JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Chemical Education Research

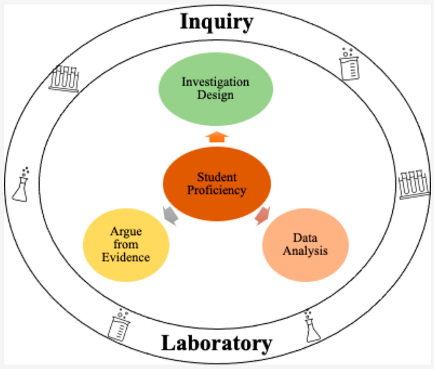
Assessment of Scientific Practice Proficiency and Content Understanding Following an Inquiry-Based Laboratory Course

Kathryn Hosbein and Joi Walker*

Cite This: J. Chem. Educ. 2022, 99, 3833−3841 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: For decades, the case has been made for inquiry-based activities that enable teaching and learning of science concepts through the process of doing science. Laboratory courses provide a unique setting with opportunities for students to learn to ask questions, plan and carry out investigations, analyze data, and construct scientific arguments. Yet, most postsecondary laboratory courses still rely heavily on confirmatory laboratory activities. One of the problems regarding the implementation of inquiry-type laboratory experiments is the issue of assessing students' achievement to demonstrate pedagogical effectiveness. The design and scoring consistency of the Investigation Design, Explanation, and Argument Assessment for the first semester of General Chemistry Laboratory (IDEAA-GC1) has been described previously. The IDEAA-GC1 is a practical laboratory assessment for General Chemistry I that measures student ability to design and conduct an investigation, analyze and interpret data, and construct an argument. This study examines the impact of an inquiry-based laboratory course in developing student proficiency with two scientific practices: (1) plan and carry out an investigation and (2) generate a scientific argument. This manuscript presents the analysis of student responses (N = 325) to the assessment which indicated proficiency with both scientific practices to varying degrees as well as the limitations based on student content knowledge.



KEYWORDS: *First-Year Undergraduate/General, Chemical Education Research, Laboratory Instruction, Testing/Assessment, Gravimetric Analysis, Titration/Volumetric Analysis*

INTRODUCTION

Science education researchers and practitioners have begun to place increasing emphasis on scientific practices by integrating authentic scientific investigations into science laboratory curricula to improve students' habits of mind, capabilities to carry out scientific investigations, and engagement in scientific reasoning.[4] These transformed curricula require assessments aligned with the change in instructional practice and the cognitive theories employed. Hofstein et al., in their 2019[1] review of inquiry-type laboratories found that assessment of student learning is a primary barrier to widespread transformation of laboratory courses (ref 1, page 522):

> *One of the most crucial problems regarding the implementation of inquiry-type laboratory experiments is the issue of assessing students' achievement and progress in such a unique learning environment. In general, large numbers of science teachers do not engage in authentic and practical assessment on a regular basis.*

Unfortunately, there are limited published assessments designed to measure student achievement and proficiency in laboratory courses, which has resulted in a call for increased research on laboratory courses generally.[5,6] Large enrollment courses rely heavily on multiple choice exams; however, recent research into the student responses on these exams has revealed that the inferred connection between a correct response and student knowledge is inconsistent at best and incorrect at worst.[7,8] The alternative to selective response questions, is an open response assessment that can provide greater insight into students' disciplinary knowledge and ability to do science. Proficiency with scientific practices alone does not address the knowledge students gain from participation in an inquiry laboratory. The notion of inquiry in teaching science can serve to stimulate both content understanding and science process understanding.[2] As faculty and teachers seek to assess proficiency with scientific practices, they do so using disciplinary content. If not addressed directly, there can be an incongruity between assessment of scientific practices and assessment of content detail.[9]

In a previous study, student performance was reported in a community college setting on a practical exam for second-semester general chemistry. The student scores were used to

Special Issue: New Visions for Teaching Chemistry Laboratory

Received: June 13, 2022
Revised: October 22, 2022
Published: November 14, 2022

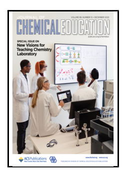


Table 1. Scientific Practice Alignment with ADI Investigations

Type of Scientific Practice	ADI Investigation	
	Chemical Reactions	Acid Concentration
	There are two possible theoretical yields.	Commercial vinegar is 5% acetic acid.
Empirical		
Ask questions[a]	What is the product of the chemical reaction?	What is the concentration of acetic acid in vinegar?
Plan an investigation	Design a process for collecting the product formed.	Design a process for determining molarity.
Carry out an investigation	Conduct the reaction of copper(ii) sulfate in aqueous solution with solid iron filings.	Conduct a titration using a standardized sodium hydroxide solution.
Analyze and interpret data	Calculate theoretical yields for two possible reactions. comparison of actual and theoretical yields.	Calculate molarity from titration data.
Representational		
Present findings to peers.	Construct and present a tentative argument.	Construct and present a tentative argument.
Construct a written argument	Oral argumentation, written lab report, and peer-review.	Oral argumentation and written argument.

