JOURNAL OF CHEMICAL EDUCATION

Article

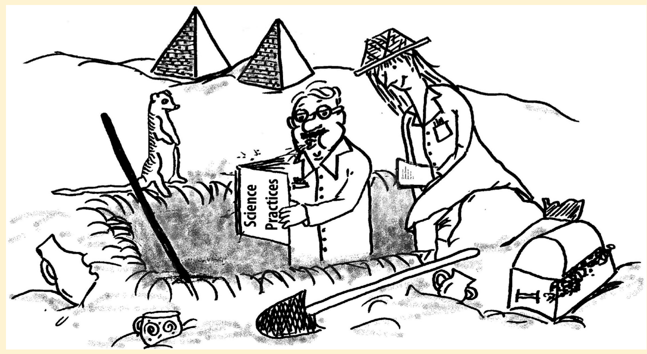
pubs.acs.org/jchemeduc

A Historical Investigation into Item Formats of ACS Exams and Their Relationships to Science Practices

Alexandra Brandriet, Jessica J. Reed, and Thomas Holme*

Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States

ABSTRACT: The release of the *NRC Framework for K-12 Science Education* and the *Next Generation Science Standards* has important implications for classroom teaching and assessment. Of particular interest is the implementation of science practices in the chemistry classroom, and the definitions established by the NRC makes these objectives much more tangible. However, this still may leave some wondering about how to begin making these changes. Mid-twentieth century chemical educators and pioneers of the first ACS exams advocated for testing science thinking and skills as early as the 1930s, and this necessitates a discussion about how early ACS exams measured these attributes. More recent debates have seen arguments that multiple-choice questions cannot measure high levels of cognitive ability in chemistry, which leaves questions about how ACS exams or instructors who write tests for large scale classrooms might try to measure science practices. The possibility that an analysis of the item formats used on ACS exams from 1934 to 1970 would help inform the creation of improved item types in testing today is investigated and presented here.



KEYWORDS: *General Public, History/Philosophy, Testing/Assessment*

■ INTRODUCTION

The recently published *National Research Council (NRC) Framework for K-12 Science Education*[1] and the *Next Generation Science Standards* (NGSS)[2,3] represent an influential set of ideas for discussions among science educators. These documents emphasize the importance of the interplay between disciplinary core ideas, crosscutting concepts, and science practices.[1−3] Of particular interest is the concept of teaching and assessing scientific practices that are central to expanding students' understandings of how scientists accomplish scientific inquiry.[1−5] The NRC Framework explicitly notes that the term *practices* is used rather than *skills* in order to emphasize that scientific development is related to both understanding and skills simultaneously.[1]

It has been argued that scientific inquiry is central to the discipline of chemistry, but the definition of inquiry often varies.[4,5] In part as a result of this variability, the NRC Framework provides a clearer picture of science practices by defining them with the skills and knowledge that are in tune with the actual conduct of science.[5] The science practices from the NRC Framework can be found as paraphrased definitions in Table 1.[1] These practices have been defined in terms of both science and engineering, but this discussion will focus more specifically on the *science* practices. It should be noted that science practices are not one specific skill or knowledge set; instead they are composed of multiple activities that collectively make up the means by which scientific discovery progresses.[1] The theory of meaningful learning suggests that students' abilities to both *understand* and *perform* must be integrated into a meaningful learning experience.[6] This is reaffirmed by the 2015 ACS Guidelines for Bachelor's Degree Programs,[7] which have indicated the importance of integrating skills that go beyond content knowledge into a college curriculum.[8] As a result, instructors may appreciate the significance of assessing science practices in the classroom and laboratory.

Relevant History of Science Education

The concepts of thinking like a scientist and gaining scientific skills have been topics of discussion among science and chemistry educators since the 19th century.[9] During the late 1800s, science education was only beginning to enter academic curriculums, and many argued its importance on the grounds that scientific thinking and skills such as "explaining the physical world" and "drawing conclusions from data" were important additions to a classical curriculum that was composed primarily of languages and grammar.[9] By the end of the 19th century, science was firmly planted within curriculums, but the courses taught were inconsistent and crowded with content.[9] By the end of World War I, a growing number of students were college-bound, which promoted a strong need for objective testing.[10] In 1921, the Division of Chemical Education within the ACS was organized, and two years later this *Journal* was founded.[11] The Division then organized the Committee on Examinations and Testing in 1930 headed by Dr. Otto Smith,[12] and the first ACS exam was released in 1934.[11,13]

Published: August 17, 2015

DOI: 10.1021/acs.jchemed.5b00459
J. Chem. Educ. 2015, 92, 1798−1806

Table 1. Science Practices and Paraphrased Definitions from the NRC Framework for K-12 Science Education[1]

