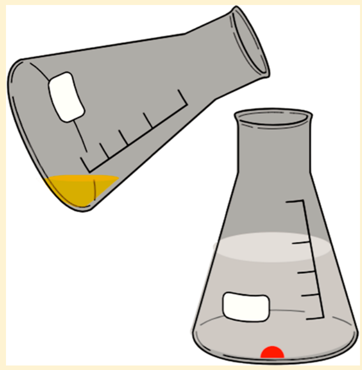
JOURNAL OF CHEMICAL EDUCATION
Article
Cite This: J. Chem. Educ. XXXX, XXX, XXX−XXX
pubs.acs.org/jchemeduc

Whether and How Authentic Contexts Using a Virtual Chemistry Lab Support Learning

Jodi L. Davenport,*,† Anna N. Rafferty,‡ and David J. Yaron§

†WestEd, Alameda, California 94501, United States
‡Carleton College, Northfield, Minnesota 55057, United States
§Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States

ABSTRACT: How can we help students develop an understanding of chemistry that integrates conceptual knowledge with the experimental and computational procedures needed to apply chemistry in authentic contexts? The current work describes ChemVLab+, a set of online chemistry activities that were developed using promising design principles from chemistry education and learning science research: setting instruction in authentic contexts, connecting concepts with science practices, linking multiple representations, and using formative assessment with feedback. A study with more than 1400 high school students found that students using the online activities demonstrated increased learning as evidenced by improved problem solving and inquiry over the course of the activities and by statistically significant improvements from pre- to posttest. Further, exploratory analyses suggest that students may learn most effectively from these materials when the activities are used after initial exposure to the content and when they work individually rather than in pairs.



KEYWORDS: *High School/Introductory Chemistry, First-Year Undergraduate/General, Chemical Education Research, Internet/Web-Based Learning, Multimedia-Based Learning, Stoichiometry, Equilibrium, Thermodynamics*

FEATURE: Chemical Education Research

In response to concerns that typical chemistry instruction focuses on isolated facts and procedures, scientists and educators continue to advocate for new approaches to science instruction. One example is the Next Generation Science Standards, that outlines a vision for science education where students learn disciplinary core ideas by engaging in authentic science practices while making connections to cross-cutting concepts, such as the flow of energy and matter, that span physical, life, and earth sciences.[1] With increased access to computers in the classrooms, interactive and simulation-based activities enable students to carry out investigations when traditional laboratory experiences are not possible. This paper reports on the design and testing of ChemVLab+, a series of online activities that enable students to learn core concepts while carrying out investigations in real-world contexts.

Design principles from research on science learning informed the design of eight ChemVLab+ activities. The activities set chemistry learning in authentic, real-world contexts, couple the chemistry content with science practices, e.g., designing experiments, analyzing data, and interpreting results, promote integration of the multiple representations of chemistry, and provide formative assessment via immediate feedback and teacher reports. Students receive just-in-time feedback based on their responses, and teachers can review reports that show student proficiencies across core concepts and inquiry skills. Our study addressed two research questions: (1) How do activities applying these design principles help students learn? and (2) In what ways does the context of classroom use influence how students learn from the activities?

■ CHEMVLAB+ ACTIVITIES

ChemVLab+ activities scaffold students as they carry out virtual lab investigations related to an authentic context. We provide a walkthrough of the Drinking Water activity as an example of the ChemVLab+ approach. The central problem in the Drinking Water activity is whether water from the school drinking fountain is safe to drink. Throughout the activity students are prompted to consider the rationale behind each step. Students are introduced to the difference between soluble and insoluble salts and use the virtual lab to mix different solutions of salts to determine which reactions form precipitates (see Figure 1). Next, students learn that the Environmental Protection Agency provides a recommended range of concentration for sulfates. They are then introduced to gravimetric analysis and learn that sulfates can react with barium chloride to form insoluble salts that can be filtered and weighed to determine the initial sulfate concentration in a sample of water. Students use the virtual lab to carry out the process of gravimetric analysis with the water sample. Once the precipitate has formed, students determine the molar mass of

Received: January 22, 2018
Revised: June 18, 2018

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX

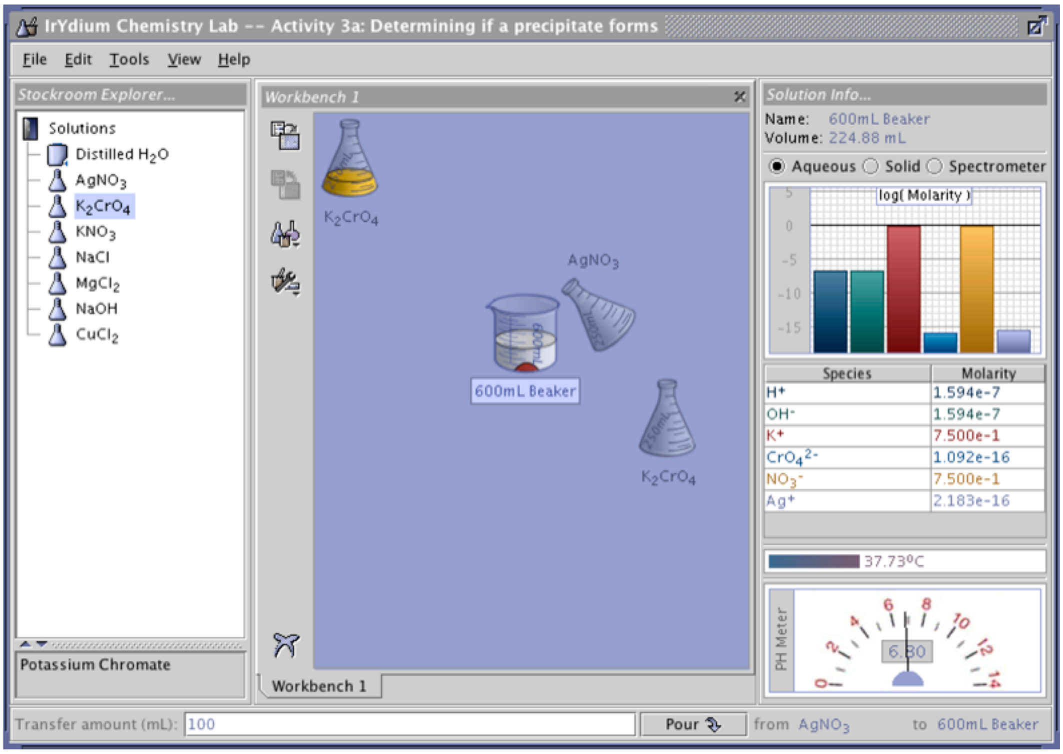


Figure 1. Screen capture of the Drinking Water activity showing a chemical reaction that produces a precipitate.

barium sulfate using information from the periodic table. Finally, students use the mole ratio and unit conversion to calculate the concentration of sulfate and determine whether the water meets the recommendations. Students repeat the analyses for a different water sample and again answer prompts to explain their actions.