[a]ADI provides a guiding question for each investigation.

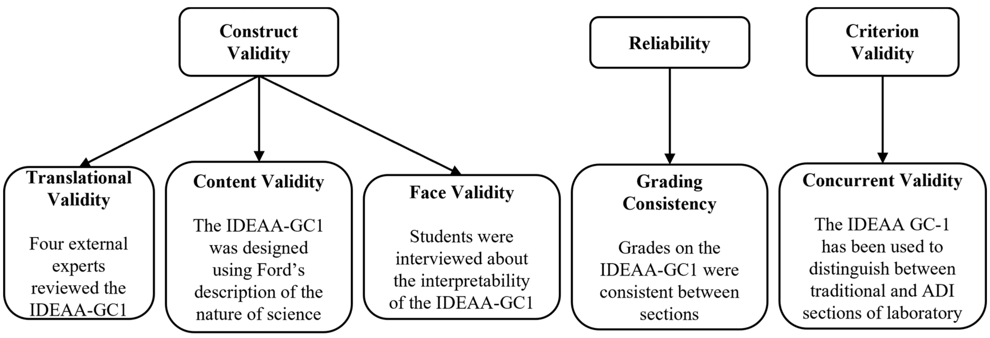


Figure 1. Validity and reliability evidence previously presented for the IDEAA-GC1

determine the extent to which the type of instruction employed during a general chemistry laboratory course affected students' proficiency with scientific practices.[10] The assessment targeted student proficiency with planning and conducting investigations, analyzing and interpreting data, and arguing from evidence. The mean scores on the practical exam were significantly higher in inquiry-based sections for all students including female students, under-represented minority (URM) students, and students with lower past academic achievement. The 2017 study compared total scores without delving into the details of student responses. This manuscript, takes a more granular approach, looking at strengths and weakness in student ability to do science and how that may be impacted by their chemistry understanding.

THEORETICAL FRAMEWORK

This study is grounded in a social constructivist theory of learning; a perspective on learning founded on the basic assumption that knowledge is actively built or "constructed" by the learner based on interaction with others.[11,12] Considering social constructivist views of learning, the chemistry laboratory can be viewed as a setting where students can interact directly with the material world, using the tools, data collection techniques, models, and theories of science. A discipline specific laboratory course, therefore, presents a unique platform for providing students an opportunity to participate in and develop proficiency in the scientific practices while simultaneously building disciplinary knowledge. Ford's description of the nature of scientific practices was used to categorize what students do in the laboratory as either empirical or representational.[13] Ford describes the "material practices" of science as having two distinct but complementary components. The first component is practices related to manipulating nature to study aspects of it, and the second component is those practices that are related to "making nature's behavior apparent" to peers (ref 13, p. 408). The first component consists of practices such as conducting investigations and analyzing data, while the second component includes those practices that scientists use to communicate findings with the community, such as argumentation.

ARGUMENT-DRIVEN INQUIRY

In previous work, the issues of expository laboratory curriculum[5,14] were addressed by defining the chemistry laboratory as a venue for students to learn and participate in the scientific practices through Argument-Driven Inquiry (ADI).[10,15−17] The ADI instructional model is designed to give a more central place to argumentation and the role of argument in the social construction of scientific knowledge while promoting inquiry. The multiweek investigations are designed to integrate the learning of scientific concepts with inquiry, argumentation, and writing in such a way that students gain proficiency through engagement in the laboratory investigations moving from investigation design to analysis and argument development, to argumentation and written argument (Table 1). ADI describes scientific argumentation using a claim, evidence, reasoning framework where students make a *claim* that they support with *evidence* and provide *reasoning* as to how the evidence supports the claim. In the ADI laboratories, the claim is in response to a question and is supported by data that students collect and analyze. The argument should include a justification (reasoning) of the

https://doi.org/10.1021/acs.jchemed.2c00578
J. Chem. Educ. 2022, 99, 3833−3841

evidence that fully connects the claim to the evidence using the relevant scientific concept.

■ INVESTIGATION DESIGN, EXPLANATION, AND ARGUMENT ASSESSMENT (IDEAA)

The Investigation Design, Explanation, and Argument Assessment (IDEAA) was designed to measure scientific practices specific to the ADI laboratory such as investigation design and argument construction.[3] The development and scoring consistency of the *Investigation Design, Explanation, and Argument Assessment* for the first semester of General Chemistry Laboratory (IDEAA-GC1) was published previously.[3] Figure 1 presents the steps taken for validity and reliability according to the credibility framework adapted from Trochim, used for the IDEAA-GC1.[18]

