Science Standards, have focused on unifying core disciplinary content with crosscutting concepts that span across science disciplines and scientific practices. With these reforms comes the challenge of creating “three-dimensional” assessments that measure beyond basic content knowledge mastery to incorporate measures of science practices and crosscutting concepts. Adding measures of science practices to traditional forms of assessments may prove to be challenging. The work herein investigates how science practices have been inherently incorporated into previously released ACS exams through analysis and classification of ACS general chemistry exam items with a developed rubric. The findings may help to inform the creation of new assessment items that are more explicit in their incorporation of science practices by building upon the implicit incorporation of science practices already present in many of the items analyzed. By creating awareness of the presence of science practices in current assessment items, exam designers may have more ease sculpting items that encompass the three dimensions outlined in the NRC Framework.

Keywords
Chemical Education Research, General Public, Testing/Assessment

Disciplines
Chemistry | Curriculum and Instruction | Elementary Education | Science and Mathematics Education | Secondary Education

Comments

Rights
This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

This article is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/chem_pubs/1057
Analyzing the Role of Science Practices in ACS Exam Items
Jessica J. Reed,† Alexandra R. Brandriet,‡ and Thomas A. Holme*‡
Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States

ABSTRACT: Recent efforts to reform K–12 science curricula, embedded within the NRC Framework for K–12 Science Education and the Next Generation Science Standards, have focused on unifying core disciplinary content with crosscutting concepts that span across science disciplines and scientific practices. With these reforms comes the challenge of creating “three-dimensional” assessments that measure beyond basic content knowledge mastery to incorporate measures of science practices and crosscutting concepts. Adding measures of science practices to traditional forms of assessments may prove to be challenging. The work herein investigates how science practices have been inherently incorporated into previously released ACS exams through analysis and classification of ACS general chemistry exam items with a developed rubric. The findings may help to inform the creation of new assessment items that are more explicit in their incorporation of science practices by building upon the implicit incorporation of science practices already present in many of the items analyzed. By creating awareness of the presence of science practices in current assessment items, exam designers may have more ease sculpting items that encompass the three dimensions outlined in the NRC Framework.

KEYWORDS: General Public, Chemical Education Research, Testing/Assessment

FEATURE: Chemical Education Research

INTRODUCTION

Reform efforts in science education have often turned attention to ways that curriculum and assessment differ. One common result of these efforts is to suggest that assessments ought to provide evidence of student performance beyond basic content knowledge.1−3 A key example of this trend can be found in A Framework for K–12 Science Education: Practices, Crosscutting Concepts and Core Ideas,4 which is a report from the National Research Council. This report is often referred to as the Framework. An important step toward implementing the ideas presented in the Framework has been published in the Next Generation Science Standards (NGSS).5 The NGSS provide specific outcomes of K–12 science education that intertwine disciplinary content, practices, and crosscutting concepts. This triad of components allows educators to identify targets for assessment beyond content proficiency.6 Within the same time frame, additional efforts have been made to enhance the quality of assessment, including those from The College Board7−10 and changes to the Medical College Admissions Test.11,12 Taken together, these assessment related efforts suggest that substantial changes are needed in order to be sure that students can use the information they gain in science courses. Large-scale assessment programs face a particular challenge when responding to calls for reform because their scale often leads to a reliance on multiple-choice questions, which are not necessarily written with more complex measurement objectives in mind. In light of the recent evidence-based reforms to curricula and assessments, and the call for reform of STEM learning in higher education,11 the work reported here explores the incorporation of measures beyond content proficiency in chemistry tests from the American Chemical Society Examinations Institute (ACS-EI). This analysis aims to garner a better understanding of how skills beyond content have been valued and evaluated in the assessment of chemistry.

Science Practices

While the idea of science practices, or process skills, is not novel,12 the Framework brings new light to these ideas by defining eight practices for all science and engineering disciplines, and suggesting that these practices are an integral component of classroom learning and thus should be emphasized and assessed similarly to course content.1,2 Previous research has suggested that even though chemistry faculty value these practices and want students to develop them, they often do not assess them. Reasons for this include a lack of awareness of assessment instruments designed to make measurements beyond content knowledge and most importantly a lack of time.13−16 Thus, assessment developers may need to consider ways to embed measures of science practices into traditional forms of content knowledge assessment while maintaining appropriate validity and reliability characteristics of the tests.