LITERATURE BASIS AND DESIGN PRINCIPLES

ChemVLab+ activities were designed to address challenges that make chemistry difficult to learn and teach. Chemistry involves entities and processes (i.e., molecules and their rearrangement during reactions) that cannot be directly observed and whose size and number are at a scale that is vastly beyond students' everyday experience. To succinctly convey ideas and explanations at multiple scales, the field uses a variety of abstract representations, notational systems, and quantitative procedures that students must learn as they simultaneously try to grasp key chemical principles. The Johnstone triangle captures three representations of chemical phenomena that must be coordinated to understand the disciplinary core ideas of chemistry: symbolic (e.g., notations of chemistry), submicroscopic (e.g., interactions of particles and forces), and macroscopic (e.g., substances or solutions in a lab).[2] Though experts move fluidly between these various representations when reasoning about chemistry, novices struggle to make connections and require thoughtfully designed instruction to develop deep understandings of the domain.[3,4]

Typical high school chemistry instruction and assessment emphasizes quantitative problem-solving activities and practice with symbolic manipulations, such as balancing chemical equations and drawing Lewis structures. These practices assume that students will learn core concepts in chemistry by manipulating numbers and symbols. However, all too often, students' assessments of procedural knowledge suggest mastery, but assessments of the associated concepts suggest many students learn procedures without understanding core principles.[5−9]

Our goal was to promote deeper conceptual understanding by prompting students to connect quantitative calculations to chemical processes at the microscopic level (e.g., the level of atoms and molecules) and to outcomes at the macroscopic level (e.g., final concentrations, color, temperature). We applied four design principles: using authentic contexts, integrating science practices, building on multiple representations, and providing formative assessment with feedback.

Setting Instruction in Authentic Chemistry Contexts

Chemistry education research demonstrates that authentic and context-based instruction helps students make connections to students' lives and promotes learning in chemistry classrooms. When students engage in solving meaningful problems, in authentic contexts such as climate change, designing medical drugs, or environmental pollutants, they are more engaged, and are more likely to use higher-order thinking skills.[10−12] Learning science research suggests that the learning benefit results from contextualized knowledge being more readily accessible, thus more memorable and more likely to transfer to new situations.[13]

As chemistry instruction often presents facts and procedures in isolation, many high school students fail to learn what chemists actually do. A study of Nobel prizes and science publications sought to categorize the activities of chemists and found the main activities included *explaining* phenomena, *analyzing* substances to reveal their chemical composition and *synthesizing* new materials. The "toolbox" of chemistry, i.e., the notations, calculations, and procedures, supports these activities. In contrast to chemistry as practiced, analyses revealed that popular chemistry textbooks focused on teaching

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX

the "toolbox" and explaining phenomena with little coverage of analysis or synthesis activities.[14]

ChemVLab+ was developed to help students connect chemistry with their lives. Eight context-based activities make up two modules: stoichiometry and equilibrium/thermodynamics. Each 45 min activity begins by posing an authentic problem to be addressed, moves through use of the tools of chemistry to solve the problem, and closes with interpretation of the findings. Table 1 shows the contexts, type of chemistry activity, and topics. Examples of guiding questions are as follows: What is the concentration of sugar in a drink? Are the factories accurately reporting their emissions? Is water safe to drink? What is the chemical formula for a bioremediation accelerator? What substances will be best to use for a hotpack or to store energy in a solar power plant?

Table 1. Contexts and Topics for Each Activity

Module	Context	Chemistry Activity and Topic
Stoichiometry		
1	Concentration of a sports drink	Analysis of concentration and color intensity
2	Evaluating factory emissions	Analysis of effects of dilution
3	Drinking water	Analysis using gravimetric analysis
4	Bioremediation of oil spills	Explanation using reaction stoichiometry
Equilibrium/Thermo		
5	Manipulating equilibrium systems	Explanation using Le Chatelier's principle
6	Making hot and cold packs	Analysis using reaction enthalpy
7	Solar energy	Synthesis using energy transfer, heat capacity
8	pH and pool safety	Analysis using acid–base chemistry

Connecting Concepts with Science Practices

The second design principle is to provide students with an interactive environment that allows them to develop and use science practice skills. The Next Generation Science Standards[1] emphasize learning core ideas and concepts by engaging in science practices such as asking questions, designing investigations, and drawing conclusions from evidence. These practices require students to have access to either physical or virtual environments where they can manipulate chemical systems, gather data, and analyze results.

Science laboratory setups allow students to gain first-hand experience learning the tools and techniques of the field. However, pragmatic constraints limit access to laboratories for many students. Many schools lack the resources to stock and maintain laboratories and restrict the types of chemicals and tools that can be used because of safety or environmental concerns. When lab-based activities are available, many still fail to engage students in true inquiry, as students follow step-by-step instructions instead of designing novel approaches or reasoning for themselves.

