JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Article

Development of the Investigation Design, Explanation, and Argument Assessment for General Chemistry I Laboratory

Kathryn N. Hosbein, Rosa Alvarez-Bell, Kristine L. Callis-Duehl, Victor Sampson, Steven F. Wolf, and Joi P. Walker*

Cite This: https://dx.doi.org/10.1021/acs.jchemed.0c01075 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: There have been multiple calls to incorporate the teaching of scientific practices within science laboratory courses over the past decade. To accomplish this goal, changes must be made to the curriculum standards, instructional programs, and assessment-evaluation systems used in laboratory courses. One instructional program that can used in a laboratory course to help students learn scientific practices such as investigation design, collecting and analyzing data, argument generation and critique, and science writing is the argument-driven inquiry (ADI) instructional model. This article describes the development of an end-of-course assessment, the Investigation Design, Explanation, and Argument Assessment for General Chemistry I Laboratory (IDEAA-GC1), that educators can use to measure students' ability to use scientific practices after incorporating the ADI instructional model into the General Chemistry I Laboratory. This new instrument has strong face and content validity as well as consistent instructor grading. The face validity of the instrument was established through iterative revisions of the IDEAA-GC1 based on faculty and student feedback. Content validity was established through the alignment of the IDEAA-GC1 with scientific practices and anchoring concepts as described by the Three-Dimensional Learning Assessment Protocol and the General Chemistry Anchoring Concepts Content Map.

KEYWORDS: *General Public, Chemical Education Research, Laboratory Instruction, Problem Solving/Decision Making, Testing/Assessment*

FEATURE: Chemical Education Research

The science laboratory is widely valued for its essential and distinctive role in providing students with a unique opportunity to learn how to do science. However, too often science laboratory courses are designed to illustrate established science or instruct students in rote techniques.[1] Previous studies on the effect of laboratories on undergraduate students' science learning indicate that traditional science laboratories have a negligible impact on students' science conceptual understanding.[2−8] Furthermore, a review of over two decades of educational research on science laboratories found that there is limited evidence that supports the idea that laboratory courses help students understand how to do science.[9] On the basis of these reports, the view of laboratory learning has shifted from a focus on verifying scientific concepts taught in lecture to mastering a systematic way of studying and making sense of the natural world, referred to as scientific practice.[9]

The *Framework for K12 Science Education* highlights scientific practices that professional scientists engage in as they study and make sense of the natural world.[10] These practices are intended to reflect what scientists actually do as they develop, share, and critique scientific knowledge, such as investigation design, collecting and analyzing data, and argument generation. The focus on scientific practices in science laboratory courses has led education researchers and practitioners to seek out instruments that can be used to measure changes in the ways students engage in these practices in order to examine the effect of making a change to an existing curriculum or adopting a new instructional approach.[11−13] A primary challenge with conducting this type of research is using an assessment instrument that will actually capture a target learning outcome because learning outcomes can range from basic process skills and generic laboratory techniques to critical thinking and conceptual understanding.[14,15] The assessment of a target learning outcome is further complicated by the low student to teacher ratio that results in numerous sections of a courses, which are often led by novice instructors.[16,17]

There is often an underlying supposition by those seeking to implement change in laboratory courses that the effectiveness of any teaching reform should be measured using the same instruments that are used to measure the outcome of traditional teaching methods, in an attempt to generate a "fair" comparison. However, this view often conflicts with the fundamental assumptions of the teaching reform because the new curriculum or instructional approach is often designed to help students reach a completely different outcome.[18,19] The

Received: August 14, 2020
Revised: December 8, 2020

students in the two different conditions, as a result, do not have an equal opportunity to learn the same thing. To complicate matters further, assessment of student learning often relies on selected-response (multiple-choice) questions which are useful for measuring some conceptional knowledge, but these types of items are not as useful for measuring how well a student can engage in scientific practices.[10,20] There have been attempts at the K−12 level to use hands-on kits to assess the ways students engage in scientific practices, but implementing and scoring these assessments can be burdensome.[10] Figure 1 depicts a congruence triangle where

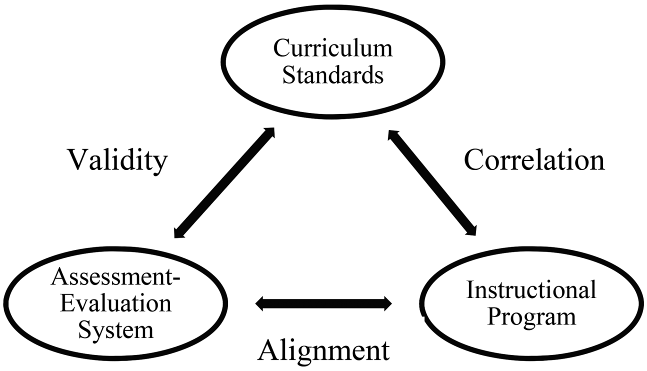


Figure 1. Congruence triangle.

standards, instruction, and assessment interact in the planning and implementing a successful educational program. If any of the three dimensions does not clearly link or interface with the other dimensions, then the fairness, credibility, validity, and utility of the assessment are compromised.[21]

A BRIEF HISTORY OF LABORATORY ASSESSMENT

Science laboratory assessments designed for traditional laboratory courses are "usually objective, paper and pencil measures".[14,17] Student learning outcomes in chemistry laboratories have been traditionally assessed through individual or group written laboratory reports and paper and pencil exams that focused on understanding concepts rather than practices.[22] There are existing assessments that measure basic process skills, generic laboratory techniques, experiment planning, or detailed observations that are made, but these assessments do not capture the actual practices because practices require the use of both knowledge and skill to make sense of the natural world.[23,24]

There have been several efforts in recent years to develop laboratory assessments that focus on more on ways students use knowledge and skills to make sense of a phenomenon. Hofstein et al., for example, developed an assessment that uses a combination of written group reports and teacher observations to examine how well students plan and carry out experiments.[25] Kirton et al. developed the *Structured Chemistry Examinations* (SChemES) to measure laboratory-based skills, and Pullen et al. developed a competency-based assessment that measures skills-acquisition, chemical knowledge and application of principles, and teamwork in a first-year laboratory program.[26,27] More recently, Stephenson et al. developed several assessment tasks that can be used to measure student proficiency with science and engineering practices.[28] These more practical-focused assessments of student learning in the laboratory, however, have not been used extensively in the literature.[23]

THEORETICAL FRAMEWORK OF SCIENTIFIC PRACTICES

Ford's description of the nature of scientific practice provides a lens to illuminate the laboratory curriculum in terms of practice.[30] Ford describes the "material practices" of science as having two distinct but complementary components: (1) practices related to manipulating nature to study aspects of it and (2) those practices that are related to "making nature's behavior apparent" to peers (p 408). This dual nature of the *practices of science* is important to recognize, as it places a premium on the role of both the natural world and the community in the enterprise of science. The first component consists of *empirical* practices such as conducting investigations and analyzing data. These are the practices that are very much a staple of undergraduate laboratory courses. The second component includes those practices that scientists engage to represent their ideas in ways that will be convincing to others in the scientific community, which we would distinguish further as *representational* and *interactive* practices. The representational aspects describe modeling and construction of explanations, while the interactive mode ultimately links the empirical and representative aspects through broader community engagement in critique, argumentation, and revision from feedback.[31]

Ford's description of scientific practices aligns with the classification of the scientific practices in the *Framework* (Figure 2). The *Framework* identifies three spheres of learning,

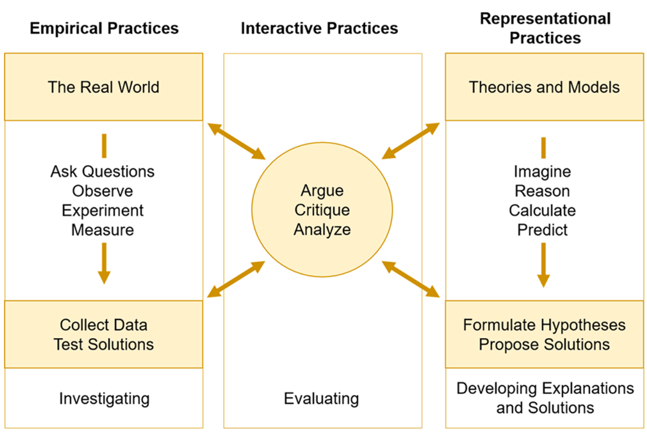


Figure 2. Alignment of theoretical framework with scientific practices. (Adapted with permission from ref 10, p 45, Figure 3.1. Copyright 2012 National Academies Press.)

the far-left sphere encompasses activities related to investigating and aligns with *empirical practices*; the right sphere comprises the practices related to developing explanations and solutions where scientists make sense of the data for themselves and the community. This sphere aligns with *representational practices*. The activities in the middle sphere connect the outer two and relate to evaluating, where scientists engage in the iterative process of developing and refining their explanation or model with the community. This sphere aligns with *interactive* practices.