Science Practice	Definition
1. Asking questions	Asking scientific questions initiates the process of investigation. Students are expected to propose questions that can be tested experimentally.
2. Developing and using models	Both mental and conceptual models are used by scientists to make predictions, process information, and explain phenomena. Students are expected to both develop and use different models in a scientific context.
3. Planning and carrying out investigations	Often discussed in the context of a laboratory, students are expected to be able to design experiments that are necessary to answer a research question or test a hypothesis. This relates to identifying the variables of interest, determining how these variables are measured (both controls and experimental), and reducing potential sources of error.
4. Analyzing and interpreting data	After data is collected, students are expected to be able to organize and interpret patterns within the data by using tables, graphs, and statistical analyses.
5. Using mathematics and computational thinking	Developing simulations, analyzing data using statistics, and using computational tools are expected in order for students to reason mathematically and make predictions.
6. Constructing explanations	Constructing explanations about the natural and physical world is a common goal in the sciences. Students are expected to use their knowledge to articulate intelligible theories and explanations that can account for underlying systems or phenomena.
7. Engaging in argument from evidence	It is common for scientists to defend their explanations and theories based on evidence. Students need to be able to coherently and logically defend their explanations as supported by their knowledge and scientific evidence.
8. Obtaining, evaluating, and communicating information	Science requires clear and persuasive communication. Students should be able to communicate their understandings verbally, in writing, using tables, graphs, and diagrams, and by having scientific discussions. Students should also be able to obtain information from scientific writing, evaluate the validity of the information, and assimilate that information.

Around the time of the early ACS exam development,[13−16] many science educators advocated for student-centered science education that aided students to develop practical skills.[9] These efforts included arguments for the adoption of the laboratory as a vital component of the chemistry curriculum.[9] By the 1930s and 1940s, it was obvious that Smith and his ACS Exams committee colleagues were also interested in the implementation of scientific skills and thinking within both teaching and assessment.[11,17−21] In 1935, Hendricks and Smith[17] (members of the early ACS exam committees) further advocated for the teaching objectives self-reported by 200 instructors, which included understanding the natural world, recognizing cause and effect, drawing generalizations from experimental data, gaining proficiency in the scientific method of thinking, and developing the ability to apply chemical principles. In this article the authors describe:

> *If these are some of our objectives how well have we succeeded in teaching our students **the significance of cause and effect, or the ability to generalize or skills as scientific thinkers**? [emphasis added]*

A 1941 memo from the ACS Examinations Institute (ACS-EI) archives written by Dr. Smith[22] regarding the plan to hold a workshop of the Committee on Examinations and Tests in Chicago notes:

> *The committee on Examinations and Tests of the ACS plans on holding a week-or-ten day conference to plan the work of the Committee for the coming two years and in particular to study the newer developments in evaluation, such as **the measurement of achievement in the ability to use the scientific method, formulate hypotheses, draw conclusions, interpret data, apply principles, and to reason logically in the chemical realm**. [emphasis added]*

This quote illustrates an interesting juxtaposition between the interests of chemical educators around the inception of the ACS Division of Chemical Education and modern trends in science education reform. Given this apparent overlap of interests, it makes sense to see how these eras in science education compare, and the existence of ACS exams provides a useful means by which this comparison can be made. ACS exams have existed for over 80 years, and their production is best described as a grassroots effort of volunteers.[23] More recently, the ACS-EI has engaged in survey research of the chemistry education community that has revealed key trends in the interests of chemistry educators.[24−32] In particular, this work has found that instruction that goes beyond chemistry content (i.e., noncontent goals) is important to instructors.[33] The combination of current interest of chemistry educators, policy statements such as the NRC Framework,[1] the NGSS,[2] and archival efforts of pioneers in chemistry education[17−22] provides an enlightening way to consider how multiple-choice (MC) items might be redefined to be capable of measuring science practices more reliably.

■ SCIENCE PRACTICES IN TESTING

Item Formats

When constructing test questions instructors must determine how to best assess course materials, and oftentimes the logistics associated with large courses can play an important role in the choices they make. Grading open-ended (OE) question responses in classrooms and laboratories with large numbers of students requires a significant amount of time,[34,35] and hand-scoring can potentially introduce additional error.[34] Further, the extra time it takes to allow students to write personal responses to OE questions may not be feasible in courses where there are time constraints due to considerable amounts of content that need to be tested.[35] As a result of these challenges, some instructors may feel compelled to choose MC item formats; however, MC testing is not as common in chemistry education as some may think.

To better understand instructors' assessment needs, the ACS-EI administered a survey to chemistry instructors from around the country.[24−29] From this survey, 1,542 instructors responded regarding their use of MC questions for course exams, and only 22.2% responded that they used *mostly* or *all* MC questions in their assessments, while 58.0% responded that they used only *a few* MC questions, and 19.7% responded that they did not use MC questions. When comparing the instructors' use of MC questions to their preferences for them, a total of 14.7% indicated that they used more MC questions than they preferred and only 2.4% indicated that they used fewer MC questions than they preferred. The remaining 82.9% were relatively satisfied in how often they used (or did not use) MC questions, but only 15.9% of the satisfied instructors used *mostly* or *all* MC questions in their assessments. These results suggest that some educators may hold negative views regarding MC question use, and there may be several reasons for these preferences. Several opinions have

DOI: 10.1021/acs.jchemed.5b00459
J. Chem. Educ. 2015, 92, 1798−1806

been stated and/or investigated in the literature regarding MC questions including that they reduce the potential for diverse and personal sets of responses,[36] they increase the possibility that students may guess the correct answers,[37,38] they cannot prepare students for the skills that they will need in their future careers,[39] and they cannot evaluate deeper dimensions of student understanding.[40,41]

Measuring Science Practices

It can be argued that science practices require a complex level of thinking that may not be accessed by rote memorization or algorithms, and writing quality MC questions that measure more than this can be very difficult. Given the number of science practices shown in Table 1 that use verbs such as construct, predict, evaluate, propose, communicate, analyze, and develop, it is unsurprising that traditional forms of MC items present challenges for assessment of these practices. Because increased student enrollment favors tests that can be electronically scored for reasons of practicality, this may leave instructors in a difficult situation. As is true for most nationally normed tests, the ACS exams released within the past decade have been composed of MC questions. However, traditional MC formats were not the most commonly used item format for early ACS exams, and these historical tests arose at a time when many of the ACS exam committee members were advocating for the implementation of scientific skills and thinking.[11,17−22]