Use of the word practices is important because it explicitly acknowledges that scientists use both knowledge and skill to accomplish their scientific inquiries. The eight science practices

Received: August 29, 2016
Revised: November 2, 2016
Published: November 26, 2016
Box 1. Eight Science Practices As Outlined in the NRC Framework for K–12 Science Education

Science Practices
1. Asking Questions
2. Developing and Using Models
3. Planning and Carrying Out Investigations
4. Analyzing and Interpreting Data
5. Using Mathematics and Computational Thinking
6. Constructing Explanations
7. Engaging in Argument from Evidence
8. Obtaining, Evaluating, and Communicating Information

(SP)s are displayed in Box 1, and further descriptions of each can be found within the Framework. The components of the practices that relate to engineering were not used in this analysis, because the test items analyzed focused on chemistry only.

Meaningful Learning
An advantage of incorporating science practices into traditional test items is that more robust assessment is often more aligned with the theory of meaningful learning. This theory stems from Constructivist roots and asserts that students are able to best learn when a component explicitly. It should be considered that if the goal is for students to learn meaningfully, and to avoid extensive rote memorization and factual recall, then science practices should be assessed.

RESEARCH QUESTIONS
The primary research questions addressed by the current study follow.
1. To what extent have established assessments in chemistry, as exemplified by ACS exams, incorporated science practices?
2. Which science practices, if any, are incorporated in ACS exam items most often?

Analysis of ACS exams also led to additional questions that are related to these two main research questions. These questions are also noted here, and they relate to how science practices are incorporated across different aspects of ACS exams such as the specific type of exam, item classifications, and chemistry content domains.

METHODS
Assessment Items
The primary source of data for this research came from standardized chemistry exam items from the ACS-EI. ACS exams carry secure exam copyrights, so no exam items can be included in this report. Instead, examples for discussion purposes will be provided by mock questions that are similar in content and construct. An ACS exam is developed by a committee of chemists who teach the course for which the exam is intended. Developers are not governed by item specifications from the ACS-EI, but because they teach a course related to the exam being developed means that exam content will be at a level that reflects national trends of content coverage.

Given the goal to investigate how science practices are being incorporated across the general chemistry curriculum and the fact that the ACS-EI produces several different versions of exams for this course, a number of different exams were analyzed in this project. In total, the incorporation of science practices was investigated for 12 released ACS exams (N = 738 items). Descriptions are provided in Table 1 of the exams used in this analysis. An exam not used in general chemistry, the Diagnostic of Undergraduate Chemistry Knowledge (DUCK), was included in the analysis because it is scenario-based and therefore could be expected to have a different level of incorporation of science practices.

Because there are slight differences in the goals and practicities of exams, a number of factors influenced which exams were included in the analysis here. For example, in general chemistry in particular, conceptual test items are often designed to eliminate certain algorithmic approaches to the content so they might incorporate science practices in different ways than traditional items. The PQ exams also