As students engage in these practices, they learn to use the content knowledge associated with a scientific field to explain natural phenomena and how to develop new ideas in a manner

Table 1. Alignment of Scientific Practices with Two ADI Investigations

Scientific Practice	ADI Investigation	
	Chemical Reactions	Acid Concentration
Ask questions[a]	What is the product of the chemical reaction?	What is the concentration of acetic acid in vinegar?
Develop/use models	Chemical equations	Solution stoichiometry
Plan and carry out investigations	Design a process for collecting the product formed.	Design a process for determining molarity.
Analyze/interpret data	Comparison of actual and theoretical yields.	Compare titration results among group members.
Use mathematics and computational thinking	Calculate theoretical yields for two possible reactions.	Calculate molarity from titration data.
Construct explanations	There are two possible theoretical yields.	Commercial vinegar is 5% acetic acid.
Engage in argumentation from evidence	Construct a tentative argument.	Construct a tentative argument.
Obtain, evaluate, and communicate information	Oral argumentation, written lab report, and peer-review	Oral Argumentation and written argument.

[a]Guiding question provided in ADI instructional model.

that is consistent with the norms, discursive habits, and epistemological commitments of the discipline. Ford describes this process of learning how to use disciplinary content knowledge (e.g., physical properties, energy, chemical reactions) and how to develop new ideas in a field as developing a "grasp of practice" and argues that such an understanding is fundamental to learning science.[30]

■ ARGUMENT-DRIVEN INQUIRY

The argument-driven inquiry (ADI) instructional model was developed for high school instruction in 2005 as an alternative to the largely prescriptive laboratory activities available.[32,33] While ADI provides a central role for argument construction, it is a process-oriented model that scaffolds empirical, representational, and interactive scientific practices through investigation design, collection and analysis of data, argument generation, and science writing. In the past 15 years, research in a variety of K12 and undergraduate contexts has demonstrated positive impacts of ADI on students' science content knowledge,[32,34] engagement in scientific writing,[35,36] dialogic argumentation,[37−39] and scientific practices generally.[29] Several assessments have been developed and their data validated with criteria appropriate at the time, for research on ADI. These include assessments for argument construction[40] as well as practical exams targeting scientific practices in General Chemistry II and Physics I courses.[29,41] The practical exams were developed and evaluated for content and face validity[42,43] as well as scoring reliability between researchers in 2011 and 2019.[29]

In previous work, we have attempted to address the issues of scripted laboratory curriculum by defining the chemistry laboratory as a venue for students to learn and participate in the *practices of science* through ADI.[29,35,37,38] As with any pedagogical shift, assessment of student learning is a need and a challenge. This paper presents the development of a scientific practice-focused, disciplinary-specific laboratory exam. Herein, we describe the development of the *Investigation Design, Explanation, and Argument Assessment* for General Chemistry I (IDEAA-GC1) through the establishment of content and face validity as well as consistent assessment grading by lab instructors.

■ METHODS

Setting and Participants

ADI has been previously aligned with scientific practices.[29] Students engage in seven of the eight scientific practices described in the NGSS, with "asking questions" being the exception. A guiding question is provided for each experiment; however, students may deviate from the question, particularly in the second-semester course. Table 1 provides an example overview of the alignment of the General Chemistry I Laboratory curriculum with the scientific practices for the two ADI investigations that the IDEAA-GC1 was modeled to reflect. Student grades are based on a combination of group assignments (investigation proposal, peer-review, argumentation sessions) and individual assignments (prelab quiz, written lab reports, lab exam).

The final iteration of the IDEAA-GC1 for General Chemistry I was administered to all students enrolled in General Chemistry I Laboratory courses during week 14 of the 15 week Spring 2019 semester (N = 393) at East Carolina University (ECU). The assessment consisted of two equally weighted parts. Both parts of the laboratory final examination were completed by the student individually. Part I was a written portion that targeted the student's ability to understand and interpret experimental data. Part II was a practical assessment that required students to use common laboratory techniques to determine the concentration of a solution. All sections were a plurality of sophomore level students according to university records (ranging from 48% to 63%). The laboratory was a corequisite for the General Chemistry I lecture course. Single laboratory sections were scheduled daily in 3 h blocks. Students registered for lecture and lab separately, so there was no alignment between lecture instructor and laboratory course.

Development Process of the IDEAA-GC1

The chemistry content integrated in the practical assessment consisted of topics covered in the laboratory course curriculum, so that students could apply the scientific theories and/or concepts learned in the laboratory course to complete the practical assessment. The specific investigation used in the practical assessment was appreciably different from the related investigations, so that students did not simply employ the existing experimental design that they used in class. The specific laboratory techniques required for the practical assessment investigation were techniques students had experienced in the laboratory course. This description is aligned with the congruence triangle in Figure 1, where the curriculum standards are to develop scientific practices, the instructional program is ADI, and the assessment is the IDEAA-GC1.

As a part of a larger curriculum transformation, IDEAA practicals from biology, physics, and chemistry were designed with the intention of aligning the assessment of scientific practices across disciplines. The alignment of scientific practices on the initial iteration of the IDEAA-GC1 was

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 2. Initial Alignment of Scientific Practices and IDEAA Assessment

Scientific Practice	Practical Exam	
Ask questions[a]	Part I: What is the identity of the red powder?	Part II: What is the molarity of the base solution?
Develop/use models	Chemical equations	Solution stoichiometry[b]
Plan and carry out investigations	Design a process for collecting the product formed.	Design a process for determining molarity.
Analyze/interpret data	Comparison of actual and theoretical yields.	NA
Use mathematics and computational thinking	Calculate theoretical yields for two possible reactions.	Calculate molarity from titration data.
Construct explanations	There are two possible theoretical yields.	NA
Engage in argument from evidence	Argument with claim, evidence and reasoning.	NA
Obtain, evaluate, and communicate Information	Student notebook information provided.	Report a reasonable molarity value.

[a]Guiding question provided in ADI instructional model. [b]Not directly assessed

performed by J.P.W. This alignment was evaluated by respective discipline-based education researchers in biology and physics until the alignment was agreed upon between researchers. This alignment established initial content validity of the IDEAA-GC1.

Face validity was established using faculty and student feedback. Faculty within the chemistry department of the study institution were recruited to provide feedback on initial iterations of the IDEAA-GC1. External feedback was solicited from chemistry education researchers at three institutions (two primary undergraduate research institutions in the midwestern United States and a research-intensive university in the southern United States). Faculty input on both the assessment construction and content was requested. Undergraduate students participating in chemistry education research at the research-intensive university in the southern United States were asked to complete the questions and provide feedback to improve clarity.

To ensure scientific practices and chemistry concept knowledge were reflected in the more scaffolded final iteration of the IDEAA-GC1, questions were assessed for scientific practices as outlined by the 3D-LAP protocol[44] and anchoring concepts as outlined by the General Chemistry Anchoring Concepts Content Map (ACCM).[44,45] The Three-Dimensional Learning Assessment Protocol (3D-LAP) is a protocol designed to aid in the development of assessment items aimed to measure student chemistry reasoning based on scientific practices integrated with core IDEAAs and crosscutting concepts.[44] 3D-LAP is based off of the standards set by the *Framework*. Core ideas are concepts that are essential to the study of a discipline, such as electrostatic and bonding interactions in chemistry. Crosscutting concepts are ideas common across scientific disciplines, and scientific practices are what students should be able to do with their scientific knowledge such as structure and function. Together, these criteria encompass 3D learning, and this rubric has been used in the development of chemistry assessments.[46]