The instrument was sent to four external experts for review to establish translational validity. Content validity was established by grounding the IDEAA-GC1 within a theoretical framework consisting of Ford's description of the nature of science[13] as well as the *Framework*[4] and aligning the assessment with scientific practices and chemistry content knowledge using the 3D-LAP rubric and the General Chemistry Anchoring Concepts Content Map (ACCM).[19,20] Face validity was shown through student interviews and ensured that the questions on the exam were being interpreted in the intended way. Evidence for criterion validity has been provided multiple times through concurrent validity. The assessment has distinguished between types of laboratory courses, traditional vs inquiry, in multiple contexts.[10]

The student responses for each question on the IDEAA-GC1 were scored *correct, partially correct, or incorrect,* which provided insight into student thinking beyond the more common assessment practice of marking answers right or wrong. In addition, scoring considered student responses to previous questions that may have not been fully correct, i.e., student reasoning was deemed correct if the reasoning was correct in the context of their incorrect value in a previous question. Establishing the IDEAA-GC1 as a credible assessment was a rigorous multiyear endeavor. This work was published separately from the student results presented here, and readers interested in that process are encouraged to read the assessment development paper.[3]

Part one (Figure 2) of the IDEAA-GC1) was focused on arguing from evidence. Students were asked to determine the identity of the given dark red powder in an unlabeled bottle on the shelf with copper and copper oxides. The information about open air oxidization reactions and some physical features of copper, copper(I) oxide, and copper(II) oxide were provided along with data about the mass of the red powder and final oxidization product as a student notebook page (Supporting Information). Students were asked to analyze data, generate a claim, and construct an argument where they justify the claim and the evidence.

Part two of the IDEAA-GC1 (Figure 3) was focused on planning and conducting an investigation. Students were asked to determine the concentration of a NaOH solution through titration. The background information about titration and included chemical solutions are provided as shown in Figure 3. Students were asked to design and perform an experiment, to collect and analyze data, and to calculate a reasonable molarity value.

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 2. Part 1 of the IDEAA-GC1.

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 3. Part 2 of the IDEAA-GC1 with scientific practices labeled.

■ RESEARCH QUESTIONS

1. To what extent do students that participate in an inquiry-based laboratory develop proficiency with arguing from evidence and planning and conducting an investigation?
2. To what extent does student chemistry content knowledge impede their ability to demonstrate proficiency with scientific practices?

■ METHODS

Data Collection and Participants

The IDEAA-GC1 was administered to all students enrolled in General Chemistry 1 Laboratory courses during week 14 of the 15-week Spring 2019 semester (N = 393). All sections were a plurality of sophomore level students according to university records (ranging from 48% to 63%). The laboratory is a corequisite for the General Chemistry Lecture course. Laboratory sections were scheduled daily in 3-h blocks, students registered for lecture and lab separately, so there is no alignment between lecture instructor and laboratory course.

https://doi.org/10.1021/acs.jchemed.2c00578
J. Chem. Educ. 2022, 99, 3833−3841

The general chemistry course introduces students to the basic principles and laws of chemistry. Topics include measurements, reactions, and stoichiometry, thermochemistry, atomic structure, periodicity, bonding, and molecular structure, and states of matter. A scoring rubric for the IDEAA-GC1 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 2, 1, 0, respectively. The full rubric with details on scoring is available in the Supporting Information.

Inter-rater Reliability

In order to provide item-level analysis and interpretation of the IDEAA-GC1, each rubric item needed to show evidence of validity and reliability. Reliability of the scoring rubric for the IDEAA-GC1 was previously explored and shown to have consistent grading by General Chemistry 1 instructors.[3] This type of reliability was used to investigate the practicality of using the rubric in practice, but reliability for each rubric item was not explored. To establish reliability for each rubric item, each score was transformed to "correct", "partial credit", and "incorrect" so that item was on a consistent scale. Rubric items that were deemed purely scaffolding were not included in this analysis, such as rubric items that addressed "calculations shown". Next, inter-rater reliability (IRR) was established in two phases (Figure 4).

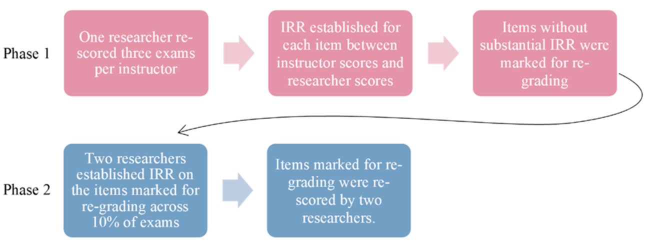


Figure 4. Process of establishing IRR for each rubric item.

During the first phase, a single researcher rescored three IDEAA-GC1 from each instructor. The three exams pulled for rescoring consisted of high, mid, and low total scores. A linearly weighted Cohen's kappa,[21,22] referred to as kappa from here on, between the instructor-scored exams and the researcher was calculated for each rubric item. Rubric items were marked for rescoring if an item did not show substantial agreement (≤0.7) through kappa.[22] The second phase consisted of rescoring the items marked for regrading. Before marked items were rescored, two researchers established IRR. To do this, three IDEAA-GC1 practicals of each version from each section were randomly selected to rescore by the researchers (10% of total data set). IRR was established once a kappa equal to or above 0.7 was reached on each rescored item. Once IRR was established between the two researchers, all subsequent rubric items marked for rescoring were rescored.

