UC RIVERSIDE

Academic Senate

GRADUATE COUNCIL

June 2, 2022

To: Gregory Beran, Graduate Advisor
Department of Chemistry

From: Don Collins, Chair
Graduate Council

Re: Designated Emphasis in Chemistry Education – Revised

The Graduate Council voted to approve this proposed new Designated Emphasis as written. The approved proposal and catalog copy are attached for your reference.

cc: Christina Youhas, Chemistry
Kara Oswood, Graduate Division
Luis Bravo, Student Affairs, Marketing & Communications
Cortney Crooms, Student Affairs, Marketing & Communications
Accreditation Liaison Officer
Registrar Dailey

May 24, 2022

Dear UCR Graduate Council,

Thank you for the feedback you provided in March 2022 regarding the Chemistry Department's proposal for a new graduate-level Designated Emphasis (DE) in Chemical Education. We sincerely appreciate your suggestions regarding incorporating classes offered by the Graduate School of Education. After considerable internal discussion of the Grad Council feedback and our department's goals, however, we believe that the existing DE proposal course requirements provide the best fit for the proposed DE. The revised proposal being submitted here does not change the previously proposed coursework requirements, but we have added a new appendix to the proposal in which we articulate the specific rationale behind those coursework requirements more clearly by relating them to specific examples of the types of research output that would be expected for students in the DE.

The following documents are included:
1) The proposal, with changes from the prior version highlighted in blue *(only trivial changes were made since last submission)*.
2) An appendix to the proposal providing more thorough justification for the proposed course requirements *(entirely new since last submission)*.
3) List of Catalog Changes *(unchanged since last submission)*.
4) A letter of support from Psychology *(unchanged since last submission)*.

If you have any questions, please contact me. Sincerely,

Gregory Beran

Prof. Greg Beran
Chemistry Graduate Advisor
gregory.beran@ucr.edu

Proposal to Establish a Designated Emphasis (DE) in Chemistry Education

Contact: Gregory Beran, Chemistry Department Graduate Advisor gregory.beran@ucr.edu

May 24, 2022

Approvals:

- Chemistry Department Faculty Vote, Revised Proposal:	19 Yes, 0 No, 1 Abstain.	2/8/2022
- Chemistry Department Faculty Vote:	21 Yes, 0 No, 1 Abstain.	11/9/2021
- Psychology Department:	Letter of Support is attached.	

A. Goals/Benefits of a Chemistry Education DE

1. Between 2014-19, 27% of Chemistry PhD graduates pursued a career in teaching immediately after graduating (mostly through community college, lecturer, and visiting professor positions). These teaching positions provide opportunities for recent graduates to conduct research in chemistry education. The opportunity to complete the DE in Chemistry Education will provide students with training the quantitative methods in education to make them more competitive for jobs with a strong teaching focus.

2. Creating the Chemistry Education DE will distinguish the UCR Chemistry PhD program from those at other UC campuses and regional universities. This will provide our department an opportunity to recruit a distinctive population of students who are interested in pursuing a career path in chemistry education. For example, completing the Chemistry Education DE would make students more competitive for chemistry education post-doctoral positions, which have been expanding rapidly over the past decade as more departments across the country have created PhD-level chemistry education research programs. Programs with specific PhD programs in Chemistry Education include: Purdue University, University of Iowa, University of Colorado-Boulder, Miami University (OH), UMass-Boston, University of Northern Colorado, University of South Florida, University of New Hampshire. Many other departments have faculty who focus on Chemistry Education research as part of their conventional PhD programs.

3. Creating the Chemistry Education DE will expand the department's opportunities to pursue NSF funding related to STEM education research. If awarded, such grants would provide new funding streams to support graduate students and increase the indirect cost return for the university. Significant funding is available through the NSF Core Research program, for example:

https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504924

Not only will the Chemistry Education DE provide a more consistent group of graduate students to work with the department's instructional faculty, but having a formal program should make the department more competitive for this funding in the eyes of the NSF.