In a recent investigation, current ACS exams were found to infrequently assess some science practices (as defined by the NRC Framework[1]), while other science practices were assessed but done so indirectly.[42] This observation is not surprising because design considerations of ACS exams do not include the measurement of science practices. These traditional ACS exam questions were coded using a rubric developed by multiple discipline-based educational researchers at Michigan State University.[43] One facet of this rubric evaluates the incorporation of science practices as defined by the NRC (Table 1).[1]

Recent work has established that examining historical ACS exam artifacts can provide insights into curriculum reform challenges,[44] and the current work expands on this premise to introduce an analysis of historical ACS general chemistry exam item formats. This work arose as a result of some challenges associated with the characterization of science practices in ACS exam MC questions.[42] Early ACS exams assessed chemistry content and ways of thinking about chemical sciences that may provide fresh insights into the assessment of higher-order cognitive abilities, like science practices, and may help inform discussions of whether or not item format inherently limits the incorporation of science practices in chemistry testing.

■ HISTORICAL ANALYSIS OF ACS EXAMS

Item formats on 20 released general chemistry ACS exams between 1934 and 1970 were analyzed. *It is extremely important to note, that because exams from this time frame are historical artifacts, far removed in time from current ACS exams, examples of specific items will be included in this publication. Items from ACS exams are generally not allowed to be published in any forum, and the assignment of example questions as "historical" as included here was adjudicated in discussions with experts in copyright law*. The 1934 and 1935 exams were referred to as provisional exams; this suggests that the committees believed that alterations would need to be made to the exams for future iterations. The provisional nature of these exams is marked with a "(P)". Each item on all 20 general chemistry exams was coded based on the item format, and this resulted in a total of 5,384 items analyzed. The item format codes were not predetermined prior to coding the items, so when each instance of a unique format was identified, a new code was added. As a result, six item formats were classified that included:

1. Multiple true/false (MTF)
2. Matching
3. Open-ended (OE)
4. Multiple-choice (MC)
5. Reasoning-tiers
6. Interpretations

The questions identified as MC were structured like the most commonly used MC format in which a question is situated within an item stem, and one response choice must be identified as the correct answer;[45] an example of this is illustrated in Figure 1. The question in Figure 1 is not found on the historical exams and, rather, is presented only to show an example of what is meant by a traditional MC question. The other item formats identified (with the exception of the OE format) could potentially fall under a broad category of MC, but for the purposes of this discussion, only traditional item formats like that shown in Figure 1 were coded as MC.

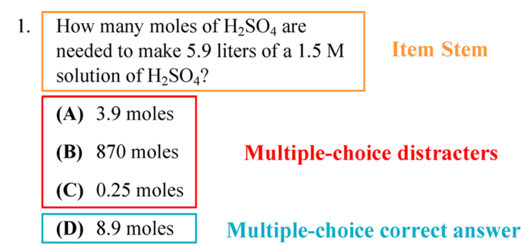


Figure 1. Example of a traditional multiple-choice item with item stem, 3 distracters, and 1 correct answer.

Similar to MC questions, MTF formats situate a question within an item stem, however, more than one response choice could possibly be correct.[46] Therefore, students had to make individual judgments about each response choice, rather than discerning which single response was the best answer, so these items may be thought of as a series of related true/false questions. An example of a MTF question can be found in Figure 2. Across this item format, multiple differences existed in how the questions probed student understanding. In some questions, two or more statements contradicted so that only certain subsets of statements could logically be together. Such is the case in Questions 10 a through d in Figure 2, where logic argues that a student could mark either a, c, or d as true, but not a combination of the three, and a student could mark b as true no matter which other responses were chosen. Because a student could theoretically mark any of the responses as true or false, the responses that fit the MTF model were coded as individual questions. In addition to the content oriented answer-tier (Questions 10 a through d in Figure 2), tests in this era often included a second part with a tier of possible reasons to justify the answer(s) to the item (Question 10 e through l in Figure 2). The reasoning-tier for this item is shown in Figure 2. Therefore, Question 10 actually is an MTF item associated with 12 responses (a through l). For the purpose of investigating quantitative trends, the questions that were part of a reasoning-tier were doubled coded as belonging to both a reasoning-tier and a specific item format, in this case the MTF format.

DOI: 10.1021/acs.jchemed.5b00459
J. Chem. Educ. 2015, 92, 1798−1806

10. The equation for the reversible reaction by which nitric oxide is prepared from nitrogen and oxygen is:

$$\text{Heat} + N_2 + O_2 \rightleftarrows 2NO$$

In one experiment nitrogen and oxygen gases are mixed using an excess of nitrogen. The gases are exposed to an electric spark at 3500°C and atmospheric pressure. In a second experiment the same quantity of oxygen is used but only enough nitrogen is used so that there is no excess of nitrogen nor of oxygen. This mixture of gases is exposed to an electric spark at 3000°C and atmospheric pressure. Will there be any difference in the reaction for the two experiments and why?