Fidelity of Implementation of ADI

Argument-driven inquiry has been the instructional model in the general chemistry curriculum since 2017. New faculty and teaching assistants attend a full-day workshop prior to fall semester that provides an overview of the research on the teaching and learning of science. The new instructors work in groups on a sample lab investigation that allows them to experience each stage of the ADI model. Throughout the semester all instructors attend orientation sessions for each investigation. When conducting research that relies on multiple individual's implementation of a teaching model, it is necessary to confirm that fidelity of implementation. The ADI-Specific Observation Protocol (ASOP) was developed to measure *fidelity of implementation* of the ADI instructional model.[23] The complete protocol is provided in the Supporting Information. The ASOP was used to observe each instructor for one three-week cycle of ADI. During one three-week investigation, two members of the research team observed the same instructor to establish inter-rater reliability. Any differences between the two observers were resolved through discussion.

Table 2. Descriptive Statistics for Rubric Items Shown to Measure Scientific Practices (N = 325)

Central Scientific Practice	Question	Rubric Item Description	Incorrect (%)	Partially Correct (%)	Correct/Difficulty (%)
Constructing an Argument	1.1	Subtracts mass of dish	2	3	95
	1.2	Product color	12	48	40
	1.3	Other element present	13	34	53
	1.4	Hypothesis 1 balanced[a]	21	29	50
		Hypothesis 2 balanced	20	38	42
	1.5	Theoretical yield	41	15	44
	1.7	Claim[a]	22	23	55
	1.8	Evidence	39	27	34
		Justification[a]	26	50	24
Planning and Carrying out an Investigation	2.1	Procedure	39	27	34
	2.2	Mass of KHP	5	16	79
		Final and initial mL	1	4	95
		mL NaOH used	12	7	82
		Volume recorded to 0.01 mL	37	12	51
		Clear organization	5	3	92
	2.3	Calculate molarity	16	17	67
	2.4	Accuracy of molarity .002	28	46	26

[a]Items rescored by researchers.

https://doi.org/10.1021/acs.jchemed.2c00578
J. Chem. Educ. 2022, 99, 3833−3841

RESULTS AND DISCUSSION

Data Cleaning and Participants

An instructor of one of the sections did not use the online rubric for scoring the IDEAA-GC1, and therefore the scores of that section were omitted from analysis. 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 IDEA assessments representing 87% of the enrolled students (N = 393).

ASOP Protocol

The ASOP scores for the 10 instructors were analyzed to determine the facilitation techniques and the time allotted on relevant tasks. All instructors were observed to maintain primarily student-centered techniques, i.e., asking guiding questions, checking on whole class or group progress, allowing students to independently analyze data. There were some instances of marginal facilitation in which instructors provided some aid and offered suggestions for data analysis. The argumentation sessions were allowed sufficient time for students to engage in meaningful argument from evidence followed by postargument discussions. The peer-reviews were conducted by the students following a brief calibration for uniform scoring. Based on these results, there were no instances of instructors deviating from the intended pedagogy.

Descriptive Statistics and Rubric Item Difficulty on Final Scores

The descriptive statistics of items used to measure the two scientific practices are listed in Table 2. Item difficulty was assessed by percent of students that received a correct score for each question. The higher is the percentage of correct answers, the lower is the difficulty.

IDEAA-GC1 Part 1 Construct an Argument from Evidence

The representational practice of generating an argument from evidence consisted of students making a claim (Table 2, Q 1.7) and backing the claim with evidence and justification (Table 2, Q 1.8). Figure 5 indicates 78% of the responses were proficient

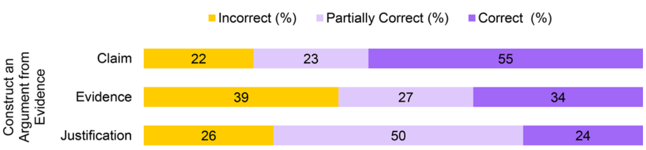


Figure 5. Distribution of responses for questions 1.7 and 1.8.

in constructing a correct or partially correct claim. Responses for using evidence and justifying the evidence indicated a wider range of proficiency.