B. Structure and Requirements for Chemistry Education DE

1. The DE will require 12 units of coursework; students could satisfy by taking a selection of the following courses:

a. Required Course:

CHEM 241: Learning Theories in Chemistry Education (4 units; no course prerequisite)

This new course will be created and submitted for Senate approval during the 2021-22 academic year. Topics will include, but will not be limited to: representational competence/Johnstone's triangle, constructivist theory of learning, Ausubel's assimilation theory, embodied learning, cognitive load theory, and the cognitive theory of multi-media learning). The cross-listed undergraduate course can act as a free elective for chemistry majors, but will also be appealing for chemistry/STEM majors in the Science-Math Initiative pre-service teacher program. Creating the cross-listed undergraduate course will also help ensure minimum enrollments are met to sustain the course. Graduate students from the Graduate School of Education STEM learning program will also be invited to enroll.

b. One graduate courses in the Department of Psychology:

PSYC 211: Statistical Inference (4 units; no course prerequisite)

c. One of the following courses (students will be advised to take the course which will be most relevant based on the type of work being done in the research project and/or the student's longer term career goals; if the student is interested in a chemistry education research career path PSYC 212 may be more appropriate, but if a student is more interested in a teaching-focused career PSYC 207C may be more appropriate):

PSYC 212: Multiple Regression and Correlation Analysis (4 units; course prerequisite is PSYC 211)

PSYC 207C: Processes of Cognitive Development (3 units; no course prerequisite)

An overall GPA of at least 3.0 must be obtained in these three courses. More detailed rationale underlying these specific course requirements and their connections to what is involved in preparing publishable chemical education research studies is provided in Appendix 1.

2. The DE will require the student to complete a chemistry education research project (CHEM 299, 4 units) in addition to their normal chemistry research. This scope of this project is expected to be such that it comprises one chapter in the final PhD dissertation. This project may be completed under the guidance of a chemistry department professor of teaching (who would act as a co-advisor overseeing the chemistry education research), or could be completed independently under the approval of the DE Program Committee and the student's research advisor.

3. A DE Program Committee will be established to oversee the program curriculum and review student applications to the DE. The committee will consist of one professor of teaching, the Department of Chemistry Graduate Advisor, and one additional ad hoc faculty member.

4. Students cannot begin the DE coursework until candidacy has been completed within the regular research program. The DE coursework should be completed within one year after the completion of candidacy, and the chemistry education research project can be completed any time prior to the final submission of the dissertation.

5. Students will apply to enter the DE program by submitting a statement of interest to the DE Program Committee. The statement should include the following: 1) why the student is interested in completing

the Chem Ed DE; 2) how the Chem Ed DE would fit within the student's career goals; and 3) what type of project the student is interested in pursuing (if the student does not have a specific project/research question in mind, that will not prevent them from being accepted into the program). The statement of interest can be submitted by students prior to their candidacy exam in order to plan in advance of the program start.

6. The DE program will initially target three students per graduate student class. This will help ensure the existing teaching faculty are not overburdened and ensure the Psychology graduate courses can accommodate the added student enrollment. This target can be increased in future years as extramural funding is obtained and/or additional teaching faculty are hired in the department.

C. Potential Impacts on Graduate Student Research/Teaching Activities

The extra coursework and research for students completing the Chemistry Education DE will have require time that might otherwise be spent on dissertation research. Nevertheless, it is expected that the program structure will not delay advancement to candidacy (since the coursework will be taken afterwards). With focus and efficient work, students should be able to complete the PhD requirements without exceeding standard program graduation timelines (normative time to PhD is 15 quarters, current mean time to degree is 15.1 quarters). Any extramural funding secured would provide additional resources to support students participating in the DE.

Appendix 1: Rationale behind the proposed DE coursework requirements

Introduction

This appendix discusses the relationships between the proposed DE course requirements and the skills students will need to acquire to perform the chemical education research required by the DE. For context, three peer-reviewed papers that were published in the *Journal of Chemical Education* and *Chemistry Education Research and Practice* are attached. These are well-regarded journals in the chemical education sub-discipline with impact factors of 3.0 and which are published by the American Chemical Society or the Royal Society of Chemistry. These three publications represent research previously carried out by Chemistry department PhD students Lance Talbert and Hoi-Ting Wu in collaboration with Professor Jack Eichler. These projects were completed through an informal collaboration with Prof. Eichler that was done in addition to their primary dissertation research projects that lie in other sub-disciplines of chemistry. The results were (for L.T.) or will be (for H.-T.W.) included as chapters in the students' final PhD dissertations. The success of informal collaborations like these inspired the department's current efforts to establish a formal Chemistry Education DE that will encourage, recognize, and support research studies like these. These three publications are representative of the type of research output that students in the DE program will be expected to produce:

Pub #1: "Revisiting the use of concept maps in a large enrollment general chemistry course: implementation and assessment." *Chem. Educ. Res. Pract*. **21**, 37 (2020).