a. More nitric oxide will be formed in the first experiment..................( )
b. The reaction in the first experiment will proceed more rapidly than in the second experiment...( )
c. The amount of nitric oxide made will be the same in both experiments......( )
d. More nitric oxide will be formed in the second experiment.................( )

From the following answers select and check the ones which indicate the line of reasoning you followed in making your predictions above.

e. An excess of one of the reactants in a reversible chemical reaction shifts the equilibrium point in the direction which tends to reduce the quantity of this reactant..( )
f. An excess of nitrogen will not change the chemical composition of nitric oxide...( )
g. An excess of one of the reactants in a reversible chemical reaction shifts the equilibrium point in the direction which tends to increase the quantity of this reactant...( )
h. If greater pressure is applied to a mixture of reacting gases in equilibrium, the equilibrium point will be shifted in the direction so that those gaseous reactants or products which occupy the smaller total volume are increased in amount..( )
i. If greater pressure is applied to a mixture of reacting gases in such a way as not to increase the temperature, the speed of the reaction will be increased...( )
j. An increase of pressure on a mixture of gases does not affect the amount of the products formed by the reaction of the gases.......................( )
k. An excess of one reactant has no effect on the amount of the products formed by the reaction...( )
l. As the temperature of a reversible chemical reaction increases the equilibrium point is shifted in such a way as to increase the quantity of the substances which absorb heat in their formation......................( )

Figure 2. Multiple true/false answer- and reason-tiers from the 1935(P) ACS exam.

Matching questions were another item format identified on the historical exams. For these questions, the students had to correctly pair potential response choices with corresponding items; this type of format was sometimes used to assess chemical terminology as shown in Figure 3.

Raw materials used to	
46. prepare ordinary glass	1. Clay and limestone
47. prepare Portland cement	2. Limestone, sand, and sodium carbonate
48. make electrodes	3. Clay and sand
49. soften water containing Mg^{++} and Ca^{++}	4. Graphite and binder
50. make bricks	5. Lime and sodium carbonate

Figure 3. Matching items from the 1942 ACS exam.

In addition to matching questions, many of the early exams had a novel item format in a section labeled as *Interpretation*.[18] Interpretation items provided interesting ways to prompt students to use scientific practices and were constructed in a variety of formats, and three examples are shown in Figures 4–6. Items of this nature often included a real-world context (e.g., Figure 4) or a short description of an experiment (e.g., Figures 5 and 6) and typically asked students to either evaluate the validity of the statement/result or make judgments about the evidence associated with an assertion. Because these items required the students to choose one correct answer, they were essentially MC questions. However, given the manner in which they were used, they each contribute to the current discussion of science practices in a distinctive manner. The item shown in Figure 4 focuses student attention on the way that experimental data can be used to assert conclusions about chemical contexts. The item shown in Figure 5 has a similar net goal of reasoning from data, but requires students to gather data from multiple sources, thereby changing the cognitive demand imposed. Finally, the item shown in Figure 6 prompts students to evaluate more theoretical characteristics of argumentation by adjudicating the nature of assumptions versus factual content in a chemical context. While these prompts appear capable of eliciting different cognitive responses, they are all examples of requiring students to interpret data or argue a scientific result based on data. As a result, for analysis purposes these items were coded only as interpretation items and not MC items. For interpretation items with multiple parts (such as a through g shown in Figure 4), each statement was coded as an individual interpretation item.

Progressive Change in Historical Item Formats

Beyond providing interesting item formats that might otherwise have remained largely forgotten, the existence of these examples from historical ACS exams also provides an interesting perspective on how assessment and curricula have changed over time. The percentages of the different item formats on each general chemistry exam from 1934 through

DOI: 10.1021/acs.jchemed.5b00459
J. Chem. Educ. 2015, 92, 1798–1806

```
1 = Reasonable interpretation of the results obtained.
2 = Interpretation might be true but insufficient facts are given to justify it.
3 = Interpretation contradicted by the results obtained.

6. Eighty freshly laid chicken eggs were obtained. The whites of five of these eggs
   were examined for the hydrogen ion concentration (pH value) they contained. The
   remaining 75 eggs were stored in ordinary room air at different temperatures: 25
   eggs at 35°C, 25 at 25°C, and 25 at 15°C. After different periods of storage, five
   eggs were taken from each of the three temperature groups and examined for the
   hydrogen ion concentration (pH value) they contained. A higher pH value means a
   lower hydrogen ion concentration. The results are recorded in the table below.
```

Number of Days Eggs Were Stored	Hydrogen Ion Concentration (pH value) of the Eggs		
	35°C	25°C	15°C
0	7.60	7.60	7.60
1	9.00	8.89	....
3	9.32	9.18	9.04
7	9.50	9.39	9.28
14	9.49	9.46	9.52
21	9.37	9.38	9.48

```
a. At the end of 3 days the whites of the eggs stored at the higher temperature
   had a higher pH value than the eggs stored at the lower temperature.........(   )
b. The hydrogen ion concentration of the whites of the eggs examined at the end
   of 1 day was greater for those stored at 35°C than those stored at 25°C.(   )
c. As the relative humidity increased the pH value of the whites of the eggs
   increased...........................................................(   )
d. At the end of one week a higher pH value was found in the whites of the eggs
   stored at 15°C than in those stored at 35°C...............................(   )
e. The pH value of the whites of eggs stored at 35°C increased more rapidly than that
   of the eggs stored at 15°C during the first week of storage.................(   )
f. At the end of 21 days the whites of the eggs stored at 35°C had a lower pH
   value than the eggs stored at 15°C.........................................(   )
g. As the per cent of carbon dioxide in the air increased the hydrogen ion
   concentration of the whites of the eggs increased..........................(   )
```

Figure 4. Interpretation items from the 1935(P) ACS exam.