The challenges students had with writing the correct balanced chemical reactions for each hypothesis are illustrated in the partially correct and incorrect responses observed for roughly 50% of the student responses (Table 2, Q 1.4). This question was confounded with writing correct formulas for diatomic oxygen and the correct formula for copper(II) oxide which was provided indirectly in the reaction description provided to students.

$$\text{Hypothesis 1:} \quad Cu_{(s)} + O_{2(g)} \rightarrow 2CuO_{(s)}$$

$$\text{Hypothesis 2:} \quad 2Cu_2O + O_{2(g)} \rightarrow 4CuO_{(s)}$$

The role of mole ratios in stoichiometric calculations was confounded by the misbalanced equations; however, the questions were scored based on the student's balanced equation. Ratios were often "flipped" which is likely a result of students incorrectly applying a mathematical model rather than the incorrect product (CuO) and reactant (Cu or Cu_2O).[24,25] The required stoichiometric calculations to obtain the theoretical yield, were problematic with 41% (Table 2, Q 1.5) of the students having an incorrect value for theoretical yield. The unique challenges students have with stoichiometry have been well documented,[8,26−28] yet at most institutions these calculations are a cornerstone of first-semester general chemistry.

Examples of student arguments with scoring rationale are presented in Table 3. In scoring the written arguments correct use of the incorrect values was considered (Table 3, example A), but these may have been so far afield as to prevent any reasonable response. Partial or incorrect scores tended to use incomplete evidence, i.e., not explicit in the values used for comparison of actual product yield to theoretical yield (Table 3, examples C−F). Similarly, partial or incorrect justification responses did not explain the relevance of theoretical yield to answering the question (Table 3, examples B and D).

The written argument required more than just determining the correct theoretical or percent yield. The challenge of scientific argumentation is the *justification* of the evidence that fully explains how the claim and evidence are connected using a scientific concept. In other words, students had to use their understanding of the limits of percent yield as justification for using yield as evidence to support their claim. The challenge for students to provide scientific justification is well documented in our research[17,29] and in argumentation research collectively.[30−32] This comes down to not just doing a yield calculation, but connecting the meaning of yield to the claim in a manner that makes sense.

IDEAA-GC1 Part 2 Planning and Carrying out an Investigation

Figure 6 presents the student proficiency with the empirical practice, Planning and Carrying out an Investigation. The 40% incorrect responses on the procedure (Table 2, Q 2.1) was based on students responses to the prompt "Describe in detail the experimental method you will use to determine the concentration of the NaOH solution." The student responses generally lacked detail, such as properly preparing the buret, removing the funnel from the buret, checking for air bubble in the tip. The percent incorrect, reflected that the method described was missing several key details.

The representational practice of communicating data was assessed through Question 2.2. The ADI instructional model rarely provides data tables for students to complete, rather students construct their own systems for recording data. The majority of students designed appropriate data tables and recorded data with the correct significant figures and units. Three examples of student work are shown in Figure 7 and illustrate the range of data tables students created. The most common error was to record titration volumes with too few decimal places. Students tend to leave off zeros, recording 49 mL rather than 49.00 mL as in example B. The correct recording of buret volume is the focus of the prelab activity for investigation and students are advised to review "use of a buret" on the exam information. Students were allowed two titrations, which is seen in example C, but replicates were not

https://doi.org/10.1021/acs.jchemed.2c00578
J. Chem. Educ. 2022, 99, 3833−3841

Table 3. Examples of Student Arguments and Scoring

	Student Argument	Scoring[a]	
		Evidence	Justification
A.	The powder used in the experiment was copper metal. After calculating a theoretical yield of 0.2832 g for Hypothesis 2, the mass of the product was compared to both theoretical yields for hypothesis 1 and hypothesis 2. The mass of the product was 1.0750 g which is over the theoretical yield, 0.28324, for hypothesis 2. Also, when calculating % yield (actual yield/theoretical yield) 100), the % yield for hypothesis 2 was 379.53% and was thus ruled out because it was invalid due to the being over 100%. It is impossible to create more product than the theoretical yield and therefore hypothesis 2 was ruled out. The mass of the product, 1.0750 g, was less than the 1.2758 g theoretical yield for Hypothesis 1 and thus best fit Hypothesis 1. Hypothesis 1 was if the red powder was copper metal, therefore, the red powder was copper metal.	2	2
		The theoretical yield of 0.28324 g was incorrect, but the argument based on the data and the student calculations was correct.	
B.	The red powder is most likely not copper metal, as the theoretical yield (1.251 g) does not closely match the weight of the product (1.185 g). However, the theoretical yield of copper(I) oxide (1.1138 g) is much closer to the final amount of product, meaning the powder in the unmarked bottle is most likely copper(I) oxide.	2	1
		Student does not explain why actual and theoretical yields count and why being "closer" is important.	
C.	The red powder is copper metal, because the physical yield of the experiment was 1.185g in comparison with the theoretical yield from copper(I) oxide. The actual yield cannot be more than the theoretical yield.	1	2
		Student used grams of product as the only evidence. Student does provide a justification for using actual and theoretical yields.	
D.	Copper metal is the unknown powder because based off the theoretical yield and actual yields a percent yield can be found by dividing the actual yield by the theoretical yield to determine which reactant formed the closest product yield it was intended to. The percent yield for hypothesis 1, copper metal came to 84% while the percent yield for copper oxide came to 58%. 84% is closer to the 100% suggesting that the unknown is the copper metal reactant.	1	1
		Student only used % yield values as evidence and did not show or discuss this calculation. The inference that percent yield should be closer to 100% is not justified.	
E.	Based off the information provided and the calculations, copper(II) oxide was produced due to a combustion reaction. There was 1.055 g produced which is calculated by theoretical yield. With this being known, it is logical to say that the unknown powder is copper(I) oxide.	1	0
		Student does not justify using actual and theoretical yields as evidence.	
F.	The unknown powder is CuO because the percent yield is 84.3% which is high, but still accounts for the lost mass in the investigation.	0	0
		Student used a single % yield value as evidence and did not show or discuss this calculation. The reasoning behind the justification for the % yield is unclear.	