While the 3D-LAP provides one resource to define chemistry-specific content knowledge, the Anchoring Concepts Content Map (ACCM) was developed to define anchoring concepts within the general discipline and subdisciplines of chemistry.[45,47−49] The ACCMs provide a standard of chemistry "big ideas", or anchoring concepts, that align with the ACS chemistry exams.[50] The ACCMs provide maps that consist of four hierarchical levels: anchoring concepts, enduring understandings, subdiscipline articulations, and content details. Anchoring concepts are described through multiple enduring understanding statements. These first two levels consist of concepts that apply throughout the chemistry undergraduate curriculum. For example, in the General Chemistry ACCM, the anchoring concept of "Chemical Reactions: Matter changes, forming products that have new chemical and physical properties" can be described through enduring understandings such as "In chemical changes, matter is conserved and this is the basis behind the ability to represent chemical change via a balanced chemical equation". The next two levels are concepts described in terms of subdiscipline articulations followed by specific content details. From the previous enduring understanding example, a subdiscipline articulation includes "A fundamental skill for chemistry is the ability to write a balanced chemical equation". Content details that describe this subdiscipline articulation include "Chemical equations represent reactions symbolically, so they must be balanced and accurately portray reactants and products". Items to measure these anchoring concepts are designed at the content detail level and are assumed to shed light on student understanding of the overall anchoring concept. The authors of the ACCM have questioned whether or not scientific practices should be incorporated to the ACCM, but this has yet to come to fruition.[45]

While both core ideas (as defined by the 3D-LAP) and anchoring concepts (as defined by the ACCM) define chemistry content knowledge, anchoring concepts were used within this study, as the laboratory course was more so aligned with chemistry content knowledge according to the ACCM when compared to core ideas.[51] Authors J.P.W. and K.N.H. independently applied the 3D-LAP protocol to the IDEAA-GC1 for scientific practices and the General Chemistry ACCM for anchoring concepts as well as relevant enduring understandings, subdiscipline articulations, and content details. Any disagreements were discussed and resolved through consensus. The use of the 3D-LAP and ACCM provided further content validity for the IDEAA-GC1.

Consistency of Instructor IDEAA-GC1 Scores

A scoring rubric was created in the learning management system (Blackboard) that provided the answer for each question that would *meet criteria*, *partially meet criteria*, and *does not meet criteria*. The points assigned were typically 2, 1, 0, respectively (see full rubric in Supporting Information). Each laboratory section had one grader with the exception of one section that had two graders. Instructors were trained to use the IDEAA GC-1 grading rubric by scoring example assessments and discussing among each other and author J.P.W. before grading the exams for their respective sections. Examples of student responses from all three levels were provided in the training. In addition, the grading took place in the same room with oversight from the lab manager and J.P.W. As questions arose, these were addressed and shared with the group. While using a statistic to measure inter-rater reliability,

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX

such as Fleiss' κ, would be an ideal method to measure grading consistency, this was not established before the instructors graded their section's lab reports. Instead, to determine if the IDEAA GC-1 grading rubric was used consistently across sections, the total scores of each individual grader were compared using a one-way ANOVA. If the ANOVA analysis produced significant results, pairwise comparisons were explored to find the graders with total scores that were significantly different.

■ RESULTS AND DISCUSSION

Initial IDEAA-GC1 Design and Development

The IDEAA-GC1 was initially designed to capture both empirical and representational practices. Empirical practices were addressed by having students plan and execute a content-specific investigation. The representational practices were captured using argument construction for a content-specific investigation. Alignment of iteration one of the IDEAA GC-1 (Figure 4) for General Chemistry I with specific scientific practices is displayed in Table 2. Of note, this iteration was designed to elicit student proficiency of scientific practices with minimal scaffolding which resulted in broad questions.

Part I
The written portion provides data for a chemical reaction which will assess the students' ability to understand and interpret experimental data. You will need to be familiar with chemical reactions, mole ratios, chemical composition and stoichiometric calculations, and theoretical yield calculations.

Part II
The practical part of the final involves carrying out common laboratory techniques while determining the concentration of a NaOH solution using KHP. You will need to be competent with using the balance, titration, molarity, and solution stoichiometry calculations.

Figure 3. Information given to students prior to taking the IDEAA GC-1.

For part I of the assessment, students were asked to determine the identity of a dark red powder in an unlabeled bottle on the shelf with copper and copper oxides. The students were provided with information about the copper/copper(I) oxide open air oxidization reaction and some physical features of copper, copper(I) oxide, and copper(II) oxide as part of the exam. Data about the mass of the red powder and final oxidization product were also provided. Students were then asked to analyze data, generate a claim, and construct an argument where they justify the claim and the evidence.

In part two of the IDEAA GC-1, students were asked to individually determine the concentration of an NaOH solution through titration. The students were provided with background information about titration and the included chemical solutions as part of the exam. Students were then asked to design and perform an experiment to collect data, demonstrate and analyze data, and apply stoichiometric models for chemical reactions in order to provide a molarity value. The information in Figure 3 was provided to all students the week before the IDEAA-GC1.

Face Validity Using Expert Feedback and Student Interviews

The face validity for part II of the IDEAA-GC1 was established through several iterations of administration and scoring.[29] The initial draft for part I was developed from a traditional laboratory experiment.[44] Those recruited to provide feedback for part I of the IDEAA-GC1 consisted of both chemistry faculty members from ECU as well as three chemistry education researchers. Iteration one of part I of the IDEAA-GC1 that was sent to ECU faculty and chemistry education researchers is provided in Figure 4.

The suggestions from the chemistry and chemistry education research faculty focused on wording and clarification of type of answer. Example suggestions are listed below:

- *You do not ask for the unit of measure* [in question 2] *for "how much", do you want grams?*
- *From my experience with the same reaction* [in question 3c], *students struggle in determining what side to put oxygen* (O_2) *on.*
- *Are you actually looking for them to calculate the empirical formula for the initial reactant or just do comparison of the % by mass? I ask because if they have NOT already done the lab in person, then question 2a. kind of comes out of nowhere. If they've walked through it before though, then it makes a bit more sense.*

Student responses revealed misunderstanding and misinterpretation for several questions. There were common issues with not using the final heated mass for the porcelain dish or not using stoichiometry to determine the amount of copper in either moles or grams. What was most striking from the answers provided by the students was the variation in student thinking on the final question. Table 3 provides some examples of student thinking.

The analysis of the student responses to the original prompt indicated that many of the students were not using the targeted

1. Examine the data provided given on the student notebook page. Complete the table below and answer the questions.

Trial	Red Powder (g)	Mass of Product (g)
1		
2		
3		

2. How much copper was in the copper (II) oxide for each trial?
3. Based on your analysis of the data, answer the following questions.
   a) What was the composition of the red powder that was used as a reactant – was it pure copper metal powder or was it likely copper(I) oxide?
   b) How do you know? Support your claim with evidence.
   c) Write a balanced chemical equation that represents what happens when the red powder is heated to produce copper(II) oxide.

Figure 4. Iteration one of part 1 in the IDEAA-GC1 sent to faculty and a set of students for pilot testing.

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 3. Sample Student Responses to Question 3

Source[a]	Response	Evaluation
UG1	(a) The powder turns darker and then became black that is evidence that copper was in the ingredient of red powder after heating. It was not pure but it was CuO (color black), copper(I) oxide was not there, no green color was spotted (b) The color was noted black and that is evidence of CuO, no green color were captured so no copper(I) oxide. Also, CuO (black copper) has weight (63.54 Cu + 1 oxygen 15) = 79.5 g, but Cu_2O = 143 g (Cu double weight).	Student completed the table for red powder mass and product mass using two different masses for the dish. Stopped at moles of CuO rather than grams of Cu which was the goal of the question. Recognized the information about the product being black indicated that CuO had formed, but inexplicably used the lack of green color to indicate no copper(I) oxide. The ratio of masses could be useful, but not applied properly.
UG2	(a) I believe the red powder that was used as a reactant was most likely pure copper metal powder. (b) I know because of the data provided on the student's notebook page. If the final mass product was lower than the actual mass of the red powder then it would most likely be copper(I) oxide.	Student did not use the dry mass for the porcelain dish. Deduced that the mass of the product would have been less if the red powder was Cu_2O. Deduction is incorrect as it ignores stoichiometry. Both possible reactants would yield an increase in the mass of product.
UG3	(a) % copper in Cu_2O: 127.092/143.092 = 0.89 (b) Cu % in red powder; trials 0.895, 0.89, 0.89	Determined the % copper in Cu_2O, but does not compare with % copper in CuO, which suggests their answer to question a is Cu_2O. Does not justify using % copper.

[a]UG = undergraduate.

scientific practices in their responses (see Table 2). To address this issue, the questions were modified to be more explicit and to encourage students to describe their thinking in light of the target scientific practice in iteration two of part I (see Figure 5). The data provided to the students as part of the prompt was also simplified to a single table with mass of red powder + dish and mass of product + dish in order to remove the heat to constant mass variation.