Virtual simulation environments provide students with opportunities to actively engage in practices when physical laboratories are not available or practical. These environments use simulations to visualize invisible processes and enable a wider range of investigations that are not limited by the constraints of physical lab setups. For instance, stand-alone simulations allow students to explore and manipulate submicroscopic processes, e.g., Phet,[15] Connected Chemistry,[16,17] the Molecular Workbench,[18,19] and the Minds and Molecules[20] project. Virtual chemistry laboratories mimic real laboratory setups and allow students to carry out investigations from anywhere they have access to a computer.[21−23] Research suggests these environments help students create models of unobservable phenomena in the context of lab investigations.[23−25] Like physical laboratories, the efficacy of simulation environments depends on how teachers use the materials, what supports are provided for students, and how students interact with the materials. Virtual laboratories require teachers or supplemental materials that link concepts together, coach students, and promote the forms of scientific experimentation and inquiry that reflects real-world chemistry research.[26] Many existing simulation environments require substantial planning from teachers to integrate into existing curricula and pose challenges for classroom management as they lack support for tracking progress or helping individual students with different abilities.[25,27,28]

ChemVLab+ activities address the difficulties of providing an interactive environment for students to engage with science practices by embedding the virtual lab experiments in self-contained instructional modules that provide support to both the students and the instructors. Our motivation is not to replace classroom lab experiences, but rather to provide additional opportunities for students to connect their knowledge to laboratory investigations. ChemVLab+ activities embed a virtual lab that allows students to select chemical reagents, manipulate them in a manner that resembles that of a physical laboratory, and examine various representations of the outcome of their experiments. The open-ended nature of the lab enables students to design and analyze results from their own experiments. As the virtual lab is embedded in a larger activity with a guiding question, students have context for carrying out investigations. The system mitigates issues of classroom management by providing students and teachers with just-in-time feedback as described below.

Connecting Multiple Representations in Johnstone's Triangle

The third design principle suggests students should be provided with opportunities to connect the multiple representations in Johnstone's triangle. Learning science research has repeatedly demonstrated that providing instruction with multiple visual representations can enhance learning, particularly when students are encouraged to actively link the representations.[27−30] When students are prompted to make predictions, observe, and create explanations based on dynamic displays, such as animations that show the motion of particles, they develop deeper understanding of complex processes.[31,32] Educational technologies in chemistry and physics often present a variety of representations, such as simulations, animations, graphs, and pictures, simultaneously on the same screen.[26] Learning to integrate across these representations is central for a deep understanding of chemistry.[33]

The ChemVLab+ activities include representations from all three corners of Johnstone's triangle. Students interact with macroscopic representations as they manipulate solutions in the virtual chemistry lab. Students engage with molecular representations as they sort collections of particles at different temperatures or concentrations. Finally, students view multiple symbolic representations including chemical reaction equations, and chemical quantities expressed in moles, grams, and concentrations. Prompts in the activities focus attention on key aspects of the representations, scaffold understanding of these

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX

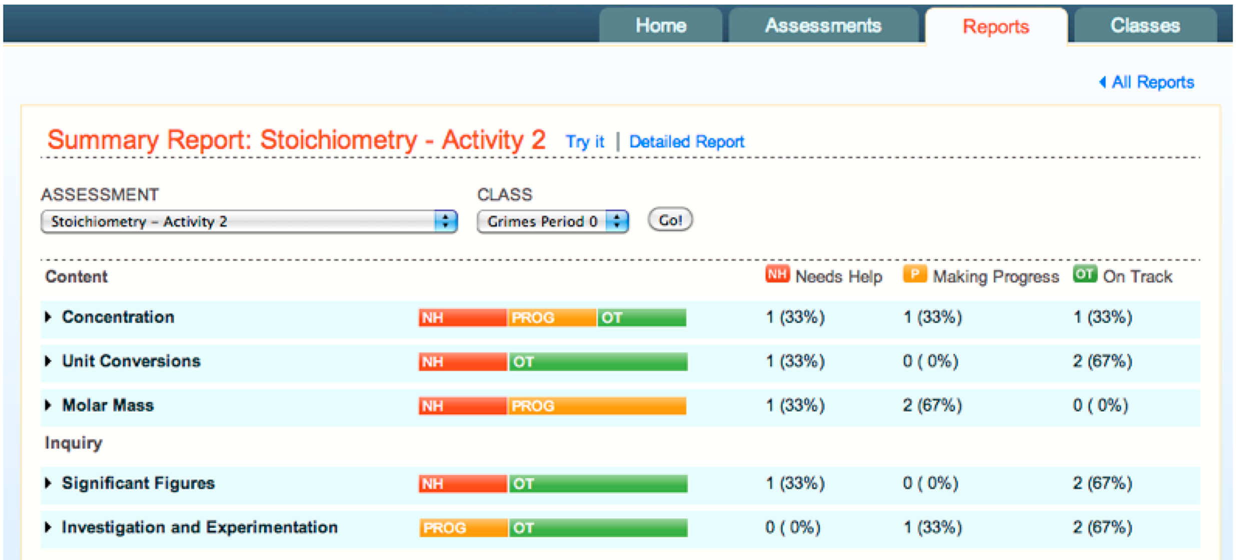


Figure 2. Example of summary report for teachers.

representations, and encourage students to make connections across representations. As the ChemVLab+ activities tie chemical representations at both the submicroscopic and macroscopic scale to an authentic context, students develop fluency connecting what is observable with what is happening at the particulate level.

Formative Assessment with Feedback

The final design principle relates to providing formative assessment with feedback. The ChemVLab+ activities apply research from cognitive and educational measurement research about the power of quizzing and formative assessment for learning. Quizzing encourages students to practice retrieving information from memory, which leads to improved retention of key information.[34,35] Quizzing can additionally enhance learning when used as formative assessments that provide timely feedback and additional instruction that is tailored to a student's current level of understanding.[36]

ChemVLab+ activities provide sequenced tasks for students to complete. Each task serves as an embedded assessment that evaluates student inputs and provides feedback about the correctness of their responses. When students carry out experiments, the system analyzes the state of the virtual lab, determines the current stage of the experiment, and provides an appropriate hint. Students can receive feedback on demand, by requesting a hint, or as needed, when attempting to progress to the next screen with errors. Hint messages provide increasing levels of support. The first hint tells students the location of the error, the second hint explains the concept behind the error, and the final hint provides student with the correct response. The specific feedback in the hint messages were derived from the literature on common chemistry errors and misconceptions.

As the students can work independently at their own pace, teachers are able to work with individual students, a practice that has been shown to be effective in research using online systems that provide customized feedback.[37]

In addition to receiving just-in-time feedback through the hint messages, students and teachers receive summative feedback at the end of each activity. Student proficiency is estimated using the number of attempts they needed before they completed tasks successfully, with the fewest attempts demonstrating the highest level of mastery. When the class completes an activity, teachers can plan future instruction using reports that summarize student performance across key concepts and skills. See an example summary report in Figure 2.