<table>
<thead>
<tr>
<th>Examination</th>
<th>Exam ID</th>
<th>Number of Items Analyzed</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Chemistry (Full Year)</td>
<td>GC</td>
<td>215</td>
<td>Associated with a year-long general chemistry course.</td>
</tr>
<tr>
<td>General Chemistry First-Term</td>
<td>GCF</td>
<td>70</td>
<td>Associated with the first term of a general chemistry sequence.</td>
</tr>
<tr>
<td>General Chemistry Conceptual</td>
<td>GCC</td>
<td>180</td>
<td>General chemistry content associated with a year-long course assessed in a conceptual manner.</td>
</tr>
<tr>
<td>Paired Questions First-Term</td>
<td>PQF</td>
<td>40</td>
<td>Pairs of questions, conceptual and traditional (algorithmic), associated with the first semester of general chemistry.</td>
</tr>
<tr>
<td>Paired Questions Second-Term</td>
<td>PQS</td>
<td>40</td>
<td>Pairs of questions, conceptual and traditional (algorithmic), associated with the second semester of general chemistry.</td>
</tr>
<tr>
<td>Laboratory (Online)</td>
<td>LAB</td>
<td>Approx 40</td>
<td>Content associated with general chemistry laboratory experiments, equipment, and procedures. Conducted via an online computer interface.</td>
</tr>
<tr>
<td>Chemistry in Context</td>
<td>CIC</td>
<td>90</td>
<td>Content associated with chemistry as it relates to real-world applications and contexts.</td>
</tr>
<tr>
<td>Diagnostic of Undergraduate Chemistry Knowledge</td>
<td>DUCK</td>
<td>60</td>
<td>Scenarios to assess content across the undergraduate chemistry curriculum.</td>
</tr>
</tbody>
</table>
have a number of conceptual items so they were included in the analysis. The GC exam series is released every other year, so the number of items available to analyze is much larger than other exams as is reflected by the number of items analyzed for this work. With many possible exams from which to choose, those with release dates near those of the GCC exams were analyzed. The GCF exam was analyzed in part because of the role it plays in other ACS-EI projects where evidence related to the incorporation of science practices on the exam may be useful. Both the DUCK and CIC exams tend to include real-world applications and help broaden the scope of chemistry content analyzed. The LAB exam is an online exam, and laboratory environments are likely to be more naturally tied to skills, so it was included both because of the content and constructs of the test items.

**Rubric for Analyzing Assessment Materials**

The analysis of whether or not items on chemistry tests contain science practices requires a rubric to determine the extent to which the science practices outlined by the Framework are present in the assessment. Researchers at Michigan State University have developed such a rubric, called the Three Dimensional Learning Assessment Protocol (3D-LAP). This tool provides a means to evaluate the three dimensions of learning (science practices, core content, and crosscutting concepts) discussed by the Framework. Use of this rubric allowed classification of whether or not chemistry items from ACS exams contained science practices.

The initial development of the rubric used an extensive and iterative process with educators from all science disciplines and included applications within K–12 education as well as postsecondary education. The rubric itself was initially envisioned primarily for free-response questions, and the version from which our efforts began required some minor modifications so that classification of chemistry-specific multiple-choice items was achievable. For example, it was necessary to add additional criteria to the way the practice of Developing and Using Models (SP 2) was articulated to account for students’ explanation of how or why they used a particular model or representation presented in a multiple-choice question.

Items were rated for incorporation of science practices by two chemical education researchers. These raters initially reviewed the assessment items individually and then convened to discuss their independent results. On the basis of the individual ratings only, agreement was present for roughly 90% of items. When item classifications arose where the raters disagreed, raters reached 100% agreement by discussing the items in light of the rubric. If these discussions met an impasse, a third rater with expertise in assessment was consulted to resolve the rating dispute. In these instances, the assessment expert would review the item(s) in question independently before hearing arguments for a particular rating from the researchers. The expert and the raters would then have a discussion until 100% agreement could be reached on the classification of the item(s) in question.

**Classification of Assessment Items**

The 3D-LAP rubric was used to analyze all of the items on the selected ACS exams noted in Table 1 for incorporation of science practices. There was no limit to the number of science practices that could be assigned in any item, and some items were identified that contained multiple science practices. In the majority of cases, however, if any science practice was found in an item, there was only one. Three was the maximum number of science practices found in a single item.

Once an item was judged to include a science practice, by using the 3D-LAP rubric, the raters determined whether the presence of the practice was “explicit” or “implicit”. The items from ACS exams essentially represent artifacts of the current state of chemistry education, so they are not designed to include science practices, even if the long-term goal of efforts such as the Framework and NGSS is to enhance assessment related to practices. In the current efforts, a practice was considered “explicit” only if all criteria in the 3D-LAP rubric were met for that practice.