1970 are displayed in Figure 7. This graph illustrates how many of the earliest exams had several of these distinctive item types with MTF being the most common format for most of the exams. The 1935 exam is an exception in which the items were distributed across MC, matching, and OE formats. Unfortunately, documentation associated with the development of these historical exams no longer exists, so it is difficult to say why this exam varies so much from the other exams.

A noticeable format change also started roughly between the 1944 and 1946 exams. After 1944, item formats other than MC declined until 1965, by which point the tests were composed exclusively of MC questions. Many of the exams prior to 1946 were composed primarily of MTF questions. These questions were often structured like Question 10 in Figure 2, where students were asked to apply their chemistry knowledge to an experiment or real-world context. These items also sometimes required students to predict the resulting outcome and indicate their reasoning. Even though this item type slowly faded from use in ACS exams, when MTF are written according to expert guidelines,[46] they still have potential for chemistry assessment. For example, they can be beneficial if there is concern that the students are guessing, but they may pose difficulty for some electronic scoring systems. Further, tests that include such questions should be designed with the understanding that they are probably more time-consuming and may impose a higher level of cognitive load from the student.

The historical use of interpretation questions (shown in Figures 4−6) is also of interest. These items had the strongest connection to science practices since many focused on analyzing and interpreting raw data (e.g., Figure 4), descriptive information (e.g., Figure 6), and linear and nonlinear graphs (e.g., Figure 5). Additionally, these questions assessed students' abilities to evaluate the validity of statements, interpret the reasonableness of results, or identify the nature of the evidence associated with experimentally based assertions. As such, the interpretation format may be an especially useful tool for measuring students' abilities to analyze and interpret data (Science Practice 4 in Table 1), engage in argumentation (Science Practice 7 in Table 1) or evaluate information (Science Practice 8 in Table 1). Importantly these questions are capable of being graded automatically, as MC questions are, which suggests that they may provide a format that is better suited for the measurement of high-level cognitive skills using readily graded questions.

Focusing on chemistry content in exams, there are some content similarities between the early exams and more current exams. Many topics, such as balancing equations, naming molecules and compounds, stoichiometry, and equilibrium, are commonly found across many of the early exams. Additionally, many of the items found on these early exams (such as Questions 46−50 shown in Figure 3) were focused on rote memorization, and others were focused on the application of algorithmic problem solving. Nonetheless, beyond content the early exams are important to consider because of the specific attempts to measure scientific thinking.[18] Arguably, using multiple item formats allowed the early exam committees to focus on how to best measure all of the objectives of interest, not just those directly related to content knowledge alone. Because item format can influence the constructs measured,[46] writing test questions that assess science practices may benefit from identifying the best format for making a specific measurement. While this discussion has shown that MC item formats may potentially be used to measure high levels of cognitive abilities and skills, accepting traditional MC items as the only item format for an exam may force test writers to limit objectives measured to fit that specific type of format. The

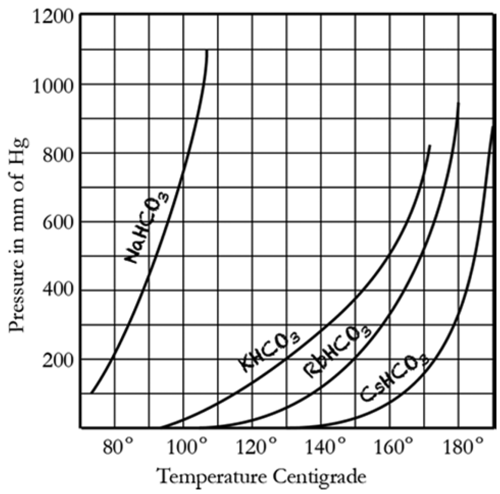
DOI: 10.1021/acs.jchemed.5b00459
J. Chem. Educ. 2015, 92, 1798−1806

When bicarbonates of the alkali metals are heated they tend to dissociate reversibly according to the reaction

$$2MHCO_3 \leftrightarrow M_2CO_3 + H_2O + CO_2$$

If this reaction is accomplished with the reaction chamber closed so the gas and vapor cannot get out, the change in pressure can be measured. Four of these bicarbonates were tested in this way and the curves at the left are a record of the tests.

Decomposition Pressure Curves



Alkali Metals and Their Properties

Element	Atomic Weight	Melting Point	Boiling Point
Lithium	6.9	186°C	1336°C
Sodium	23.0	97°C	880°C
Potassium	62.1	62°C	760°C
Rubidium	85.5	38°C	700°C
Cesium	132.8	28°C	670°C