■ RESEARCH QUESTIONS

We predicted that completing ChemVLab+ activities designed with learning principles would increase students' understanding of key chemistry concepts. Our two research questions were as follows: (1) What evidence is there that these types of activities help students learn? and (2) How does the context of use affect student learning?

In the current study, we explored whether ChemVLab+ activities improve student learning in California high schools with diverse student populations. We hypothesized that the combination of authentic problem-solving contexts, emphasis on science practice skills, focus on connecting multiple representations, and formative assessment with immediate feedback had the potential to improve student learning of chemistry concepts for a wide range of students.

As prior research suggests that differential effects may be found on the basis of how online activities are used in classroom settings, we also carried out exploratory analyses to investigate whether the timing of using the activities (e.g., before introducing a topic, during an instructional unit, or at the end of a unit) or the mode of assignment (e.g., as homework, to individuals in class, or in pairs in class) had effects on student learning.

■ METHODS

Participants

Fourteen teachers and 1473 students from 12 San Francisco Bay Area high schools participated in the study. An additional 19 students declined to participate. IRB approval was obtained, and students were given the option to opt out of the study. The schools represented a diverse range of settings, including urban, suburban, and rural, with free and reduced lunch status ranging from 1−66%. All teachers used the equilibrium/thermodynamics module with their students; however, due to

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX

scheduling constraints, one teacher did not use the stoichiometry module, resulting in 1334 students participating in the four stoichiometry activities.

Design and Procedures

To measure student learning across activities we compared student performance at posttest with performance on the same pretest. To measure student learning within activities, we used computer log file data to compare the number of attempts students needed to complete a task the first and second time. Finally, to measure the effects of context on student learning, we integrated data from teacher interviews and logs with the pre- and posttest scores to identify how the activities were used and carried out exploratory analyses. More details about the data sources are provided below.

Before students used the activities, participating teachers attended a 3 h workshop to learn about the activities and options for integrating them with their teaching. Teachers were able to select when in their instructional sequence they introduced the activities and how they assigned them to their classes (e.g., individually as homework, individually in class, or in pairs in class).

As detailed in Table 1, the ChemVLab+ activities were split into two modules, stoichiometry and equilibrium/thermodynamics. An assessment was created for each module and the same assessment was given at pretest and posttest. For each module, teachers administered the pretest to their students. Next, students completed the four activities in the module. Each activity was designed to take approximately 45 min to fit in a single class period. Approximately half of the teachers chose to interleave the activities with periods of other classroom instruction, the other half of the teachers chose to have students complete the activities consecutively with no additional instruction. After students completed the four activities in the module, students individually took the same assessment as a posttest. With the exception of the teacher that did not use the stoichiometry module, all teachers used the stoichiometry module before the equilibrium/thermodynamics module.

Data Sources

Data sources for the current analyses included assessments used as pretests and posttests, computer log files, and teacher logs and interviews.

Two assessments were created to measure student learning, one covering topics related to stoichiometry and one covering topics related to thermodynamics and equilibrium. The activities in each module are detailed in Table 1. To avoid floor and ceiling effects, items were selected to reflect a range of difficulty with a target of approximately 50% correct. Items were sourced from released standardized tests including the California Standards Test in Chemistry, the SAT II Chemistry Subject exam, and the New York Regents Examination, or were researcher-generated.

The stoichiometry prepost test consisted of 15 items. As some items had multiple subparts, the assessment was scored for a total of 26 points. Subparts of the items were aligned to five learning targets: concentration and dilution (6), unit conversion (4), using molar mass (5), balancing reactions (4), and using stoichiometry (7). The equilibrium/thermodynamics prepost test consisted of 25 items, and was scored for a total of 34 points. Subparts of the items were aligned to four learning targets: heat and temperature (8), experimentation and problem solving (9), equilibrium (7), and acid−base chemistry (10). Some complex items were included in multiple categories.

As the assessments were researcher-created, we evaluated the two tests for validity and reliability. To ensure validity, the alignment of items with learning objectives was reviewed by a chemist, cognitive scientist, and an assessment development expert. To ensure reliability, we field tested the assessments in high school classrooms the year before using them for our study. None of the students in the field test participated during the study year. For the field test, posttest data was gathered from 337 students on the stoichiometry assessment and 220 students on the thermodynamics assessment. Overall, IRT analyses found the tests to have good reliability. For the stoichiometry assessment, the EAP reliability was 0.80 and Cronbach's Alpha was 0.76, and for equilibrium and thermodynamics assessment, both EAP reliability and Cronbach's alpha were 0.84.

Another source of data was computer log files from the activities that indicated the overall number of attempts students needed to complete a task and whether students were correct on the first try. Activities were designed to have paired tasks with similar demands. Having a second opportunity to demonstrate knowledge and skills provided students with additional practice and allowed us to track learning within an activity. Our hypothesis was that if students were learning as they progressed through an activity, the second time they encountered a similar task the overall number of attempts it took students to successfully complete the task would be reduced and more students would complete it correctly on the first try. For example, paired tasks in the sports drink activity required students to create drinks with different specified concentrations, paired tasks in the water safety activity required students to carry out gravimetric analysis on different samples of water, and paired tasks in the acid base activity required students to test and adjust water pH for different pools. The paired tasks had slight differences; the first task in the pair provided more scaffolding for students than the second task in the pair, such as suggestions for how to perform the lab activity. In all cases, these small differences required students to work more independently in the second task, meaning that the second task was at least as difficult as the first task.

To determine when and how teachers used the activities, teachers completed online instructional logs after completing each module with their students and participated in structured phone interviews. In the instructional logs, teachers reported which virtual lab activities they used, whether they had technical difficulties, and how they integrated the activities into their teaching. In the structured phone interviews, teachers provided details of how they used the activities, how they used reports for formative assessment, and how students reacted to the activities. For the current study, information from the logs and phone interviews was used to determine the timing of use of the two modules (e.g., before introducing a topic, interleaved with instruction, or as review after completing a topic), and the mode of administering the activities (e.g., individually as homework, individually in the classroom, or in pairs in the classroom).