[a]Scores were 0—incorrect, 1—partially correct, 2—correct.

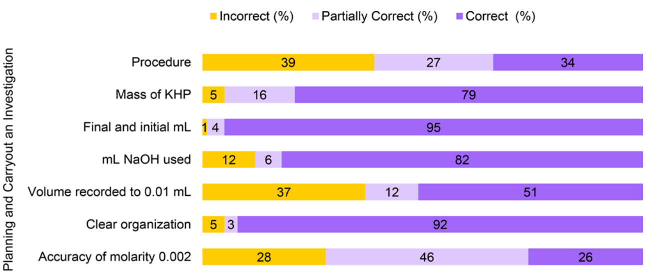


Figure 6. Distribution of responses for questions 2.1, 2.2, and 2.4.

required due to time constraints. Still, more than a third of the students did not grasp the distinction between a number and a measured value, which speaks to a much larger issue that needs to be addressed in the laboratory.

The 46% partially correct and 26% correct accuracy of molarity (Q 2.4, Table 2) was based on the instructor's calculation of the molarity using the student's titration data, which prevented a 2-fold penalty from calculation of molarity (Table 2, Q 2.3). This finding, 72% of students conducted a respectable titration, is noteworthy. The issues students had with using stoichiometry in this calculation again highlights the need to focus on conceptual understanding of stoichiometric calculations within the scientific practice rather than mathematical manipulations.[8,26−28]

■ LIMITATIONS

Practitioners

This assessment was designed for specific scientific practices and content represented within the laboratory curriculum at the study institution. The terminology for argumentation (claim, evidence, justification) is aligned with the ADI instructional model. Given these stipulations, laboratory coordinators seeking to use a practice-focused laboratory assessment will need to consider alignment of their curriculum with this assessment.

Researchers

This research was performed in a single context, at a single university. Further studies should be performed to investigate

https://doi.org/10.1021/acs.jchemed.2c00578
J. Chem. Educ. 2022, 99, 3833−3841

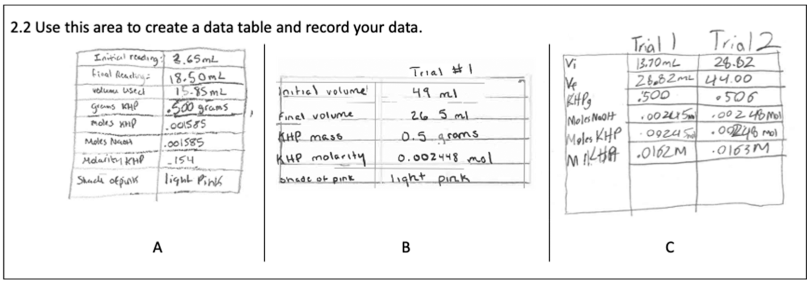


Figure 7. Examples of student data tables.

the validity and reliability of IDEA assessment data in other contexts to further support its use in measuring scientific practices. The short- and multistep open-ended questions on the IDEA assessment allowed for students to provide a variety of answers as compared to a multiple-choice assessment. Expanding the scoring to include partially correct, allowed us to capture more nuance in student answers as well as to give credit for less than perfect responses.

■ CONCLUSIONS AND IMPLICATIONS

There are limited published science laboratory assessments designed to measure student achievement and progress from participation in inquiry-type laboratories. This lack of authentic assessment has been identified as a significant barrier to widespread laboratory transformation.[1,3] This research presented student performance on the IDEAA-GC1, a practical laboratory assessment for General Chemistry 1 that measures proficiency with both empirical and representational scientific practices[13] using disciplinary content knowledge. The goal in designing this assessment was to scaffold the questions such that student thinking moved from generalized concepts into finer levels of detail within an activity that demonstrated proficiency with a scientific practice. This assessment seeks to infer both what students "know" and what students are able to "do". The reliability of this inference relies upon structure of that task.[27] Returning to the research questions:

1. To what extent do students that participate in an inquiry-based laboratory develop proficiency with arguing from evidence and planning and conducting an investigation?
2. To what extent does student chemistry content knowledge impede their ability to demonstrate proficiency with scientific practices?