The **curves** and **data** above indicate that:	**True (1)**	**Probably True (2)**	**Impossible to Judge (3)**	**Probably False (4)**	**False (5)**
12. **$RbHCO_3$** has a decomposition pressure of one atmosphere (**760 mm**) at **170°C**.......	( )	( )	( )	( )	( )
13. The decomposition pressure of **$NaHCO_3$** at **100°C** is **600 mm**.......................	( )	( )	( )	( )	( )
14. The decomposition pressure of **$RbHCO_3$** at **170°C** is greater than that of **$KHCO_3$**....	( )	( )	( )	( )	( )
15. **K_2CO_3** decomposes less than does **Na_2CO_3** when both are heated to the same temperature...........................	( )	( )	( )	( )	( )
16. At **130°C**, the decomposition pressure of **$RbHCO_3$** is approximately one-third that of **$KHCO_3$**...............................	( )	( )	( )	( )	( )
17. In a mixture of equivalent quantities of **$KHCO_3$** and **$CsHCO_3$** at **150°C**, more than two-thirds of the decomposition pressure would be due to **$KHCO_3$**.........	( )	( )	( )	( )	( )
18. At **90°C**, **$CsHCO_3$** has a greater decomposition pressure than **$LiHCO_3$**.....	( )	( )	( )	( )	( )
19. The decomposition pressure of an alkali bicarbonate is directly proportional to the temperature........................	( )	( )	( )	( )	( )

Figure 5. Interpretation items from the 1941 ACS exam.

variety of machine gradable item types from early ACS exams may provide insight into ways to enhance the measurement of science practices.

■ SUMMARY

The recent incorporation of science practices into science education frameworks from the NRC has the potential to make a positive impact on both chemistry curricula and assessment. However, the belief that science is defined by a set of practices has been emerging over the past 60 years,[1,47] and the concepts of scientific thinking and skills date back even further.[17,18] Early ACS exam leaders and writers specifically advocated the necessity of curricula and tests that helped students develop essential scientific skills and ways of thinking.[17−22] An investigation of historical ACS exams identified specific item formats that helped accomplish the measurement of practices. The item formats from early ACS exams were predominantly composed of MTF questions, and these items were mostly used to assess students' abilities to predict experimental results and real-world outcomes. Also of specific interest are the interpretation items like those shown in Figures 4−6, because these items were used to assess concepts that are now articulated as science practices defined within the NRC Framework.[1] The historical item types found in early ACS exams suggest that machine gradable questions may be used to assess more than low-level cognitive function.

Despite the creative nature of these early items, changes in item format were observed largely starting with the 1946 ACS exam. Ultimately, the content started to more closely resemble the standardized chemistry testing that is familiar today. Many historical events occurred during this era that may have contributed to the need to standardize testing, and this may have led to more focus on chemistry content rather than science practices. At the conclusion of World War II and with

DOI: 10.1021/acs.jchemed.5b00459
J. Chem. Educ. 2015, 92, 1798−1806

A student analyzes a clear liquid to determine whether it contains sulfate, $SO_4^{=}$. He adds a clear barium chloride solution and obtains a white precipitate. He finds that the precipitate is insoluble in hydrochloric acid. The student concludes that the unknown contains a sulfate. This conclusion, however, may or may not be valid. It is based not only upon the experimental results but also upon some assumptions, i.e., things taken for granted, as well as upon knowledge of general chemistry.

Classify each of the following statements (110-116) by marking answer space
1. if it is an assumption which **must be made** to make the conclusion valid.
2. if it is an assumption that **has no bearing** on the validity of the conclusion.
3. if it is **not an assumption** because it is a statement of fact.

110. Barium sulfate is the only salt of barium which is insoluble in water.

111. Barium sulfate is the only sulfate which is insoluble in water.

112. The barium chloride solution did not contain a sulfate.

113. Barium sulfate is the only barium salt which is insoluble in water and in hydrochloric acid.

114. Barium sulfate is insoluble in water.

115. Barium chloride is the only barium salt that produces a precipitate with a sulfate.

116. The precipitate that is obtained on adding $BaCl_2$ to the unknown will not dissolve in concentrated H_2SO_4.

Figure 6. Interpretation items from the 1946 ACS exam.

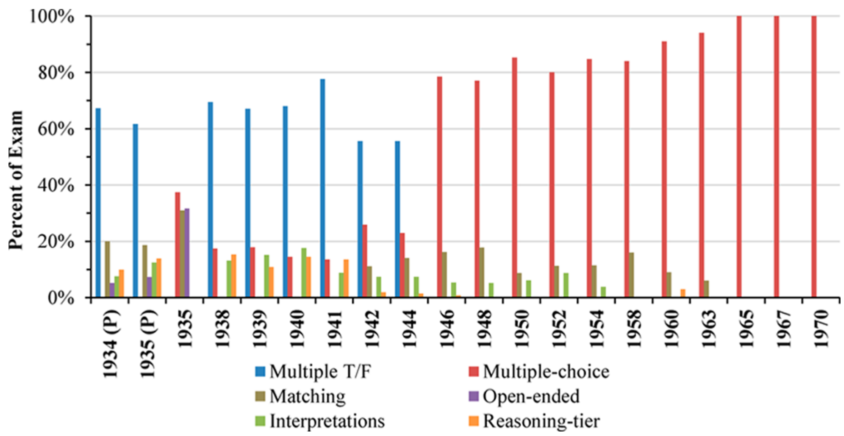


Figure 7. Percent of item formats on the historical ACS general chemistry exams. (P) indicates the provisional nature of the 1934 and 1935 exams.

the introduction of the GI Bill student enrollment in higher education increased.[11] Within the early precursors of the ACS-EI, major changes occurred during this period including collaborations with the U.S. armed forces to create military versions of the exams,[11,21,48] branching away from the Collaborative Test Services to develop exams,[11] and changes in leadership.[11,12] These developments contributed to changes in the model for testing that have survived largely intact since. Nonetheless, the earlier history of testing may provide insight anew for measuring all the things that students of today need to know.