Iteration two of the practical assessment was administered in the Fall of 2017 to 581 students enrolled in General Chemistry I. The lab instructors were once again trained to score the exams using a standardized rubric. The means on part I and part II of the IDEAA-GC1 were 46% (SD = 15%) and 75% (SD = 18%), respectively. Student performance on part I was lower than anticipated. In a discussion with faculty and evaluation of student answers, we identified three main issues: incorrect formulas for copper(I) oxide and copper(II) oxide, confusion over which side of the chemical reaction to put the oxygen, and mixed up values for actual and theoretical yields to use in their argument. All of these issues were addressed by adding additional scaffolding to the prompt. The third and final iteration of part I is provided in Figure 6.

This final iteration has been administered for two years using two different versions. Version 1 had data consistent with hypothesis 1 as the correct claim, and version 2 had data consistent with hypothesis 2. These two versions were used to address faculty concern over student "cheating" or the exam "getting out". To address exam security, exams are not returned to students, nor are students allowed to see their scored exam. Students may request that the score be reviewed by the laboratory manager if there is concern over instructor grading. The third iteration of the IDEAA-GC1 has been administered for two years with fairly consistent means.

Content Validity Using the General Chemistry ACCM and Scientific Practices

After establishing face validity, the most recent iterations of part I and part II of the IDEAA-GC1 were analyzed using the 3D-LAP to ensure that the practical was targeting scientific practices as originally intended. The 3D-LAP was originally used in a way to apply a single scientific practice to a cluster of questions. While it may be appropriate to teach students a performance described by a scientific practice, such as to graph a trend, it is critical to teach students how the practices interact with one another to explain a natural phenomenon.[52] Therefore, the clusters of questions within parts I and II were designed to address a single anchoring concept but multiple, interrelated, scientific practices. This was reflected in the application of the 3D-LAP.

Scientific practices within chemistry are integrated with chemistry content knowledge.[10,30] The General Chemistry ACCM was used to describe what specific anchoring concepts were integrated with the scientific practices on the IDEAA-GC1. Part I and part II of the IDEAA-GC1 were each assigned an anchoring concept. Relevant enduring understanding categories, subdisciplinary articulations, and content details were then assigned to each part and described accordingly.

IDEAA-GC1 Part I Analysis, Evidence, and Reasoning. The detailed application of the General Chemistry ACCM and 3D-LAP pertaining to the alignment of part I of the IDEAA-GC1 with anchoring concepts and scientific practices is shown in Table 4. The anchoring concept in part I was Chemical Reactions: Matter changes, forming new products that have

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX

1. Examine the data provided given on the student notebook page. Complete the table below and answer the questions.

Trial	Red Powder (g)	Mass of Product (g)	Sample Calculations Trial 1
1			
2			
3			

2. What information are you given that identifies the product as copper(II) oxide?
3. What was the source of the copper? How do you know?
4. What other element is present in the product? What was the source or sources of the other element? How do you know?
5. What mass (g) of copper was in the copper(II) oxide for Trial 1? Show your work for full credit.
6. Assuming the red powder was copper(I) oxide write a balanced chemical equation that represents what happened when the red powder was heated.
7. Determine the theoretical yield for copper(II) oxide using your equation and Trial 1 data. Show your work for full credit.
8. Using the answers above, what was the composition of the red powder that was used as a reactant – was it pure copper, copper(I) oxide or a mixture of copper oxides? How do you know? Support your claim with evidence and provide a justification for the evidence used.

Figure 5. Iteration two of part I within the IDEAA-GC1 administered in Fall 2017.

1. Examine the data provided given on the student notebook page. Complete the table below, Show your work.

Trial	Red Powder (g)	Mass of Product (g)
1		

2. What information are you given that identifies the product as copper(II) oxide (CuO)?
3. The combustion of the red powder requires another reactant. What is the chemical formula for the second reactant and what was the source of the second reactant?
4. Using your answers to questions 2 and 3, for each hypothesis write a balanced chemical equation that represents what happened when the red powder was heated.

   Hypothesis 1: Red powder is copper metal

   _____ Cu(s) + ____________ → ________________

   Hypothesis 2: Red powder is copper(I) oxide

   _____ Cu_2O(s) + ____________ → ________________

5. Based on your balanced equation for **Hypothesis 1** and the grams of red powder used in Trial I, calculate how many grams of copper(II) oxide (CuO) should have been produced, i.e. the theoretical yield? Show your work.
6. Complete the table below using your calculations for #1 and #5. The theoretical yield for Hypothesis 2 is given.

Mass of Product Trial 1	Theoretical Yield of CuO Hypothesis 1	Theoretical Yield of CuO Hypothesis 2
		1.133 g

7. What is your claim for the question – ***What is the identity of the red powder?***
8. Write an argument using your data as evidence and provide a justification for the evidence used.

Figure 6. Iteration three of part I within the IDEAA-GC1 first administered in Fall 2018.

new chemical and physical properties. Part I of the IDEAA-GC1 asked students to identify the reactant of a combustion reaction, an unknown red powder, that would result in the formation of copper(II) oxide. In order to complete this task, students needed to have a basic knowledge of how substances react within a combustion reaction and be able to take observations about this reaction along with two hypothesized balanced chemical equations to form a claim about the identity of the unknown red powder reactant. Students were required to write balanced chemical equations for two possible red powder identities that could have combusted to form copper(II) oxide. The reaction that formed copper(II) oxide was found by applying the law of conservation of mass through stoichiometric calculations using the two possible balanced chemical equations. Of note, questions 1, 2, and 6 did not particularly apply to any of the content details within the anchoring concept. These questions acted as scaffolding for students.

Multiple scientific practices were addressed within part I of the IDEAA-GC1. These practices include Analyzing and Interpreting Data, Developing and Using Models, Using Mathematical and Computational Thinking, and Constructing Explanations and Engaging in Argumentation (Figure 7). Questions 1−3 for part I addressed Analyzing and Interpreting Data. Students needed to directly assess observational data from the combustion of the red powder to determine what pieces of the data helped in the identification of the product and reactant in addition to the red powder. Question 4 addressed Developing and Using Models. Within this question, students had to investigate two hypothesized identities of the red powder through the completion of balanced chemical equations. Using Mathematical and Computational Thinking was addressed within questions 5 and 6.

The assignment of Using Mathematical and Computation Thinking to questions 5 and 6 was the only disagreement among authors K.N.H. and J.P.W. when applying the 3D-LAP to the IDEAA-GC1. Questions 5 and 6 could also be seen as Analyzing and Interpreting Data because students analyzed the data from questions 1−4 through stoichiometric calculations to find a theoretical yield. Upon further discussion, the authors decided that questions 5 and 6 assessed the practice of Using Mathematical and Computational Thinking because the

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 4. Detailed Application of the General Chemistry ACCM and the 3D-LAP to Part I of the IDEAA-GC1