Pub #2: "Efficacy of an Asynchronous Online Preparatory Chemistry Course: An Observational Study." *J. Chem. Educ.* **97**, 4287 (2020).

Pub #3: "Incorporating concept development activities into a flipped classroom structure: using PhET simulations to put a twist on the flip." *Chem. Educ. Res. Pract*. **22**, 842 (2021).

The next section discusses how the material taught in the required and elective courses relates to the skills involved in completing studies like Pubs #1-3.

Relationship between DE coursework requirements and necessary chemical education research skills

1. To carry out studies such as Pubs #1-3, students need to be familiar with the broader theoretical frameworks of learning. This background knowledge aids in the experimental design process and is also required when publishing educational research studies. See the Introduction sections of the Pubs #1-3 for representative examples of how the theoretical frameworks must be used as the research scaffold. All students in the DE will be required to take the new graduate-level CHEM 241 course (see catalog description below) which will discuss the theoretical frameworks of learning that are most relevant to chemistry, such as Johnstone's Triangle and Chemical Representations, Cognitive Load Theory, Cognitive Theory of Multimedia Learning, Embodied Learning and Spatial Reasoning, etc. CHEM 241 will also introduce experimental designs in educational research, assessment, and measuring educational outcomes that will prepare students how to design and carry out their chemical education studies.

> CHEM 241 Foundations of Chemistry Education Research (4) Lecture, 3 hours; discussion, 1 hour. Prerequisite(s): Consent of Instructor; graduate standing. Provides an overview of the theoretical frameworks of learning relevant to chemistry and the typical research methodologies used in chemistry education research. Includes a discussion of experimental design considerations and an introduction to quantitative data analysis.

2. The quantitative analyses employed in Pubs #1-3 are common in chemistry education research and are modeled on quantitative social science methods. The required PSYC 211 course will provide training in

hypothesis testing, human subjects experimental design, measurement of psychological data (including data related to learning), and inferential statistical analysis (see course catalog description below). This course will be critical in providing the Chemistry Education DE students the background needed to carry out the type of research reported in all three attached publications. Note that chemistry PhD students generally have strong quantitative skills and have received some statistics training (in the context of laboratory studies), but they often have not been exposed to the types of statistical inference and other techniques used in social science and educational research. The introduction to quantitative data analysis provided in CHEM 241 is intended to refresh students' memories in statistics and provide a bridge that prepares them to learn the topics taught in PSYC 211.

> PSYC 211 Statistical Inference (4) Lecture, 3 hours; discussion, 1 hour; laboratory, 2 hours. Prerequisite(s): graduate standing in Psychology or consent of instructor. Examines basic issues related to the application of statistical inference, effect size estimation, and significance tests to various research paradigms in psychology. Discusses aspects of psychological measurement and the appropriateness of particular statistical techniques to different types of psychological data.

3a. Students who opt to carry out a project that is more heavily quantitative and/or which requires more sophisticated data modeling/analysis will be encouraged to take PSYC 212 as their third (elective) graduate-level course for the DE. This course covers multiple regression analysis, analysis of variance, and multivariate analysis (see course catalog description below); such methods are sometimes used in chemical education research, as exemplified by Pub #2.

> PSYC 212 Multiple Regression and Correlation Analysis (4) Lecture, 3 hours; discussion, 1 hour; laboratory, 1 hour. Prerequisite(s): graduate standing in Psychology, PSYC 211; or consent of instructor. Multiple regression, the general linear model, their relationship to analysis of variance, and extensions to multivariate analysis. The use of assorted computer statistical packages.

3b. Students whose DE research requires a deeper understanding of the theories of human learning, cognition, and information processing will be encouraged to take PSYC 207C as their third (elective) graduate-level course for the DE (instead of PSYC 212). PSYC 207C builds on the material from CHEM 241 and supports the DE learning objectives of being able to design chemical education research projects that are grounded in the current understanding of learning processes. For example, studies such as Pub #3 require a strong understanding of cognitive load, attention, information processing, and other topics which are covered in detail by PSYC 207C.

> PSYC 207C Processes of Cognitive Development (3) Lecture, 3 hours. Prerequisite(s): consent of instructor. Examines the cognitive changes in humans throughout the life cycle. Topics include Piagetian theory and memory, information processing, attention, and intelligence with a focus on the changes that occur in these skills.