■ FINDINGS

What evidence is there that these types of activities help students learn? To analyze student understanding and learning over the course of each module, we examined performance at

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX

two levels of granularity: pre- to posttest performance after completion of the activities and performance within the activities. For all t tests we report both p value, that indicates whether the effect is significant, and the effect size, that describes the magnitude of the effect. We calculated effect sizes using Cohen's d which divides the difference in means between conditions by the pooled standard deviation. As significance levels, i.e., p values, increase with sample size, calculating the effect size is essential for understanding whether an effect has practical significance for education. Cohen's d describes the magnitude of a difference using standard deviation units. A small effect size of 0.2 represents one-fifth of a standard deviation, a medium effect size of 0.5 represents one-half of a standard deviation, and a large effect size of 0.8 or greater represents 8/10ths of a standard deviation.[38] Another way of understanding the expected magnitude of an effect is to compare to typical effect sizes in similar circumstances using similar measure. Relevant to our current work, the typical effect size of student learning in their first year of high school as measured by nationally normed science tests was 0.19, and mean effect sizes from studies using researcher developed measures in high school was 0.39.[39] As our assessments reflected a mix of items sourced from large scale standardized tests and created by researchers, effect sizes greater than 0.3 likely reflect substantial practical significance.

Evidence of Learning from Pretest to Posttest

Of the 1334 students who participated in at least some of the stoichiometry activities, 1185 (89%) completed both pre- and posttests. Only data from students completing both assessments were used for our analyses. A paired t test comparing scores on the stoichiometry assessment at pre- (M = 10.67) and posttest (M = 13.22) found that student scores after completing the stoichiometry activities improved, on average, by 24% of the pretest score, $t(1184) = 23.4$, $p < 0.00$, $d = 0.48$.

To ensure the effects were general, rather than reflecting an improvement on just a few items, we used t tests to compare pre- and posttest scores on items related to each learning target. Scores improved the most on items aligned with learning targets in the sports drink and factory activities related to concentration and dilution, $t(1184) = 19.4$, $p < 0.001$, d = 0.52, and unit conversion, $t(1184) = 18.6$, $p < 0.001$, $d = 0.56$. Items related to the other three learning targets also showed statistically significant improvements from pre- to posttest, but more modest effect sizes: molar mass, $t(1184) = 8.65$, $p <$ 0.001, $d = 0.25$, balancing reactions, $t(1184) = 13.5$, $p < 0.001$, $d = 0.33$, and using stoichiometry, $t(1184) = 7.25$, $p < 0.001$, d = 0.21.

Of the 1473 students who participated in at least some of the equilibrium and thermodynamics activities, 1195 (81%) completed both pre- and posttests. As with the stoichiometry module, only data from students completing both assessments were used for our analyses. A paired t test comparing scores on the equilibrium/thermodynamics assessment at pre- (M = 13.85) and posttest (M = 16.11) found that student scores after completing the module improved, on average, by 16% of the pretest score, $t(1194)$=18.0, $p < 0.001$; Cohen's $d = 0.38$. Similar to stoichiometry, we carried out paired t tests comparing pre- and posttest scores on each of the learning targets to ensure that students improved overall, and not for a subset of items. Students made the largest improvement on items related to acid base chemistry, $t(1194) = 15.3$, $p < 0.001$, d = 0.40, followed by experimentation and problem solving, $t(1194) = 15.0$, $p < 0.001$, $d = 0.34$, heat and temperature $t(1194) = 13.0$, $p < 0.001$, $d = 0.29$, and equilibrium, $t(1194)$ = 8.12, $p < 0.001$, $d = 0.23$.

Overall effect sizes show the improvements from pre- to posttest were of practical significance for both stoichiometry (d = 0.48) and equilibrium/thermodynamics ($d = 0.38$). That is, student scores improved nearly half a standard deviation between pre- and posttest.

As a final indicator of student learning, we evaluated whether the number of activities completed by students correlated with higher posttest scores. Students that completed more activities tended to have higher posttest scores for both stoichiometry, $r(1183) = 0.08$, $p < 0.01$, and for equilibrium/thermodynamics, $r(1193) = 0.098$, $p < 0.001$.

Evidence of Learning within Activities

To provide evidence that students were learning over the course of the activities, we used computer logs to examine changes in performance across pairs of similar tasks. Specifically, we looked at whether students were more likely to be successful on the first try or require fewer attempts the second time they performed a task. All activities required students to correctly complete each task before they could move on to the rest of the activity.

The four stoichiometry activities contained 26 paired tasks. We examined the 30,503 cases in the log files that corresponded to a student completing both tasks in a pair. Because each student could potentially complete all 26 pairs, the same student is represented multiple times across these cases. A paired t-test comparing performance on the second task in a pair with performance on the first task, found that students were more likely to be correct on their first try for the second task in the pair (53.0%), than for the first task in the pair (41.3%), $t(30{,}502) = -34.84$, $p < 0.0001$. Even when students were not correct on the first try, the average number of attempts decreased from a mean of 3.82 for the first task to a mean of 2.77 for the second task, ($t(30{,}502) = 38.96$, $p <$ 0.0001), with a median of two attempts for the first task and one attempt for the second. These findings suggest that students learned from the first task and applied this understanding the next time they were presented with a similar task.

The equilibrium/thermodynamics activities contained 22 paired tasks. We examined the 24,209 cases in the log files where a student completed both tasks in a pair. A paired t test comparing performance on the second task in a pair with performance on the first task found students were more likely to be correct on the first try on the second task (54.6%) than for the first task in a pair (46.2%), $t(24{,}208) = -21.87$, $p <$ 0.0001). The average number of attempts also decreased from 3.42 on the first task to 2.72 on the second task ($t(24{,}208)$ = 25.31, $p < 0.0001$). These results mirror the findings from the stoichiometry activities.

Overall, these analyses provide evidence that students are learning within each activity.

Contextual Mediators of Learning

How does the context of use affect student learning? To address our second research question, we analyzed whether the timing or mode of use affected student posttest performance in the stoichiometry module. We used ANCOVAs for these analyses with pretest as a covariate to account for differences in pretest scores due to prior opportunities to learn the material.