The detailed analysis addresses both questions and generally establishes that students do gain proficiency with the targeted scientific practices, but the limitations imposed by student content knowledge can be confounding. By assessing student chemistry concept knowledge and scientific practice proficiency in this way, it is possible to suggest ways to improve student understanding in future iterations of the course, such as focusing more on the conceptual aspects of stoichiometry or to change course learning outcomes to reflect the current research on stoichiometric calculations.[8,28]

The scores on the IDEAA-GC1 highlighted ways that the practical exam could be modified. The *Constructing an Argument from Evidence* items (Figure 5) suggested that students were more proficient with providing a claim than they were with evidence and justification. This is consistent with previous research.[15] These results have led to adjustment of the IDEAA by removing the second version for Part I, and using the data related to hypothesis 1 which is consistent with the student lab experiment. In addition, there have been changes to the related laboratory investigation so that students do not calculate theoretical yield until after they conduct the experiment. A question has been added to Part II of the assessment that asks students to explain or justify the use of an indicator which will also accomplish the goal to further understand student conceptual thinking.

Design and validation of the IDEAA began in 2016 and therefore represents change in assessment and scoring practices that may be seen as insufficient by chemistry education researchers. However, this assessment has been instrumental in bringing about assessment change within the department and the institution where it was developed. In 2016, there were many who questioned student ability to complete a practical exam as well as the ability to administer and reliably score such an assessment. This manuscript presenting student performance combined with the development manuscript[3] presents a path toward assessment reform in a system that is often resistant to change. Indeed, the physics and biology departments developed their own versions of the IDEAA which are more directly aligned with an ADI investigation, requiring students to plan and carry out an investigation on a topic not covered directly in their laboratory course and then use their own evidence to generate a written argument.[33] The lead PI on this project developed similar assessments for a study comparing course-based undergraduate research experiences with traditional laboratory courses in organic and analytical chemistry.[34] These assessments in other disciplines and chemistry courses have enabled science education researchers to make the case for reform broadly and resulted in expansion of the inquiry-based laboratory courses and research-intensive courses across multiple disciplines and colleges at the study institution. This broader impact of assessment practice has and is resulting in institutional transformation that would not be possible without direct evidence of student knowledge and proficiency.

Implications for Practitioners

These types of assessments are challenging to develop and to administer; however, the benefit is what can be learned about student thinking and proficiency. Researchers of 3D LAP have provided several publications to guide and inform development of such assessments through transformation of existing exam items.[7,9,26] This manuscript suggests that using a combination of single answer questions followed by questions that provide students an opportunity to not just indicate what they know, but how they know may provide meaningful insight into student content understanding and proficiency with scientific practices.

https://doi.org/10.1021/acs.jchemed.2c00578
J. Chem. Educ. 2022, 99, 3833−3841

Implications for Researchers

Design of this specific assessment took place over a five year period.[3] The process was complicated by concurrent curriculum development and the alignment of exam questions across disciplines. Assessments that integrate more than one concept are often more complex, and it will be more difficult to establish the standard psychometrics.[7] Therefore, there is a tension between the need for reliable/valid assessments and the need for those assessment items to measure more than facts, algorithmic calculations, and pattern recognition. Rather than use psychometric analyses, validation relied on expert feedback, student responses, and the in-depth contextual analysis presented here.[3] If researchers would like to use the IDEAA-GC1 to assess scientific practices, additional validity evidence should be collected within their specific contexts.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.2c00578.

ASOP protocol, the IDEAA-GC1 scoring rubric, and the student notebook page for Part 1 (PDF,DOCX)

■ AUTHOR INFORMATION

Corresponding Author

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

Author

Kathryn Hosbein − *Department of Chemistry, Middle Tennessee State University, Murfreesboro, Tennessee 37132, United States;* orcid.org/0000-0001-6771-3660

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.2c00578

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. Rosa Alvarez-Bell and Eli Hvastkovs as well as all the laboratory instructors who supported this project.

■ REFERENCES

(1) Hofstein, A.; Dkeidek, I.; Katchevitch, D.; Nahum, T. L.; Kipnis, M.; Navon, O.; Shore, R.; Taitelbaum, D.; Mamlok-Naaman, R. Research on and Development of Inquiry-Type Chemistry Laboratories in Israel. *Isr. J. Chem.* **2019**, *59* (6−7), 514−523.

(2) National Research Council. *Inquiry and the National Science Education Standards*; National Academy Press: Washington, DC, 2000.

(3) Hosbein, K. N.; Alvarez-Bell, R.; Callis-Duehl, K. L.; Sampson, V.; Wolf, S. F.; Walker, J. P. Development of the Investigation Design, Explanation, and Argument Assessment for General Chemistry I Laboratory. *J. Chem. Educ.* **2021**, *98* (2), 293−306.

(4) National Research Council. *A Framework for K-12 Science Education: Practices, Croscutting Concepts, and Core Ideas*; National Academies Press, 2012; *1*. DOI: 10.17226/13165.