Since the early days of the ACS Division of Chemical Education, advocates for chemistry education believed that science showed unique promise for teaching students scientific thinking and skills that have now been further conceptualized and explicitly defined by the NRC.[1] It should be emphasized that while this discussion focused explicitly on science practices, this is only one dimension of the 3-dimensional model presented by the NRC.[1] In addition to science practices, the NRC[1] proposes the importance of also interweaving disciplinary core ideas and crosscutting concepts in order to achieve meaningful learning in science and engineering.[1−3] It should not be assumed that students are gaining these understandings and practices if they are not *explicitly* taught and assessed in the classroom. However, making this type of assessment possible, particularly in large enrollment courses, presents significant challenges. In addition to using assessment development resources such as the NRC report on *Developing Assessments for the Next Generation Science Standards*,[3] examining historical ACS exams for item formats suggests ways that changing question structure may make the assessment of science practices more achievable.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: taholme@iastate.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by NSF DUE #1323288; 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).

DOI: 10.1021/acs.jchemed.5b00459
J. Chem. Educ. 2015, 92, 1798−1806

REFERENCES

(1) National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education; The National Academies Press: Washington, DC, 2012. http://www.nap.edu/openbook.php?record_id=13165 (accessed April 2015).

(2) Achieve, Inc. *Next Generation Science Standards*; National Research Council: Washington, DC, 2013. http://www.nextgenscience.org/next-generation-science-standards (accessed April 2015).

(3) National Research Council. *Developing Assessments for the Next Generation Science Standards*; Committee on Developing Assessments of Science Proficiency in K-12. Board on Testing and Assessment and Board on Science Education; Pellegrino, J. W., Wilson, M. R., Koenig, J. A., Beatty, A. S., Eds.; Division of Behavioral and Social Sciences and Education; The National Academies Press: Washington, DC, 2014. http://www.nap.edu/download.php?record_id=18409 (accessed April 2015).

(4) Osborne, J. Teaching Scientific Practices: Meeting the Challenge of Change. *J. Sci. Teach. Educ.* **2014**, *25* (2), 177−196.

(5) Cooper, M. M. Chemistry and the Next Generation Science Standards. *J. Chem. Educ.* **2013**, *90* (6), 679−680.

(6) Novak, J. D. *Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations*, 2nd ed.; Routledge: New York, NY, 2010.

(7) *Undergraduate Professional Education in Chemistry: ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs*. http://www.acs.org/content/dam/acsorg/about/governance/committees/training/2015-acs-guidelines-for-bachelors-degree-programs.pdf (accessed June, 2015).

(8) Wenzel, T. J.; McCoy, A. B.; Landis, C. R. An Overview of the Changes in the 2015 ACS Guidelines for Bachelor's Degree Programs. *J. Chem. Educ.* **2015**, *92* (6), 965−968.

(9) DeBoer, G. E. *A History of Ideas in Science Education: Implications for Practice*, 1st ed.; Teachers College Press: New York, NY, 1991; pp 1−296.

(10) Ashford, T. A. A Brief History of Objective Tests. *J. Chem. Educ.* **1972**, *49* (6), 420−423.

(11) Ashford, T. A. Contributions of the ACS Examinations Committee to Chemical Education. *J. Chem. Educ.* **1965**, *42* (9), 496−501.

(12) Ashford, T. A. Otto M. Smith−A Tribute. *J. Chem. Educ.* **1979**, *56* (12), 804.

(13) Smith, O. M.; Hendricks, B. C.; Kirk, R. E.; Phelan, E. W.; Powers, S. R.; Reed, R. D.; Wade, F. B. The 1935−36 College Chemistry Testing Program: Report of the Committee on Examinations and Tests. *J. Chem. Educ.* **1937**, *14* (5), 229−231.

(14) Phelan, E. W. The 1936−37 College Chemistry Testing Program. *J. Chem. Educ.* **1937**, *14* (12), 586−590.

(15) Foster, L. S. The 1939−1940 College Chemistry Testing Program. *J. Chem. Educ.* **1941**, *18* (4), 159−164.

(16) Ashford, T. A. The 1940−1941 College Chemistry Testing Program. *J. Chem. Educ.* **1942**, *19* (3), 116−121.

(17) Hendricks, B. C.; Smith, O. M. Service Test for Chemistry. *Sch. Sci. Math.* **1935**, *35* (5), 488−491.

(18) Ashford, T. A.; Shanner, W. M. Are We Teaching Our Students to Distinguish Between Fact and Theory? *J. Chem. Educ.* **1940**, *17* (7), 306−309.

(19) Ashford, T. A.; Shanner, W. M. Objective Test Items of the Recognition Type that Test Reasoning and Minimize Guessing. *J. Chem. Educ.* **1942**, *19* (2), 86−89.

(20) Ashford, T. A. The Problem of Chemistry in General Education. *J. Chem. Educ.* **1942**, *19* (6), 260−263.

(21) Ashford, T. A. The College Chemistry Test in the Armed Forces Institute. *J. Chem. Educ.* **1944**, *21* (8), 386−392.

(22) Smith, O. M. *Examination and Test Conference: University of Chicago*; Examination and Testing Committee Memo; ACS Examination and Testing Committee: Chicago, IL, 1941.

(23) Holme, T. Assessment and Quality Control in Chemistry Education. *J. Chem. Educ.* **2003**, *80* (6), 594−596.