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX

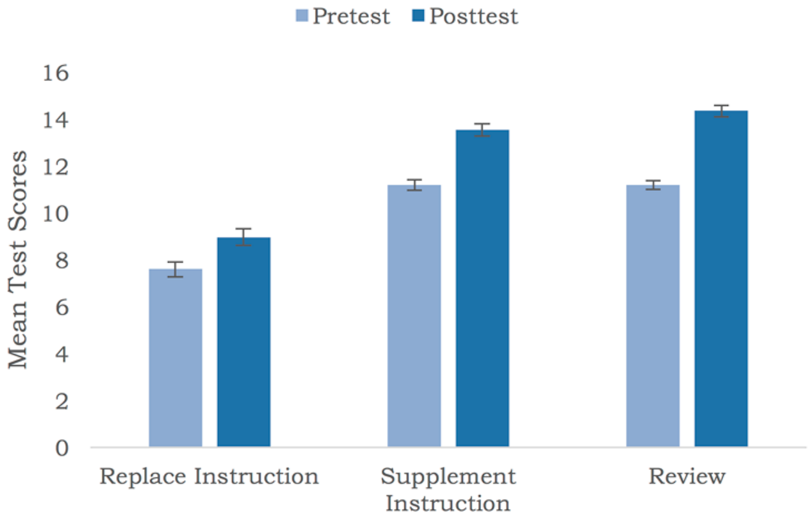


Figure 3. Pre- and posttest scores by the timing of activities. All types had significant improvements from pre- to posttest, and improvements were largest for teachers that used the activities as review.

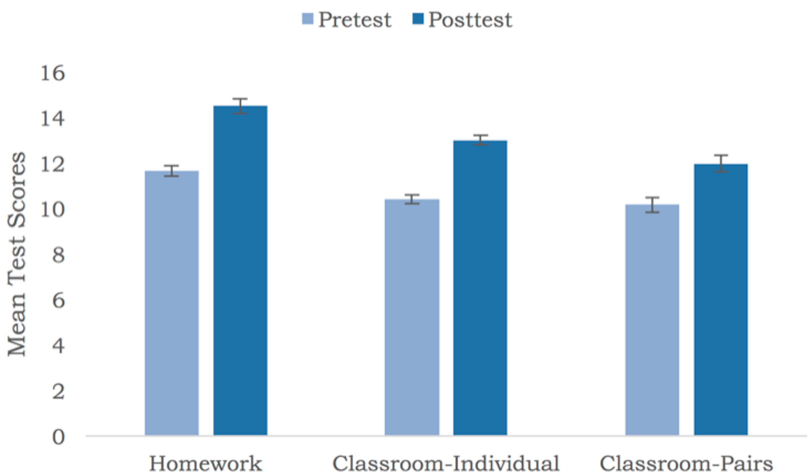


Figure 4. Pre- and posttest scores by how the activities were completed. All use types had significant improvements from pre- to posttest, and improvements were largest for students using the activities individually either at home or in the classroom.

Timing of Use

Teachers had flexibility in how they integrated the ChemVLab + activities with other types of instruction. Using data from teacher logs and interviews, we categorized teacher activity use into three categories: *replace instruction*, *supplement instruction*, or used as *review*. Teachers using the activities to *replace instruction*, presented the ChemVLab+ activities as the only exposure to the material and gave no additional instruction between pre- and posttest. Teachers using the activities to *supplement instruction* interleaved the activities with their own lectures and stoichiometry activities. Finally, teachers using the activities for *review*, presented the activities to the students after the topics had already been covered in class to "strengthen previously taught concepts", but similar to teachers in the replace instruction condition, gave no additional instruction between pre- and posttest. Of the 13 teachers, two teachers used the activities as a replacement for instruction, six used the activities to supplement instruction, and five used the activities as review of instruction.

As expected, pretest scores for students that had little instruction on these topics prior to using the activities were lower than scores in the other conditions. Thus, for activity use, we used an ANCOVA with pretest score as a covariate and found significant differences between posttest scores in each category, F (2, 1181) = 22.54, $p < 0.001$. The increase from pre- to posttest was 1.37 points (d = 0.31) when used to *replace instruction*, 2.35 points (d = 0.43) when used to *supplement instruction*, and 3.16 points (d = 0.64) when used as *review*. See Figure 3. Students demonstrated larger improvements from pre- to posttest when given the activities as *review* despite the fact that students in the *supplement* instruction received additional instruction between pre- and posttest, and that students in the *replace* condition had the most opportunity for growth. Further, students seemed to benefit most from the activities alone if they had already had prior instruction on the content.

How Activities Were Used

In addition to studying the timing of use, we also explored the effects of teachers' choices in assigning students to work on the activity at school or as homework. At school, some teachers had students work in pairs and others had students work independently. At home, students who completed the activities as homework were presumed to have worked independently. The majority of teachers chose to have students complete the

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX

activities individually during class (nine teachers), two chose to assign the activities to be completed in pairs in class, and two teachers chose to assign the activities as homework. We analyzed results for *assignment type* similarly to activity use.

For assignment type, an ANCOVA of posttest scores with pretest score as a covariate found significant differences between each category, F (2, 1181) = 5.63, $p < 0.01$. The mean improvement from pre- to posttest for *classroom-pairs* was 1.37 points ($d = 0.31$), the mean improvement from pre- to posttest for *classroom-individual* 2.35 points ($d = 0.43$) and the mean improvement from pre- to posttest for *homework* was 3.16 points ($d = 0.64$). See Figure 4. Posthoc comparison showed that students in the *classroom-pairs* demonstrated less improvement than students using the materials individually. The results suggest students benefited more when they worked through the activities independently.

■ LIMITATIONS

Though the current study suggests that activities that use authentic contexts, engage students in science practices, prompt mapping between representations of the Johnstone's triangle, and provide just-in-time feedback to help students learn chemistry, the design of the study limits the nature of the conclusions that we can draw.

First, the activities were designed according to four principles that have been shown to improve student learning in past work. As our aim was to study the synergistic effects of applying these principles, our design did not allow us to make claims about the relative contributions of different features on improvement from pre- to posttest.

Second, the teachers varied in how they implemented the activities in their classes. In classrooms that used the activities to replace instruction or as review of materials previously taught, no additional instruction was provided to students and any improvements from pre- to posttest may be attributed to the use of the ChemVLab+ activities. In contrast, in classrooms where the activities were used to supplement instruction, teachers did provide students with additional instruction that may have contributed to increased performance at posttest. As scores from students receiving no additional instruction (in the replace or review conditions) increased between 0.3 and 0.6 standard deviations from pre- to posttest, the findings suggest the activities improved student learning.