Criteria	How Criteria Were Met
Anchoring Concept: Chemical Reactions: Matter Changes, Forming New Products That Have New Chemical and Physical Properties	
Relevant enduring understanding: In chemical changes, matter is conserved, and this is the basis behind the ability to represent chemical change via a BCE.[a]	Students investigate an unknown red powder through two BCEs[a] to predict the outcome of two possible reactions. Students must use the conservation of mass to compare provided data with the stoichiometric results obtained using the two balanced equations to choose the correct identity of the red powder.
	Relevant Subdisciplinary Articulations
(1) A fundamental skill for chemistry is the ability to write a BCE.	(1) Students complete two BCEs.
(2) An important observation of reactions is that mass is conserved.	(2) Students use the conservation of mass to make a claim about the red powder through the comparison of the experimental yield with two theoretical yields.
(3) The chemical equation provides key information for quantitative problem solving in stoichiometry.	(3) Students use the two BCEs to solve for two theoretical yields through stoichiometry.
	Relevant Content Details
(1a) Chemical equations represent reactions symbolically, so they must be balanced and accurately portray reactants and products.	(1a) Question 4: Students complete two BCEs.
(1b) Predicting the reactivity of chemicals is a key skill that ultimately involves the ability to write a BCE.	(1b) Question 3: Students must know the chemicals that react in a combustion reaction to answer this question and use this information in question 4. Questions 7 and 8: Students make a claim related to the identity of a red powder through the reactivity of Cu and copper(II) oxide with oxygen.
(2a) Conservation of mass in a chemical reaction can be observed in the laboratory or calculated on the basis of the masses of the reactants and products.	(2a) Question 7: Students must state the identity of the red powder based on the mass of the red powder compared to two possible theoretical yields.
(3a) Stoichiometric calculations of chemical reactions are based on mole ratios determined from BCEs.	(3a) Question 5: Students must calculate the theoretical yield for one of the BCEs in question 4.
(3b) Because chemical equations provide mole ratios, stoichiometry problems involving masses require the use of molar mass conversions.	(3b) Question 5: Students must use mole ratios and molar mass conversions to solve for theoretical yield.
Scientific Practice: Analyzing and Interpreting Data (Questions 1−3)	
(1) Question gives a scientific question, claim, or hypothesis to be investigated.	(1) The guiding question provided: What is the identity of the red powder?
(2) Question gives a representation of the data (e.g., table or graph, or list of observations) provided to answer the question or test the claim or hypothesis.	(2) A description of the experimental procedure describing combustion of a red powder to form copper(II) oxide with sample data was provided.
(3) Question gives an analysis of the data or asks student to analyze the data.	(3) In question 2, students were asked to identify what specific part of the data provided evidence that the product was copper(II) oxide. Question 3 asked students to fill in a missing piece of the data by asking what reacted with the red powder to form copper(II) oxide.
(4) Question asks student to interpret the results or assess the validity of the conclusions in the context of the scientific question, claim, or hypothesis.	(4) This is done in question 8, where students were asked to use evidence and justification to support the identity of the red powder using all of the information from part I.
Scientific Practice: Developing and Using Models (Question 4)	
(1) Question gives an event, observation, or phenomenon for the student to explain or make a prediction about.	(1) The question asked students what happened when the red powder was heated.
(2) Question gives a representation or asks student to construct a representation.	(2) The question provides the skeleton of possible chemical equations and asks students to complete the BCEs.
(3) Question asks student to explain or make a prediction about the event, observation, or phenomenon.	(3) The question asked students to make two predictions of what happened to the red powder when it was heated through two possible balanced chemical reactions.
(4) Question asks student to provide the reasoning that links the representation to their explanation or prediction.	(4) This is done in question 8, where students were asked to use evidence and justification to support the identity of the red powder using all of the information from part I.
Scientific Practice: Using Mathematics and Computational Thinking (Questions 5−6)	
(1) Question gives an event, observation, or phenomenon.	(1) Students still used the overall guiding question provided, "What is the identity of the red powder?", but at this point had completed two hypothesized balanced equations that act as their observation for this question.
(2) Question asks student to perform a calculation or statistical test, generate a mathematical representation, or demonstrate a relationship between parameters.	(2) Students were asked to use one of the hypothesized balanced equations from question 4 to calculate how many grams of copper (II) oxide would have been produced.
(3) Question asks student to give a consequence or an interpretation (not a restatement) in words, diagrams, symbols, or graphs of their results in the context of the given event, observation, or phenomenon.	(3) This is done in question 8, where students were asked to use evidence and justification to support the identity of the red powder using all of the information from part I.

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 4. continued

Criteria	How Criteria Were Met
Scientific Practice: Constructing Explanations and Engaging in Argument from Evidence (Questions 7−8)	
(1) Question gives an event, observation, or phenomenon.	(1) Questions 1−6 and the provided data described the event.
(2) Question gives or asks student to make a claim based on the given event, observation, or phenomenon.	(2) Question 7 asked students to construct a claim based on the data evaluation and analysis from questions 1−6.
(3) Question asks student to provide scientific principles or evidence in the form of data or observations to support the claim.	(3) Question 8 asked students to use evidence to support the claim used.
(4) Question asks student to provide reasoning about why the scientific principles or evidence support the claim.	(4) Question 8 asked students to provide justification for why the evidence supports the claim.

[a]BCE = balanced chemical equation.

questions were more focused on using mathematics to calculate a theoretical yield through stoichiometry rather than specifically asking students to identify pieces of the data that were pertinent to analysis.

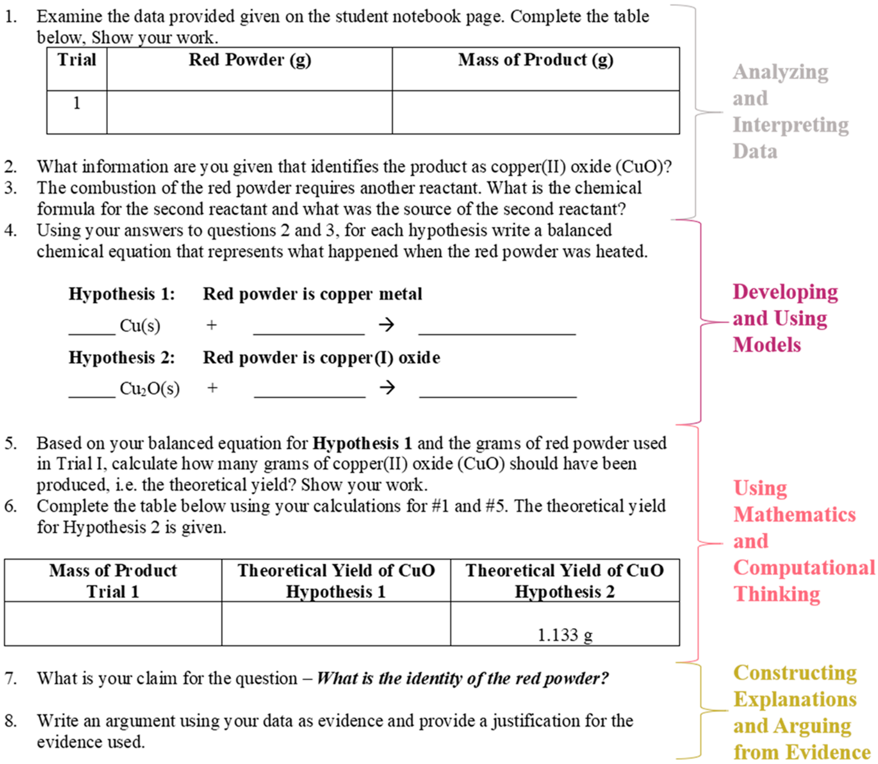
Constructing Explanations and Engaging in Argument from Evidence were assessed in questions 7 and 8. Within the original description of scientific practices in Table 2, as outlined by the NGSS, Constructing Explanations and Engaging in Argument from Evidence were described separately. Within the 3D-LAP rubric, these scientific practices were combined. Our analysis followed that of the rubric, so the two scientific practices were combined for our purposes. Within questions 7 and 8, students were asked to construct an argument consisting of a claim that was backed by evidence and justification, and the claim is considered to be the explanation of the phenomenon. Students needed to state the identity of the red powder and use evidence and justification from their results in questions 1−6. Of note, the reasoning components of all scientific practices were incorporated within question 8. Students needed to reason through the entirety of part I to come up with appropriate evidence and justification. This is an illustration of how the scientific practices are interrelated. More details on the alignment between scientific practices and specific questions can be seen in Table 4.

IDEAA-GC1 Part II Investigation Design, Data Collection, and Analysis. The detailed application of the General Chemistry ACCM and 3D-LAP to part II of the IDEAA-GC1 can be seen in Table 5. The anchoring concept addressed within part II was *Experiments, Measurement, and Data: Chemistry is generally advanced via experimental observations*. Students gained practice with performing an experiment in order to obtain an observation, in this case, a titration to obtain a more precise concentration of a sodium hydroxide solution. As students carried out the titration, they were scored on their use of lab equipment within their experimental design.

Two scientific practices were addressed within part II of the IDEAA-GC1: Planning Investigations and Using Mathematical and Computational Thinking (Figure 8). Questions 1 and 2 required students to plan and carry out an experiment to find the molarity of a sodium hydroxide solution. Questions 3 and 4 required students to use mathematical and computational thinking to calculate their sodium hydroxide solution and report the value with the correct units and significant figures. Both of these scientific practices lacked a reasoning component according to the 3D-LAP. We therefore labeled these scientific practices as peripheral. We define a peripheral science practice as skills that require content knowledge but do not require a great deal of reasoning to accomplish. Participation in peripheral scientific practices is a still an important part of doing science but should be considered a fundamental aspect of the work of professional scientisists.[52] More details of the alignment of scientific practices with each question can be seen in Table 5.

The scientific practice Developing and Using Models was not directly assessed in part II of the IDEAA-GC1 (Table 2). Students needed to Develop and Use Models as part of their molarity calculation because they had to use stoichiometry. However, because Developing and Using Models was not the main focus of any of the questions, it was ultimately decided that this section of part II could only be used to assess the practice of Using Mathematics and Computational Thinking. This issue, however, underscores how many aspects of

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX



1. Examine the data provided given on the student notebook page. Complete the table below. Show your work.

Trial	Red Powder (g)	Mass of Product (g)
1		

2. What information are you given that identifies the product as copper(II) oxide (CuO)?
3. The combustion of the red powder requires another reactant. What is the chemical formula for the second reactant and what was the source of the second reactant?
4. Using your answers to questions 2 and 3, for each hypothesis write a balanced chemical equation that represents what happened when the red powder was heated.