(24) Emenike, M. E.; Schroeder, J. D.; Murphy, K. L.; Holme, T. A Snapshot of Chemistry Faculty Members' Awareness of Departmental Assessment Efforts. *Assess. Update* **2011**, *23* (4), 1−2, 14−16.

(25) Emenike, M. E.; Holme, T. A. Classroom Response Systems Have Not "Crossed the Chasm": Estimating Numbers of Chemistry Faculty Who Use Clickers. *J. Chem. Educ.* **2012**, *89* (4), 465−469.

(26) Emenike, M. E.; Schroeder, J.; Murphy, K.; Holme, T. Results from a National Needs Assessment Survey: A View of Assessment Efforts within Chemistry Departments. *J. Chem. Educ.* **2013**, *90* (5), 561−567.

(27) Emenike, M.; Raker, J. R.; Holme, T. Validating Chemistry Faculty Members' Self-Reported Familiarity with Assessment Terminology. *J. Chem. Educ.* **2013**, *90* (9), 1130−1136.

(28) Raker, J. R.; Emenike, M. E.; Holme, T. A. Using Structural Equation Modeling to Understand Chemistry Faculty Familiarity of Assessment Terminology: Results from a National Survey. *J. Chem. Educ.* **2013**, *90* (8), 981−987.

(29) Raker, J. R.; Holme, T. A. Investigating Faculty Familiarity with Assessment Terminology by Applying Cluster Analysis to Interpret Survey Data. *J. Chem. Educ.* **2014**, *91* (8), 1145−1151.

(30) Linenberger, K. J.; Holme, T. A. Results of a National Survey of Biochemistry Instructors to Determine the Prevalence and Types of Representations Used during Instruction and Assessment. *J. Chem. Educ.* **2014**, *91* (6), 800−806.

(31) Linenberger, K. J.; Holme, T. A. Biochemistry Instructors' Views toward Developing and Assessing Visual Literacy in Their Courses. *J. Chem. Educ.* **2015**, *92* (1), 23−31.

(32) Holme, T. A.; Luxford, C. J.; Brandriet, A. Defining Conceptual Understanding in General Chemistry. *J. Chem. Educ.* **2015**, DOI: 10.1021/acs.jchemed.5b00218.

(33) Reed, J. J.; Holme, T. A. The Role of Non-Content Goals in the Assessment of Chemistry Learning. In *Innovative Uses of Assessments for Teaching and Research*; Kendhammer, L. K., Murphy, K., Eds.; ACS Symposium Series 1182; American Chemical Society: Washington, DC, 2014; pp 147−160.

(34) Bodner, G. Statistical Analysis of Multiple-Choice Exams. *J. Chem. Educ.* **1980**, *57* (3), 188−190.

(35) Lukhele, R.; Thissen, D.; Wainer, H. On the Relative Value of Multiple-Choice, Constructed Response, and Examinee-Selected Items on Two Achievement Tests. *Journal of Educational Measurement.* **1994**, *31* (3), 234−250.

(36) Reja, U.; Manfreda, K. L.; Hlebec, V.; Vehovar, V. Open-ended vs. Close-ended Questions in Web Questionnaires. *Dev. Appl. Stat.* **2003**, *19*, 159−177.

(37) Martinez, M. A Comparison of Multiple-Choice and Constructed Figural Response Items. *Journal of Educational Measurement.* **1991**, *28* (2), 131−145.

(38) Campbell, M. L. Multiple-Choice Exams and Guessing: Results from a One-Year Study of General Chemistry Tests Designed to Discourage Guessing. *J. Chem. Educ.* **2015**, *92*, 1194.

(39) Rotfeld, H. Are We Teachers or Job Trainers? *AMS Q.* **1998**, *2*, 2.

(40) Martinez, M. E. Cognition and the Question of Test Item Format. *Educational Psychologist.* **1999**, *34* (4), 207−218.

(41) Kuechler, W. L.; Simkin, M. G. Why is Performance on Multiple-Choice Tests and Constructed-Response Tests Not More Closely Related? Theory and an Empirical Test. *Decision Sciences Journal of Innovative Education* **2010**, *8* (1), 55−73.

(42) Reed, J. J. *Analyzing the Role of Science Practices in General Chemistry Courses and Assessments*. PhD Dissertation, Iowa State University, 2015.

(43) Underwood, S. M.; Cooper, M. M.; Krajcik, J.; Caballero, D.; Ebert-May, D. *Designing a Rubric to Characterize Assessments*; presented

at the 248th ACS National Meeting and Exposition, San Francisco, CA, August 10−14th, 2014.
(44) 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* (11), 1437−1442.
(45) Towns, M. H. Guide to Developing High-Quality, Reliable, and Valid Multiple-Choice Assessments. *J. Chem. Educ.* **2014**, *91* (9), 1426−1431.
(46) Haladyna, T. M.; Downing, S. M.; Rodriguez, M. C. A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Appl. Meas. Educ.* **2002**, *15* (3), 309−334.
(47) Kuhn, T. S. *The Structure of Scientific Revolutions*; The University of Chicago Press: Chicago, IL, 1962.
(48) Ashford, T. A. The Testing Program of the Division of Chemical Education of the American Chemical Society: A Committee Report. *J. Chem. Educ.* **1948**, *25* (5), 280−290.

DOI: 10.1021/acs.jchemed.5b00459
J. Chem. Educ. 2015, 92, 1798−1806