The current study was conducted in classroom settings that only allowed us to collect data from pre-posttests, system logs, and qualitative data from teachers. We can infer how students were learning from the activities, but future work could supplement the use of activities with student interviews that probe more deeply into whether and how student conceptions develop as they use the eight activities.

Finally, our exploratory study revealed correlations that suggest ways the context of use may influence learning. Because we were unable to randomly assign teachers to use the activities in a particular way, we cannot draw causal conclusions as other variables may have contributed to the effects. Future research is needed to better understand optimal learning conditions for these types of activities.

■ IMPLICATIONS FOR TEACHING AND RESEARCH

Our work has a number of implications for teaching and research. First, we provide a model for creating online activities that applies design principles from the learning sciences. The structure of the activities showcases an approach to using simulations and virtual laboratories that integrate authentic science contexts, engage students in science practices, promote connections across multiple representations, and offer embedded assessments with immediate feedback. The ChemVLab+ activities differ from many existing simulation environments that focus solely on molecular visualizations or virtual laboratories without making connections between the two, fail to provide help for students working at different levels, and require effort from teachers to construct learning sequences and real-world applications for the tasks.

A second contribution of the work is the novel method of using paired tasks to investigate learning within activities. Providing opportunities for "paired tasks" across an online activity has multiple advantages. Students can solidify their learning by engaging in practice in similar tasks with increasing difficulty, instructional developers can have early indications of whether the activities they create are effective, and researchers may develop new insights related to learning progressions as they investigate how patterns of student errors change across an activity.

Finally, our exploratory analyses offer early indications how the timing and context of activities may impact learning. The context of use is important to consider for learning technology as the utility may vary depending on when and how students engage with the materials. Our exploratory data analyses found that the way teachers used the activities in the classroom was differentially associated with student learning. The analyses revealed significant differences between the three ways teachers used the activities: to replace instruction, to supplement instruction, or as review. Students that received the activities as review made the largest improvements from pre- to posttest. Using the activities as a means to reinforce and integrate previously learned concepts may be more effective than using the activities as a replacement for classroom instruction. Activities requiring a range of science content knowledge and practice skills to be applied to real contexts may have the most impact after students have had the opportunity to be exposed to some of the content earlier. Additional research is needed to understand what types of instructional sequences best support student learning.

Past work suggests that collaborative learning is generally more effective than learning individually and that the benefits of collaboration are moderated by a number of factors including learning objectives, structure of collaboration, culture, and the structure of the pairs.[40−42] In contrast, we found that students using the activities independently, either individually in class or as homework, appeared to learn more than students working in pairs. Though our design does not allow us to establish causality, we offer several hypotheses for this seemingly discrepant finding. First, the benefits of pair-based learning may differ for online, interactive activities. As the comparison condition for the majority of studies showing large benefits for pair-based learning was lecture-based instruction, the strong effects of pairs may not hold when control activities also require ongoing engagement. Second, collaborative learning may be less effective in systems that provide customized feedback during instruction. ChemVLab+ activities provide hints that are specific to the actions students took in the system. As different students may have different instructional needs, the hints may not have been optimally effective for both students in a pair. Finally, prior work suggests that effective pair-based learning requires students to actively

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX

participate in instructional activities.[40,41] As the ChemVLab+ system was not designed to support pair work, students within pairs may have differed in their use of the system, with one student taking the lead on moving through the activities. Future work is needed to better understand how different instructional tools can be used most effectively. Though discussion is clearly an essential part of chemistry classrooms, some activities may be most effective when students have the time to focus individually.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jdavenp@wested.org.

ORCID

Jodi L. Davenport: 0000-0001-9091-9616
David J. Yaron: 0000-0001-8485-8685

Notes

The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305A100069 and R305A170049 to WestEd.

■ REFERENCES

(1) National Research Council. *Next Generation Science Standards: For States, By States*; The National Academies Press: Washington, D.C., 2013; DOI: 10.17226/18290.

(2) Johnstone, A. H. Why is science difficult to learn? Things are seldom what they seem. *J. Compu. Assist. Learn* **1991**, *7* (2), 75−83.

(3) Taber, K. S. Revisiting the chemistry triplet: drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chem. Educ. Res. Pract.* **2013**, *14* (2), 156−168.

(4) Sjöström, J.; Talanquer, V. Humanizing Chemistry Education: From Simple Contextualization to Multifaceted Problematization. *J. Chem. Educ.* **2014**, *91* (8), 1125−1131.

(5) Nakhleh, M. B.; Mitchell, R. C. Concept learning versus problem solving: There is a difference. *J. Chem. Educ.* **1993**, *70* (3), 190.

(6) Bunce, D. M. Enhancing and assessing conceptual understanding. In *Sputnik to Smartphones: A Half-Century of Chemistry Education*; ACS Publications: Washington, D.C., 2015; pp 107−119.

(7) Phelps, A. J. Teaching to enhance problem solving: it's more than the numbers. *J. Chem. Educ.* **1996**, *73* (4), 301.

(8) Smith, K. J.; Metz, P. A. Evaluating Student Understanding of Solution Chemistry through Microscopic Representations. *J. Chem. Educ.* **1996**, *73* (3), 233.

(9) Taskin, V.; Bernholt, S. Students' Understanding of Chemical Formulae: A review of empirical research. *Int. J. Sci. Educ.* **2014**, *36* (1), 157−185.

(10) Mahaffy, P. G.; Holme, T. A.; Martin-Visscher, L.; Martin, B. E.; Versprille, A.; Kirchhoff, M.; McKenzie, L.; Towns, M. Beyond "Inert" Ideas to Teaching General Chemistry from Rich Contexts: Visualizing the Chemistry of Climate Change (VC3). *J. Chem. Educ.* **2017**, *94* (8), 1027−1035.

(11) Prins, G. T.; Bulte, A. M. W.; Pilot, A. An Activity-Based Instructional Framework for Transforming Authentic Modeling Practices into Meaningful Contexts for Learning in Science Education. *Sci. Educ.* **2016**, *100* (6), 1092−1123.

(12) Broman, K.; Parchmann, I. Students' application of chemical concepts when solving chemistry problems in different contexts. *Chem. Educ. Res. Pract.* **2014**, *15* (4), 516−529.

(13) Bransford, J. D.; Donovan, M. S. Scientific inquiry and how people learn. In *How Students Learn: History, Mathematics, and Science in the Classroom*; The National Academies Press: Washington, D.C., 2005; pp 397−420; DOI: 10.17226/10126.