(5) Bretz, S. L. Evidence for the Importance of Laboratory Courses. *J. Chem. Educ.* **2019**, *96* (2), 193−195.

(6) 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.

(7) 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.

(8) Ralph, V. R.; Lewis, S. E. An Explanative Basis for the Differential Performance of Students with Low Math Aptitude in General Chemistry. *Chem. Educ. Res. Pract.* **2019**, *20* (3), 570−593.

(9) 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.

(10) 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.

(11) Bransford, J.; Brown, A.; Cocking, R.*How People Learn: Brain, Mind, Experience and School*; Bransford, J., Brown, A., Cocking, R., Series Eds.; National Academy of Science Press: Washington, DC, 1999.

(12) Vygotsky, L. S.*Mind in Society. The Development of Higher Psychological Processes*; Harvard University Press: Cambridge, 1978.

(13) Ford, M. J. Disciplinary Authority and Accountability in Scientific Practice and Learning. *Science Education* **2008**, *92* (3), 404−423.

(14) Sansom, R.; Walker, J. P. Investing in Laboratory Courses. *J. Chem. Educ.* **2020**, *97* (1), 308−309.

(15) 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. *Journal of Research in Science Teaching* **2013**, *50* (5), 561−596.

(16) 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.

(17) 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.

(18) Trochim, W. M.*The Research Methods Knowledge Base*, 2nd ed.; Atomic Dog: Cincinnati, OH, 1999.

(19) 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), e0162333.

(20) Holme, T.; Luxford, C.; Murphy, K. Updating the General Chemistry Anchoring Concepts Content Map. *J. Chem. Educ.* **2015**, *92* (6), 1115−1116.

(21) Cohen, J. Weighted Kappa: Nominal Scale Agreement Provision for Scaled Disagreement or Partial Credit. *Psychological Bulletin* **1968**, *70* (4), 213−220.

(22) Hallgren, K. A. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in quantitative methods for psychology* **2012**, *8* (1), 23−34.

(23) Smith-Joyner, A.*Graduate Teaching Assistant Fidelity of Implementation in Introductory Chemistry and Physics Laboratories: Impact on Science Practice Proficiency*. Thesis, East Carolina University, 2020.

(24) Stowe, R. L.; Scharlott, L. J.; Ralph, V. R.; Becker, N. M.; Cooper, M. M. You Are What You Assess: The Case for Emphasizing Chemistry on Chemistry Assessments. *J. Chem. Educ.* **2021**, *98* (8), 2490−2495.

(25) Ralph, V. R.; Lewis, S. E. An Explanative Basis for the Differential Performance of Students with Low Math Aptitude in General Chemistry. *Chem. Educ. Res. Pract.* **2019**, *20* (3), 570−593.

(26) Stowe, R. L.; Scharlott, L. J.; Ralph, V. R.; Becker, N. M.; Cooper, M. M. You Are What You Assess: The Case for Emphasizing Chemistry on Chemistry Assessments. *J. Chem. Educ.* **2021**, *98* (8), 2490−2495.

(27) Stowe, R. L.; Esselman, B. J.; Ralph, V. R.; Ellison, A. J.; Martell, J. D.; DeGlopper, K. S.; Schwarz, C. E. Impact of Maintaining Assessment Emphasis on Three-Dimensional Learning as Organic Chemistry Moved Online. *J. Chem. Educ.* **2020**, *97* (9), 2408−2420.

(28) Ralph, V. R.; States, N. E.; Corrales, A.; Nguyen, Y.; Atkinson, M. B. Relevance and Equity: Should Stoichiometry Be the Foundation of Introductory Chemistry Courses? *Chem. Educ. Res. Pract.* **2022**, *23*, 662.

(29) Hosbein, K. N.; Lower, M. A.; Walker, J. P. Tracking Student Argumentation Skills across General Chemistry through Argument-Driven Inquiry Using the Assessment of Scientific Argumentation in the Classroom Observation Protocol. *J. Chem. Educ.* **2021**, *98* (6), 1875−1887.

(30) Sandoval, W. A.; Millwood, K. A. The Quality of Students' Use of Evidence in Written Scientific Explanations. *Cognition and Instruction* **2005**, *23* (1), 23−55.

(31) Moon, A.; Stanford, C.; Cole, R.; Towns, M. The Nature of Students' Chemical Reasoning Employed in Scientific Argumentation in Physical Chemistry. *Chem. Educ. Res. Pract.* **2016**, *17* (2), 353−364.

(32) 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. Technol.* **2018**, *55*, 1264.

(33) 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.

(34) Allen, W. E.; Hosbein, K. N.; Kennedy, A. M.; Whiting, B.; Walker, J. P. Embedding Research Directly into the Chemistry Curriculum with an Organic to Analytical Sequence. *J. Chem. Educ.* **2021**, *98* (7), 2188−2198.

https://doi.org/10.1021/acs.jchemed.2c00578
J. Chem. Educ. 2022, 99, 3833−3841