Hypothesis 1: **Red powder is copper metal**

_____ Cu(s) + ___________ → _______________

Hypothesis 2: **Red powder is copper(I) oxide**

_____ Cu_2O(s) + ___________ → _______________

5. Based on your balanced equation for **Hypothesis 1** and the grams of red powder used in Trial I, calculate how many grams of copper(II) oxide (CuO) should have been produced, i.e. the theoretical yield? Show your work.
6. Complete the table below using your calculations for #1 and #5. The theoretical yield for Hypothesis 2 is given.

Mass of Product Trial 1	Theoretical Yield of CuO Hypothesis 1	Theoretical Yield of CuO Hypothesis 2
		1.133 g

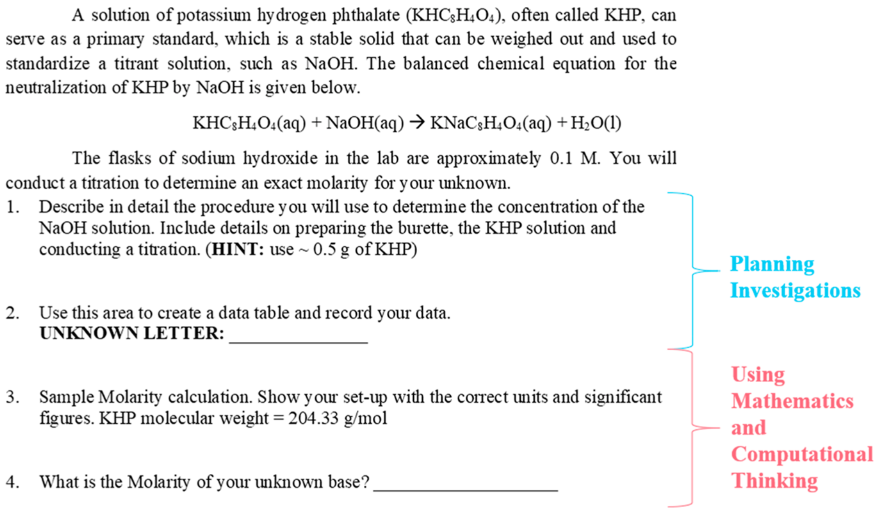
7. What is your claim for the question – ***What is the identity of the red powder?***
8. Write an argument using your data as evidence and provide a justification for the evidence used.

Figure 7. Breakdown of scientific practices in part I of the IDEAA-GC1

Table 5. Detailed Application of the General Chemistry ACCM and the 3D-LAP to Part II of the IDEAA-GC1

Criteria	How Criteria Were Met
Anchoring Concept: Experiments, Measurement, and Data: Chemistry Is Generally Advanced via Experimental Observations.	
Relevant enduring understanding: Observations are verifiable, so experimental conditions, including considerations of the representativeness of samples, must be considered for experiments.	Students are asked to determine molarity of a NaOH solution which requires a range of experimental conditions and techniques.
Relevant Subdisciplinary Articulations	
(1) Laboratory methods can be devised so that conditions limit the possibilities of measurement errors.	Students must describe a laboratory procedure to find the unknown concentration of a NaOH solution.
(2) The preparation and accurate determination of solutions and their concentrations is an important component of defining experimental conditions.	Students must perform their procedure and relevant calculations to determine the unknown concentration of a NaOH solution.
Relevant Content Details	
(1a) Techniques in setting up and carrying out laboratory experiments are an important component of instruction in introductory college chemistry.	Question 1: Students must describe a detailed procedure that they will then carry out to find the concentration of a NaOH solution.
(1b) Gaining familiarity with laboratory instruments and glassware is a key skill for the development of appropriate laboratory design methods.	Students are scored during their experimentation on the use of a balance, preparation of solutions, buret preparation, titration technique, and color of their solution at the end point.
(2a) Correct uses of volumetric glassware and/or dilution methods are important techniques in the preparation of solutions for chemical experimentation.	Students are scored during their experimentation on preparing solutions using volumetric glassware.
(2b) Determination of solution concentration is often carried out via titration with a standard.	Question 3−4: Students calculate the molarity of the NaOH solution based on their titration with a standard.
Peripheral Scientific Practice: Planning Investigations (Questions 1 and 2)	
(1) Question poses a scientific question, claim, or hypothesis to be investigated.	(1) Students are asked to determine the molarity of a NaOH solution using a primary standard, KHP.
(2) Question asks student to describe or design an investigation, or identify the observations required to answer the question or test the claim or hypothesis.	(2) Students are asked to outline the procedure for determining molarity using the titration setup provided
(3) Question asks student to justify how their description, design, or observations can be used to answer the question or test the claim or hypothesis.	(3) NA: Students carry out their outlined procedure to find the molarity of an unknown NaOH solution to answer the question.
Peripheral Scientific Practice: Using Mathematics and Computational Thinking (Questions 3 and 4)	
(1) Question gives an event, observation, or phenomenon.	(1) Students are asked to determine the molarity of an NaOH solution using a primary standard, KHP.
(2) Question asks student to perform a calculation or statistical test, generate a mathematical representation, or demonstrate a relationship between parameters.	(2) Students are asked to calculate the molarity of their NaOH solution.
(3) Question asks student to give a consequence or an interpretation (not a restatement) in words, diagrams, symbols, or graphs of their results in the context of the given event, observation, or phenomenon.	(3) NA: Students are asked to report their final NaOH molarity with correct units and significant figures.

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX


A solution of potassium hydrogen phthalate ($KHC_8H_4O_4$), often called KHP, can serve as a primary standard, which is a stable solid that can be weighed out and used to standardize a titrant solution, such as NaOH. The balanced chemical equation for the neutralization of KHP by NaOH is given below.

$$KHC_8H_4O_4(aq) + NaOH(aq) \rightarrow KNaC_8H_4O_4(aq) + H_2O(l)$$

The flasks of sodium hydroxide in the lab are approximately 0.1 M. You will conduct a titration to determine an exact molarity for your unknown.

1. Describe in detail the procedure you will use to determine the concentration of the NaOH solution. Include details on preparing the burette, the KHP solution and conducting a titration. (**HINT:** use ~ 0.5 g of KHP)
2. Use this area to create a data table and record your data.
UNKNOWN LETTER: _______________
3. Sample Molarity calculation. Show your set-up with the correct units and significant figures. KHP molecular weight = 204.33 g/mol
4. What is the Molarity of your unknown base? ___________________

Figure 8. Breakdown of scientific practices in part II of the IDEAA-GC1.

Table 6. Characteristics of IDEAA-GC1 Graders

Grader	University Position	Number of ADI Laboratories Previously Taught	Number of ADI Training Sessions Attended
A	Adjunct faculty	3	1
B	GTA[a]	3[b]	2
C	UTA	0[c]	1
D	GTA	3	1
E	UTA	2[c]	1
F	UTA	1	1
G	Adjunct faculty	2	1
H	Teaching faculty	4	1
I	UTA	1	1

[a]GTA: graduate teaching assistant. UTA: undergraduate teaching assistant. [b]Previously taught as a UTA. [c]Participated in ADI lab as an undergraduate student.

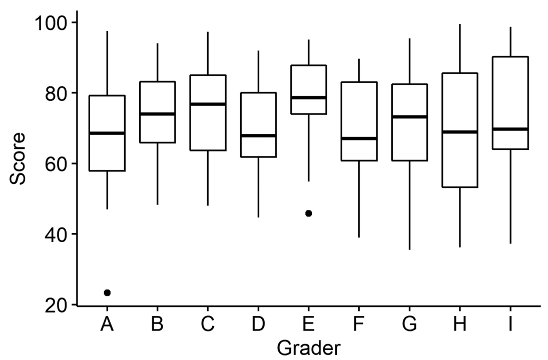


Figure 9. Box and whisker plots for individual IDEAA-GC1 graders.

scientific practices overlap. For example, students must sometimes use models as part of the practice of mathematical and computational thinking, but we cannot directly assess this scientific practice unless we directly ask them to discuss how they used a model to make sense of the phenomenon.

Consistency of Instructor Grading

One instructor of one of the sections did not use the online rubric, and therefore, the scores of that section were omitted from analysis. The characteristics of each grader included in the analyses are provided in Table 6. Students with missing exams were deleted from the data set. Exam questions with no answer were scored as "incorrect". After cleaning, there were a total of 325 completed IDEAA-GC1 practicals. Within the cleaned data, all sections had a majority of sophomore level students ranging from 53% to 65%.