The use of the 3D-LAP can be demonstrated using mock ACS items such as those shown in Boxes 2 and 3. These items are designed to exemplify ACS exam items which cannot be published because of their secure copyright. They were developed separately from the current project for use on a national survey conducted by the ACS-EI. Mock Item 1 (Box 2) was designed to represent testing of conceptual knowledge. The item contains several elements, including a particulate nature of matter (PNOM) diagram, graphs, and a balanced chemical equation. To answer this item, a student would have to access information cohesively from all three sources. The analysis of this, or any item, begins when raters independently review the questions, then converge to discuss whether a science practice was present. In this item, the science practice Developing and Using Models (SP 2) is present, because a student would need to interpret the PNOM diagram (a model of atomic/molecular behavior) and relate it to the graphs. Thinking again about the rubric, because the question does not ask students to provide the reasoning link between the representation and their response choice, the science practice in this case would be deemed “implicit”. In addition, in order to answer the question, students must connect the chemical equation and the PNOM diagram, two different symbolic representations of the reaction, and then translate that chemical information into a graphical representation. Raters determined that this set of tasks contains the science practice Obtaining, Evaluating, and Communicating Information (SP 8). Once again, the item contains an implicit practice because no explicit justification or explanation is prompted from the student.

Mock Item 2 (Box 3) is designed to exemplify a traditional “algorithmic” item. The item requires the calculation of the new volume of an ideal gas-filled balloon upon an increase in temperature. Because most students solve this type of problem using a mathematical algorithm, raters using the 3D-LAP rubric find it does not contain a science practice. This is not to say that these types of calculations do not have their place in assessments, but they are not considered to be an example of the practice related to using mathematical reasoning (SP 5) because students approach the problem with a goal to get a number, and they are not asked to interpret or understand that number. No further consideration of the item is carried out because it is classified as not containing a science practice.

These two mock ACS items represent items commonly found on tests from the ACS-EI, and in test banks associated with textbooks. They do not span the full variety of ACS exam
Box 2. Mock ACS exam item which incorporates a science practice.

Two chemicals, A and B₃, are placed in a container and react according to the equation:

\[ A + B_3 \rightleftharpoons AB + B_2 \]

The reaction reaches equilibrium and the chemicals present at equilibrium are diagrammed below.

Which graph best describes the reaction progress of the above reaction?
(A) \( \text{Graph A} \)  
(B) \( \text{Graph B} \)  
(C) \( \text{Graph C} \)  
(D) \( \text{Graph D} \)

Box 3. Mock ACS Exam Item Without a Science Practice

A balloon is filled with 2.0 liters of gas at 22°C. If the temperature increases from 22°C to 30°C, what would be the new volume of the balloon be at constant pressure?
(A) 2.7 L  
(B) 0.37 L  
(C) 2.1 L  
(D) 3.0 L

■ RESULTS AND DISCUSSION

Presence of Science Practices

As noted earlier, the present analysis includes 738 unique items from 12 exams that were released for use by chemistry educators. Of the 738 items analyzed with the 3D-LAP rubric, 360 (48.8%) contained at least one science practice. For items where a science practice was found, approximately 10% had more than one such practice. Importantly, the majority of science practices were rated as implicitly present (81%) rather than explicitly present (19%). While the use of science practices, as outlined in the Framework,¹ is intended to be explicit, these results from current assessment practice in chemistry suggest that there is a foundation available to build more explicit measures of science practices. Choices related to test specifications for ACS exams are made by exam development committees,² and the implicit value of science practices already present in test items could be made explicit by these groups if they choose to do so.

A distribution of the number of occurrences of the eight science practices can be seen in Table 2. Clearly, the science practice of Developing and Using Models (SP 2) occurs in ACS exam items far (five times) more often than any other science practice. Model constructs such as Lewis structures, graphs, and PNOM diagrams are frequently used in items on ACS exams. As a result, a science practice involving the use of models is at least implicitly needed in student approaches to items that contain these features. The only science practice not observed within the items analyzed was Asking Questions (SP 1). This observation is unsurprising, because multiple-choice items are seldom constructed from the perspective of asking a question. Nonetheless, this omission of the Asking Questions science practice does point to an area where greater development on multiple-choice assessments is likely warranted.