(14) Evans, K. L.; Leinhardt, G.; Karabinos, M.; Yaron, D. Chemistry in the Field and Chemistry in the Classroom: A Cognitive Disconnect? *J. Chem. Educ.* **2006**, *83* (4), 655.

(15) Moore, E. B.; Chamberlain, J. M.; Parson, R.; Perkins, K. K. PhET Interactive Simulations: Transformative Tools for Teaching Chemistry. *J. Chem. Educ.* **2014**, *91* (8), 1191−1197.

(16) Stieff, M. Improving representational competence using molecular simulations embedded in inquiry activities. *J. Res. Sci. Teach.* **2011**, *48* (10), 1137−1158.

(17) Stieff, M.; Wilensky, U. Connected Chemistry—Incorporating Interactive Simulations into the Chemistry Classroom. *J. Sci. Educ. Technol.* **2003**, *12* (3), 285−302.

(18) Xie, C.; Tinker, R.; Tinker, B.; Pallant, A.; Damelin, D.; Berenfeld, B. Computational Experiments for Science Education. *Science* **2011**, *332* (6037), 1516−1517.

(19) Pallant, A.; Tinker, R. F. Reasoning with Atomic-Scale Molecular Dynamic Models. *J. Sci. Educ. Technol.* **2004**, *13* (1), 51−66.

(20) Plass, J. L.; Milne, C.; Homer, B. D.; Schwartz, R. N.; Hayward, E. O.; Jordan, T.; Verkuilen, J.; Ng, F.; Wang, Y.; Barrientos, J. Investigating the effectiveness of computer simulations for chemistry learning. *J. Res. Sci. Teach.* **2012**, *49* (3), 394−419.

(21) Tatli, Z.; Ayas, A. Effect of a Virtual Chemistry Laboratory on Students' Achievement. *J. Educ. Technol. Soc.* **2013**, *16* (1), 159−170.

(22) Schwab, Z. Growing Stem Students: How Late Nite Labs' Online Platform is Spreading Science and Saving Schools' Resources. *J. Educ. Technol. & Sys.* **2013**, *41* (4), 333−345.

(23) Yaron, D.; Karabinos, M.; Lange, D.; Greeno, J. G.; Leinhardt, G. The ChemCollective—Virtual Labs for Introductory Chemistry Courses. *Science* **2010**, *328* (5978), 584−585.

(24) Winberg, T. M.; Berg, C. A. R. Students' cognitive focus during a chemistry laboratory exercise: Effects of a computer-simulated prelab. *J. Res. Sci. Teach.* **2007**, *44* (8), 1108−1133.

(25) Donnelly, D.; O'Reilly, J.; McGarr, O. Enhancing the Student Experiment Experience: Visible Scientific Inquiry Through a Virtual Chemistry Laboratory. *Res. Sci. Educ.* **2013**, *43* (4), 1571−1592.

(26) Rutten, N.; van Joolingen, W. R.; van der Veen, J. T. The learning effects of computer simulations in science education. *Comput. Educ.* **2012**, *58* (1), 136−153.

(27) Clark, R. C.; Mayer, R. E. *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*; John Wiley & Sons: Hoboken, NJ, 2016.

(28) Goldman, S. R. Learning in Complex Domains: When and Why Do Multiple Representations Help? Commentary. *Learn. Instr.* **2003**, *13* (2), 239−244.

(29) Ainsworth, S. DeFT: A conceptual framework for considering learning with multiple representations. *Learn. Instr.* **2006**, *16* (3), 183−198.

(30) Ainsworth, S. The functions of multiple representations. *Comput. Educ* **1999**, *33* (2), 131−152.

(31) McElhaney, K. W.; Chang, H.-Y.; Chiu, J. L.; Linn, M. C. Evidence for effective uses of dynamic visualisations in science curriculum materials. *Stud. Sci.. Ed.* **2015**, *51* (1), 49−85.

(32) Ploetzner, R.; Lowe, R. Looking Across Instead of Back and Forth: How the Simultaneous Presentation of Multiple Animation Episodes Facilitates Learning. In *Learning from Dynamic Visualization*; Springer: Cham, Switzerland, 2017; pp 51−68.

(33) Rau, M. A. Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. *Educ. Psychol. Rev.* **2017**, *29* (4), 717−761.

(34) Roediger, H. L.; Karpicke, J. D. Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychol. Sci.* **2006**, *17* (3), 249−255.

(35) Pashler, H.; Bain, P. M.; Bottge, B. A.; Graesser, A.; Koedinger, K.; McDaniel, M.; Metcalfe, J. *Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007−2004; National Center for Education Research*; Institute of Education Sciences, U.S. Department of Education, Washington, D.C., 2007.

(36) Black, P.; Wiliam, D. Assessment and Classroom Learning. *Assess. Edu. Princ. Polym. Pract* **1998**, *5* (1), 7−74.

(37) Koedinger, K. R.; Aleven, V. An Interview Reflection on "Intelligent Tutoring Goes to School in the Big City. *Int. J. Artif. Intell. Educ* **2016**, *26* (1), 13−24.

(38) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Erlbaum Associates: Hillsdale, NJ, 1988.

(39) Lipsey, M. W.; Puzio, K.; Yun, C.; Hebert, M. A.; Steinka-Fry, K.; Cole, M. W.; Roberts, M.; Anthony, K. S.; Busick, M. D. *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. National Center for Special Education Research*; Institute of Education Sciences, U.S. Department of Education: Washington, D.C., 2012.

(40) Freeman, S.; Eddy, S. L.; McDonough, M.; Smith, M. K.; Okoroafor, N.; Jordt, H.; Wenderoth, M. P. Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (23), 8410−8415.

(41) Kyndt, E.; Raes, E.; Lismont, B.; Timmers, F.; Cascallar, E.; Dochy, F. A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educ. Res. Rev.* **2013**, *10*, 133−149.

(42) Apugliese, A.; Lewis, S. E. Impact of instructional decisions on the effectiveness of cooperative learning in chemistry through meta-analysis. *Chem. Educ. Res. Pract.* **2017**, *18* (1), 271−278.

DOI: 10.1021/acs.jchemed.8b00048
J. Chem. Educ. XXXX, XXX, XXX−XXX