The total scores for each grader are represented in box and whisker plots seen in Figure 9. A Levene test was performed to choose the most appropriate pairwise comparison statistic. The result of the Levene test revealed a violation of homogeneity of variance, $F(8) = 3.13$, $p = 0.002$. To account for this, Welch's F test was used to test differences between groups. The Welch test was significant, $F(8) = 2.79$, $p = 0.007$, suggesting there were differences present among the graders. Games–Howell post hoc tests were chosen as a the pairwise comparison statistic, as they do not assume homogeneity of variance.

Upon visual inspection of Figure 8, grader E had the highest median and smallest interquartile range compared to the other graders. This was confirmed to be significant through the post hoc comparisons. Grader E was significantly different from graders A ($p = 0.027$), D ($p = 0.016$), and F ($p = 0.024$). Course sections were assumed to be similar, with a plurality of all students listed as sophomores by the university. Grader E was teaching the course for the third time as an undergraduate teaching assistant (UTA) who had previously taken the ADI laboratory as an undergraduate student but had no obvious characteristics that would suggest this grader would be biased in any way. Although there was one grader that showed significant differences in total scores between three other graders, there were no other significant differences between graders.

■ LIMITATIONS AND FUTURE RESEARCH

This assessment was designed to target specific scientific practices and anchoring concepts included in the laboratory curriculum at the study institution. The terminology for argumentation (claim, evidence, justification) was aligned with the ADI instructional model. Given these stipulations, laboratory coordinators seeking to use a practice-focused laboratory assessment like the IDEAA-GC1 will need to consider alignment of their curriculum with this assessment. In addition, future iterations of part II of the IDEAA-GC1 may

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX

need to be modified to assess all the scientific practices as described by the 3D-LAP.

While instructor grading consistency did not justify analysis to interpret student performance on scientific practices, it did provide evidence that consistent grading among sections was achieved, which is a practical goal for new exams. Follow-up research will include interpretations of scores among students. This will be done through the comparison of researcher IDEAA-GC1 rubric scores and instructor IDEAA-GC1 rubric scores. If there is not acceptable agreement between the two, this will entail the regrading of IDEAA-GC1 by researchers to provide more valid student scores for further interpretation.

CONCLUSIONS AND IMPLICATIONS

There are limited published assessments that are designed to measure essential scientific practices. Within this research, we have presented the IDEAA-GC1; a practical laboratory assessment for General Chemistry I that measures multiple scientific practices. Content validity was established by grounding the IDEAA-GC1 within a theoretical framework consisting of Ford's description of the nature of science[30] as well as the *Framework*[10] and aligning the assessment with 3D learning through the 3D-LAP and the General Chemistry ACCM.[44,45] Face validity was shown through expert and student interviews and ensured that the questions on the exam were being interpreted in the intended way.

The IDEAA-GC1 was shown to have almost complete consistency among total scores between graders. This is unlike the results reported by our team in physics, where more than half of the graders were significantly different on the basis of pairwise tests.[41] There was a single grader (E) who produced total scores that were significantly higher than three other graders (A, D, and F). Grader E was a UTA who had previous experience participating in, and teaching, the ADI laboratory. Any suggested reasonings for the significantly higher scores would be speculation, and interviews would be needed to explore the cause. While there was a single grader who showed significantly higher total scores, all other graders had total scores that were not significantly different from each other. There can be institutional barriers to assessment that require subjective scoring due to concerns over consistent scoring with multiple instructors. We have demonstrated that when using a detailed rubric coupled with calibration of scorings, variation between graders is minimal.

Science laboratory education researchers and practitioners could design and develop their own science laboratory practical assessments with investigations related to the development of scientific practices in the curriculum. The misalignment of points on congruence triangle (Figure 1) could give the appearance that the reform teaching methods do not significantly impact learning. For example, if the curriculum standards are to develop student scientific practices, but the instructional practice consists of traditional laboratories, an assessment of scientific practices may show low student performances because the traditional laboratory was not designed to develop scientific practices. Efforts to revise undergraduate instruction will continue to meet resistance from critics and skeptics until researchers demonstrate the value of the pedagogical change by aligning curriculum standards, instructional practice, and assessments. Furthermore, if the validity and value of alternative assessments can be established, experience has shown us, in education, what gets measured gets taught (NSB, 2006). By changing what is measured and therefore prioritized, education researchers can begin to impact the teaching and learning of science in the laboratory.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.0c01075.

Rubric used to grade the IDEAA-GC1 (PDF, DOCX)

AUTHOR INFORMATION

Corresponding Author

Joi P. Walker − *Department of Chemistry, East Carolina University, Greenville, North Carolina 27858, United States;* orcid.org/0000-0001-7783-4706; Email: walkerjoi15@ecu.edu

Authors

Kathryn N. Hosbein − *Department of Chemistry, East Carolina University, Greenville, North Carolina 27858, United States;* orcid.org/0000-0001-6771-3660

Rosa Alvarez-Bell − *Department of Chemistry, East Carolina University, Greenville, North Carolina 27858, United States*

Kristine L. Callis-Duehl − *Education Research and Outreach, Donald Danforth Plant Science Center, St. Louis, Missouri 63017, United States*

Victor Sampson − *Department of Curriculum and Instruction, The University of Texas at Austin, Austin, Texas 78712, United States*

Steven F. Wolf − *Department of Physics, East Carolina University, Greenville, North Carolina 27858, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.0c01075

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Science Foundation (DUE 1725655). We would like to acknowledge the contributions of Drs. Jeffrey Raker at the University of South Florida, Dawn Del Carlo at the University of Northern Iowa, and Andrea van Duzor at Chicago State University.

REFERENCES

(1) Hofstein, A.; Lunetta, V. N. The Laboratory in Science Education: Foundations for the Twenty-First Century. *Sci. Educ.* **2004**, *88* (1), 28−54.

(2) Holmes, N. G.; Olsen, J.; Thomas, J. L.; Wieman, C. E. Value Added or Misattributed? A Multi-Institution Study on the Educational Benefit of Labs for Reinforcing Physics Content. *Physical Review Physics Education Research* **2017**, *13* (1), 010129.

(3) Elliott, M. J.; Stewart, K. K.; Lagowski, J. J. The Role of the Laboratory in Chemistry Instruction. *J. Chem. Educ.* **2008**, *85* (1), 145−149.

(4) Herrington, D. G.; Nakhleh, M. B. What Defines Effective Chemistry Laboratory Instruction? Teaching Assistant and Student Perspectives. *J. Chem. Educ.* **2003**, *80*, 1197.

(5) Hofstein, A.; Lunetta, V. N. The Role of the Laboratory in Science Teaching: Neglected Aspects of Research. *Review of Educational Research* **1982**, *52* (2), 201−217.