Science Practices across Exam Types

Incorporation of science practices varied by the type of exam analyzed, and Figure 1 summarizes by percentage the items on various exams with and without science practices. The LAB, GCC, and DUCK exams had relatively high percentages of items that contained a science practice, with 84%, 76%, and 68%, respectively. By contrast, the GC and GCF exams only had 25% and 21%, respectively, of items with a science practice. These differences tend to reflect design choices that were made when the exams were produced. The GCC and DUCK are somewhat newer exams and are largely intended to assess conceptual understanding of chemistry, while the GC and GCF have been produced for decades and tend to reflect traditional content coverage which usually includes a large number of algorithmic, problem solving items.

In addition to looking at science practice incorporation across exam types, it is also interesting to consider which specific science practices appear on which ACS exams. In Figure 2, the percentages of test items that contain a specific science practice, as defined by the 3D-LAP, are plotted as a histogram. Developing and Using Models (SP 2) shows the greatest incorporation on ACS exams, including on the DUCK and

Table 2. Distribution of Science Practices Found in ACS Exam Items

<table>
<thead>
<tr>
<th>Science Practices</th>
<th>Number of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Asking Questions</td>
<td>0</td>
</tr>
<tr>
<td>2. Developing and Using Models</td>
<td>176</td>
</tr>
<tr>
<td>3. Planning and Carrying Out Investigations</td>
<td>45</td>
</tr>
<tr>
<td>4. Analyzing and Interpreting Data</td>
<td>28</td>
</tr>
<tr>
<td>5. Using Mathematics and Computational Thinking</td>
<td>29</td>
</tr>
<tr>
<td>6. Constructing Explanations</td>
<td>35</td>
</tr>
<tr>
<td>7. Engaging in Argument from Evidence</td>
<td>24</td>
</tr>
<tr>
<td>8. Obtaining, Evaluating, and Communicating Information</td>
<td>23</td>
</tr>
</tbody>
</table>
GCC exams, where 40% and 36% of items, respectively, included this practice. The DUCK exam uses scenarios to present all items, and visual representations or models of data are common within the scenario, so SP 2 shows up fairly often largely because of the way this exam was designed. Practices other than SP 2 appeared much less often. Constructing Explanations (SP 6) and Engaging in Argument from Evidence (SP 7) occurred most frequently in the conceptual items from the GCC and PQ exams, but even then they appear in less than 10% of analyzed items. The practice of Using Mathematics and Computational Thinking (SP 5) requires more than being able to obtain a number on an algorithmic problem, so items with this practice also show up on the paired questions exams more than any other test. These exams pair traditional computation items with conceptual items, and the latter often require students to identify the reasoning used in otherwise computationally oriented problems. Items that were found to include the practice of Obtaining, Evaluating, and Communicating Information (SP 8) required students to translate between types of representations. Once again, the GCC exams, which contain items where translating between PNOM diagrams and chemical equations are common, provide the most cases of this practice. Except for the LAB exam, practices related to Planning and Carrying Out Investigations (SP 3), and Analysis and Interpretation of Data (SP 4), were not commonly found on tests. Other ACS exams do contain items related to the laboratory, but the multiple-choice format tends to limit the ability to design items related to experimental design. Notably, the ACS Laboratory Exam is an online only test, and leverages unique capacities related to the online format that allows for items that assess science practices.

Science Practices across ACCM Big Ideas

Over the past several years, the ACS-EI has been working with chemistry instructors to identify the content domain of assessment in undergraduate chemistry courses. The result of this process is the development of the Anchoring Concepts Content Map (ACCM).26,27 The 10 Big Ideas of the current ACCM can be argued to represent the disciplinary core ideas of chemistry. As a result, it may be useful to identify which areas of chemistry content are most likely to include science practice assessment. A subset of items (N = 465) in the current work have been aligned to the ACCM in a previous study.22 This comparison inherently involves two separate alignment processes; therefore, the assignment of items to the 10 Big Ideas reflects nuances in how items are classified by both the ACCM and the science practices on the 3D-LAP. For example, items that align with either multiple practices or multiple Big Ideas merit additional

Figure 1. Percentage of exam items with and without science practices as compared across ACS exam types.