(6) Kirschner, P. A.; Meester, M. A. M. The Laboratory in Higher Science Education: Problems, Premises and Objectives. *Higher Education* **1988**, *17* (1), 81−98.
(7) Lazarowitz, R.; Tamir, P. Research on Using Laboratory Instruction in Science. In *Handbook of Research on Science Teaching and Learning: A Project of the National Science Teachers Association*; Gabel, D. L., Ed.; Macmillan: New York, NY, 1994; pp 94−128.
(8) White, R. T. The Link between the Laboratory and Learning. *International Journal of Science Education* **1996**, *18*, 761−774.
(9) National Research Council. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*; National Academies Press: Washington, DC, 2012. DOI: 10.1002/jee.20030
(10) National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press, 2012; Vol. *1*. DOI: 10.17226/13165.
(11) Bruck, A. D.; Towns, M. Development, Implementation, and Analysis of a National Survey of Faculty Goals for Undergraduate Chemistry Laboratory. *J. Chem. Educ.* **2013**, *90*, 685−693.
(12) Hofstein, A.; Mamlok-Naaman, R. The Laboratory in Science Education: The State of the Art. *Chem. Educ. Res. Pract.* **2007**, *8* (2), 105−107.
(13) Reid, N.; Shah, I. The Role of Laboratory Work in University Chemistry. *Chem. Educ. Res. Pract.* **2007**, *8* (2), 172−185.
(14) Bretz, S. L.; Fay, M.; Bruck, L. B.; Towns, M. H. What Faculty Interviews Reveal about Meaningful Learning in the Undergraduate Chemistry Laboratory. *J. Chem. Educ.* **2013**, *90* (3), 281−288.
(15) Bruck, A. D.; Towns, M. Development, Implementation, and Analysis of a National Survey of Faculty Goals for Undergraduate Chemistry Laboratory. *J. Chem. Educ.* **2013**, *90* (6), 685−693.
(16) Wheeler, L. B.; Maeng, J. L.; Whitworth, B. A. Teaching Assistants' Perceptions of a Training to Support an Inquiry-Based General Chemistry Laboratory Course. *Chem. Educ. Res. Pract.* **2015**, *16* (4), 824−842.
(17) Sandi-Urena, S.; Gatlin, T. Factors Contributing to the Development of Graduate Teaching Assistant Self-Image. *J. Chem. Educ.* **2013**, *90*, 1303−1309.
(18) Hickey, D.; Zuiker, S. A New Perspective for Evaluating Innovative Science Programs. *Sci. Educ.* **2003**, *87* (4), 539−563.
(19) Cooper, M. M.; Sandi-Urena, S. Twenty Years of Learning in the Cooperative General Chemistry Laboratory. In *Trajectories of Chemistry Education Innovation and Reform*; Holme, T., Cooper, M. M., Varma-Nelson, P., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2013; Vol. *1145*, pp 47−64. DOI: 10.1021/bk-2013-1145.ch004
(20) National Research Council. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academy Press: Washington, DC, 2001.
(21) Doran, R.; Chan, F.; Tamir, P. *Science Educators Guide to Assessment*; National Science Teachers Association: Arlington, VA, 1998.
(22) Hofstein, A. The Laboratory in Chemistry Education: Thirty Years of Experience with Developments, Implementation, and Research Laboratory. *Chem. Educ. Res. Pract.* **2004**, *5*, 247−264.
(23) Kempa, R. F.; Ward, J. E. The Effect of Different Modes of Task Orientation on Observational Attainment in Practical Chemistry. *J. Res. Sci. Teach.* **1975**, *12* (1), 69−76.
(24) Ben-Zvi, R.; Hofstein, A.; Samuel, D.; Kempa, R. F. The Effectiveness of Filmed Experiments in High School Chemical Education. *J. Chem. Educ.* **1976**, *53* (8), 518.
(25) Hofstein, A.; Levy Nahum, T.; Shore, R. Assessment of the Learning Environment of Inquiry-Type Laboratories in High School Chemistry. *Learning Environments Research* **2001**, *4* (2), 193−207.
(26) Kirton, S. B.; Al-Ahmad, A.; Fergus, S. Using Structured Chemistry Examinations (SChemEs) As an Assessment Method To Improve Undergraduate Students' Generic, Practical, and Laboratory-Based Skills. *J. Chem. Educ.* **2014**, *91* (5), 648−654.
(27) Pullen, R.; Thickett, S. C.; Bissember, A. C. Investigating the Viability of a Competency-Based, Qualitative Laboratory Assessment Model in First-Year Undergraduate Chemistry. *Chem. Educ. Res. Pract.* **2018**, *19* (2), 629−637.
(28) Stephenson, N. S.; Duffy, E. M.; Day, E. L.; Padilla, K.; Herrington, D. G.; Cooper, M. M.; Carmel, J. H. Development and Validation of Scientific Practices Assessment Tasks for the General Chemistry Laboratory. *J. Chem. Educ.* **2020**, *97* (4), 884−893.
(29) Walker, J. P.; Sampson, V.; Southerland, S.; Enderle, P. J. Using Laboratory to Engage All Students in Science Practices. *Chem. Educ. Res. Pract.* **2016**, *17* (17), 1098−1113.
(30) Ford, M. J. Disciplinary Authority and Accountability in Scientific Practice and Learning. *Sci. Educ.* **2008**, *92* (3), 404−423.
(31) Chi, M. T. H.; Wylie, R. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychology* **2014**, *49* (4), 219−243.
(32) Sampson, V.; Walker, J. P.; Grooms, J. Argument-Driven Inquiry: A Way to Promote Learning during Laboratory Activities. *Science Teacher* **2009**, *76* (8), 42−47.
(33) Walker, J. P.; Sampson, V.; Zimmerman, C. Argument-Driven Inquiry: An Introduction to a New Instructional Model for Use in Undergraduate Chemistry Labs. *J. Chem. Educ.* **2011**, *88* (10), 1048−1056.
(34) Walker, J. P.; Sampson, V.; Grooms, J.; Zimmerman, C.; Anderson, B. Argument-Driven Inquiry in Undergraduate Chemistry Labs: The Impact on Students' Conceptual Understanding, Argument Skills, and Attitudes toward Science. *Journal of College Science Teaching* **2012**, *41* (4), 74−81.
(35) Walker, J. P.; Sampson, V. Using the Laboratory to Improve Undergraduates' Science Writing Skills through Meaningful Science Writing, Peer-Review and Revision. *J. Chem. Educ.* **2013**, *90* (10), 1269−1274.
(36) Sampson, V.; Walker, J. P. Learning to Write in the Undergraduate Chemistry Laboratory: The Impact of Argument-Driven Inquiry. *International Journal of Science Education* **2012**, *34* (10), 1443−1485.
(37) Walker, J. P.; Sampson, V. Learning to Argue and Arguing to Learn: Argument-Driven Inquiry as a Way to Help Undergraduate Chemistry Students Learn How to Construct Arguments and Engage in Argumentation during a Laboratory Course. *J. Res. Sci. Teach.* **2013**, *50* (5), 561−596.
(38) Walker, J. P.; Van Duzor, A. G.; Lower, M. A. Facilitating Argumentation in the Laboratory: The Challenges of Claim Change and Justification by Theory. *J. Chem. Educ.* **2019**, *96* (3), 435−444.
(39) Grooms, J.; Sampson, V.; Enderle, P. How Concept Familiarity and Experience with Scientific Argumentation Are Related to the Way Groups Participate in an Episode of Argumentation. *J. Res. Sci. Teach.* **2018**, *55* (9), 1264−1286.
(40) Walker, J. P.; Sampson, V.; Grooms, J.; Zimmerman, C. A Performance-Based Assessment for Limiting Reactants. *J. Chem. Educ.* **2011**, *88* (8), 1243−1246.
(41) Wolf, S.; Sprague, M.; Li, F.; Smith-Joyner, A.; Walker, J. P. Introductory Physics Laboratory Practical Exam Development: Investigation Design, Explanation, and Argument. In *The PER Conference Proceedings*; Provo, UT, 2019.
(42) Arjoon, J. A.; Xu, X.; Lewis, J. E. Understanding the State of the Art for Measurement in Chemistry Education Research: Examining the Psychometric Evidence. *J. Chem. Educ.* **2013**, *90* (5), 536−545.
(43) Barbera, J.; VandenPlas, J. R. All Assessment Materials Are Not Created Equal: The Myths about Instrument Development, Validity, and Reliability. *Investigating classroom myths through research on teaching and learning* **2011**, *1074*, 177−193.
(44) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS One* **2016**, *11* (9), No. e0162333.
(45) Holme, T.; Luxford, C.; Murphy, K. Updating the General Chemistry Anchoring Concepts Content Map. *J. Chem. Educ.* **2015**, *92* (6), 1115−1116.

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX

(46) Underwood, S. M.; Posey, L. A.; Herrington, D. G.; Carmel, J. H.; Cooper, M. M. Adapting Assessment Tasks To Support Three-Dimensional Learning. *J. Chem. Educ.* **2018**, *95* (2), 207−217.
(47) Raker, J.; Holme, T.; Murphy, K. The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map II: Organic Chemistry. *J. Chem. Educ.* **2013**, *90* (11), 1443−1445.
(48) Marek, K. A.; Raker, J. R.; Holme, T. A.; Murphy, K. L. The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map III: Inorganic Chemistry. *J. Chem. Educ.* **2018**, *95* (2), 233−237.
(49) Holme, T. A.; Reed, J. J.; Raker, J. R.; Murphy, K. L. The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map IV: Physical Chemistry. *J. Chem. Educ.* **2018**, *95* (2), 238−241.
(50) 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* (6), 715−720.
(51) Cooper, M. M.; Posey, L. A.; Underwood, S. M. Core Ideas and Topics: Building Up or Drilling Down? *J. Chem. Educ.* **2017**, *94* (5), 541−548.
(52) Ford, M. J. Educational Implications of Choosing "Practice" to Describe Science in the Next Generation Science Standards: EDUCATIONAL IMPLICATIONS OF CHOOSING "PRACTICE. *Sci. Educ.* **2015**, *99* (6), 1041−1048.

https://dx.doi.org/10.1021/acs.jchemed.0c01075
J. Chem. Educ. XXXX, XXX, XXX−XXX