Figure 2. Percentage of SP items by exam type distributed across the eight science practices.
The 10 Big Ideas is shown in Figure 3. In most cases, Big Ideas aligned items, there were 189 (38.7%) that contained a science mapping of science practices across content. Of the ACCM alignments of content on the ACCM yields a more robust item. Allowing multiple alignments of practices but only single ACCM is used to identify only the major content theme of the only the primary location. This choice was made so that the ACCM aligned to more than one ACCM location are identi that aligned to more than one ACCM location are identified by the total number of practices aligned to 488. By contrast, items each science practice individually. This choice e in the science practice data set to emphasize the importance of to include more than one science practice count multiple times in the discussion. For the work reported here, items that were judged to include more than one science practice count multiple times in the science practice data set to emphasize the importance of each science practice individually. This choice effectively brings the total number of practices aligned to 488. By contrast, items that aligned to more than one ACCM location are identified by only the primary location. This choice was made so that the ACCM is used to identify only the major content theme of the item. Allowing multiple alignments of practices but only single alignments of content on the ACCM yields a more robust mapping of science practices across content. Of the ACCM aligned items, there were 189 (38.7%) that contained a science practice in this analysis.

The manner in which science practices are distributed among the 10 Big Ideas is shown in Figure 3. In most cases, Big Ideas were found to include multiple science practices, with the exception of Bonding (II). The Big Idea with the largest number of SP inclusive items was Intermolecular Forces (IV) which may reflect the manner in which the use of models is incorporated rather extensively into the pedagogy of this topic. Because SP 2 was the most common science practice present overall, it also is the most prevalent in several Big Ideas including of Bonding (II), Structure and Function (III), and Visualization (X). These areas tend to lend themselves to the use of representations and models, so the number of times in which SP 2 is incorporated makes sense. The construction of explanations is SP 6, and this practice is often found in Big Ideas I and IV, related to Atoms and Intermolecular Forces, respectively. Other big ideas where explanations would seem appropriate include Structure and Function (III) and Equilibrium (VIII), and the relative lack of SP 6 in items related to this content may represent an important opportunity for future test item development efforts. Overall, the way that science practices are distributed across the disciplinary core ideas of chemistry is arguably promising from an assessment standpoint, especially when considering that the test items that were analyzed made no explicit attempt to incorporate science practices. This observation suggests that at least some science practices have long been fundamentally important in the teaching of chemistry content.

Because the analysis presented here refers to items from ACS exams, there are item statistics available to gauge student performance. In principle, these data could be informative about the relative difficulty of items that include science practices versus those that do not. ACS exams are developed as norm referenced tests, and the test development process prioritizes items that perform well within this context. Thus, the primary effect observed in any item analysis of ACS exams is most often that they tend to provide a good spread in student performances first and foremost, and other attributes are less pronounced. This observed effect is true for the analysis of performance of students on items that include science practices, such that there are hints about characteristics that might merit further study, but no conclusive differences about relative difficulty of items that incorporate science practices.

### Science Practices by Item Type

Another way to categorize items is by type. In addition to content analysis, the subset of items described previously were also classified as algorithmic (A), conceptual (C), or recall (R) questions.22,25,28 This typology does produce interesting distinctions among items with and without science practices, as shown in Figure 4. For items that do not contain a science practice (N = 283), the percentages of conceptual and algorithmic items are the same at roughly 45%. By contrast, when items include a science practice (N = 205), 79% of the time they are found to be conceptual items. While chemistry content is often considered to include both algorithmic and conceptual aspects, student performances have often been measured to be stronger on algorithmic items. The preponderance of science practices in conceptual items where lower performance is common is therefore intriguing. Recall items comprise less than 5% of items that contained a science practice. Those recall items that did include a science practice usually included a representation or model about which a student would have to recall information in order to use the representation to answer the question.

Early observations about the differences between conceptual learning and problem solving were made by Nurrenbern and Pickering29 and have been expanded upon often in the literature.30,31 Because science practices combine content knowledge with skills, the result that items that include a science practice are more often classified as conceptual items makes sense because these items tend to require higher order thinking. By definition, science practices require students to have content knowledge and also demonstrate what they can do with that knowledge. Items that need only recall and the use of algorithms are more likely to have the former component and lack the latter. This does not imply that algorithmic or recall items are unimportant in testing.
rather that student performance gleaned from testing with these categories of items is less likely to indicate what content a student understands well enough to apply that knowledge, perhaps even in novel situations.

- CONCLUSIONS

Efforts to reform curricula inevitably put pressure on assessments to change. At present, the incorporation of science practices into new assessments represents an important reform effort, one that can benefit from understanding how multiple-choice items already incorporate such concepts. The study reported here has addressed this idea by looking at a large sample of items from the ACS-EI to determine if they incorporate science practices. Roughly 50% of ACS Exam items analyzed contained a science practice, at least implicitly. While implicit incorporation of science practices is not apt to provide specific evidence of how students engage with a particular practice, this observation provides support for the premise that science practices are valued by the chemistry community. This holds true even though the exams that were studied were developed at a time when science practices were not widely known, and certainly not discussed in terms of goals for testing. This is not to suggest that science practices are currently an innate component of exam item construction or that writing items to incorporate science practices is a straightforward process. Rather, a possible foundation for incorporating science practices more explicitly into chemistry assessment endeavors may already be in place should assessment designers choose to build upon it.

At present, for ACS exams, the practice of Developing and Using Models (SP 2) is most frequently incorporated in test items. Analysis of science practices across key areas of content revealed that this practice was the only science practice to be evidenced in relation to "bonding". While this prevalence is not particularly surprising given the content of general chemistry, finding ways to incorporate other science practices more frequently, within more topics in chemistry, may be worth considering in order to enhance efforts to measure aspects of chemistry beyond chemical facts.

- IMPLICATIONS FOR TEACHING AND FUTURE WORK

ACS exams have existed for over 80 years, and have not been designed with newer curricular ideas such as science practices in mind. The present work nonetheless finds some evidence that science practices are implicit within many items on ACS exams across most key topics taught in general chemistry. Given that the exams were not designed with practices in mind, these findings suggest that an understanding of chemistry naturally includes many concepts incorporated in the practices. It is also important that ACS exams use multiple-choice formats and are used by a relatively large group of chemistry educators. This combination suggests that incorporating science practices into tests may make it possible to improve large-scale assessment with more intentional efforts to devise items that explicitly incorporate the practices.

Design of items that explicitly measure science practices remains an area, however, where more research is needed. The results reported here suggest that science practices may be included within multiple-choice items, but corroborating evidence needs to be obtained to assert that measurements that include practices test student mastery beyond the use of algorithms. Additionally, the determination of the role that specific chemistry content plays in which science practices are measurable merits more careful study.

- LIMITATIONS

There are limitations associated with the findings presented in this study that should be noted. First, ACS exams are an example of standardized multiple-choice exams, and most of the tests analyzed were for general chemistry courses. Therefore, generalizing the results found here to other item types (e.g., free response) or other areas of chemistry or science without further investigation is not merited. Because ACS exams are developed by committees of instructors, and are improved via trial testing and multiple editing rounds, the items are generally high quality. This level of quality is evident even for a large number of items that are developed but not used on released exams. Therefore, items developed with a less rigorous process might yield different results when analyzed using the 3D-LAP rubric. The conclusions from this study also focus strongly on science practices in chemistry assessments, with some attention focused on how disciplinary core ideas are incorporated alongside the science practices. An additional component of three-dimensional learning, and part of the 3D-LAP rubric, crosscutting concepts, were not part of this study. Studies that provide empirical evidence for how crosscutting concepts can...
be incorporated in to test items merit further work to help evaluate the role they play within the chemistry curriculum.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: taholme@iastate.edu.

ORCID

Thomas A. Holme: 0000-0003-0590-8848

Present Addresses

1Department of Chemistry and Biochemistry, University of Wisconsin—Milwaukee, Milwaukee, WI, 53201.
2Department of Chemistry, University of Iowa, Iowa City, IA, 52242.

Notes

Any opinions, findings, conclusions, and/or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF). The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Melanie Cooper and Sonia Underwood for their instruction and feedback on the use of the 3D-LAP rubric. The graphical abstract was drawn by Julie Adams, and her efforts are gratefully acknowledged. This work was supported by NSF DUE 1323288.

■ REFERENCES