In summary, the required CHEM 241, PSYC211, and either PSYC 207C or PSYC 212 courses will provide students in the Chemistry Education DE with a strong foundation in learning theories, experimental design in social science, and the statistics/data analysis tools that will prepare them to complete the research components of the DE.

Chemistry Education Research and Practice

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Educ. Res. Pract.*, 2020, **21**, 37

Revisiting the use of concept maps in a large enrollment general chemistry course: implementation and assessment†

Lance E. Talbert, James Bonner, Kiana Mortezaei, Cybill Guregyan, Grace Henbest and Jack F. Eichler *

In an effort to improve student conceptual understanding and help students better connect pre-existing knowledge to new ideas, a concept map assignment was implemented in a first-year college level general chemistry course. This implementation included a quasi-experiment that was carried out in discussion group recitation sections within a third-quarter general chemistry course. Students enrolled in a single section of the course were divided into two groups in which a concept map treatment was compared to a control group that completed short journal entries. Comparison of a concept inventory post-test using an independent samples t-test indicates students in the concept map treatment appear to perform better than the students in the journal control group (t = 2.34, mean difference = 0.844, $p < 0.05$). However, a multi-variable regression analysis in which the concept inventory post-test scores were compared between the treatment and control groups, while traits related to incoming academic preparation were held constant, suggests there was no significant difference in performance (unstandardized b = 0.222, p = 0.540). The quality of the students' concept maps was also evaluated and correlated to student performance on the concept inventory, and it appears students who were better at concept mapping made greater gains in conceptual understanding (Pearson's r = 0.295, $p < 0.05$). When the relationship between the quality of concept mapping and concept inventory post-test was determined while holding constant covariates related to incoming academic preparation, the unstandardized B coefficient was positive, but was not significant at the p = 0.05 level (unstandardized b = 0.215, p = 0.134) This study does not provide unequivocal evidence that a concept map treatment leads to greater gains in conceptual understanding compared to a control population, or that students with better concept mapping skills performed better on the concept inventory instrument. Nevertheless, a template for implementing a concept map assignment in a large enrollment course is provided, and the results presented herein might prompt chemistry instructors to consider including concept map assignments in their instructional toolbox.

Received 26th February 2019,
Accepted 24th June 2019

DOI: 10.1039/c9rp00059c

rsc.li/cerp

Introduction

General chemistry is often the first science class taken by students in their undergraduate science curriculum. Unfortunately, historically high failure rates have given this class "gatekeeper" status for students who wish to major in STEM fields. One of the factors contributing to the lack of student success is the fact students often enter their first undergraduate chemistry course possessing misconceived mental models that thereby become a barrier to learning the foundational concepts covered in the general chemistry curriculum (Cros *et al.*, 1986; Mulford and Robinson, 2002; Harrison and Treagust, 2018). Even students who have performed well on typical classroom assessments in their general chemistry courses have been found to struggle when asked to provide conceptual explanations for questions related to core learning objectives in the general chemistry curriculum (Cooper *et al.*, 2013).

To overcome the limitations typical classroom assessments possess in regards to helping students develop clear and cogent mental models and conceptual understanding, considerable effort has been given to develop and employ metacognitive interventions (Novak, 1990; Rickey and Stacy, 2000). Metacognition is the process utilized to evaluate and monitor one's understanding and performance of the material. There are currently several strategies which are used to enhance metacognition, including paraphrasing and rewriting, working on homework problems, previewing material, and pretending to teach information (Rickey and Stacy, 2000). While the use of these strategies has been reported in the educational research literature

Department of Chemistry, University of California, Riverside, 501 Big Springs Road, Riverside, CA 92521, USA. E-mail: jack.eichler@ucr.edu; Tel: +1-(951)-827-3794

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9rp00059c

(Cook *et al.*, 2013), there remains a need to develop more widely applicable implementation strategies that can both help develop and assess student understanding of conceptual ideas (Johnstone, 1993; Gabel, 1999; Galloway and Bretz, 2015).

Anecdotal evidence has broadly confirmed that students in the University of California-Riverside (UCR) general chemistry program also continue to struggle to provide scientifically acceptable conceptual explanations for many fundamental learning objectives. We therefore decided to revisit the use of concept maps as both an intervention to help students better develop conceptual models and provide a means to measure this type of learning outcome. One of the first reports on the use of concept maps and their impact on student learning was published by Novak (1984), and concept maps have been subsequently proposed numerous times as a strategy to increase student retention and learning (Nicoll *et al.*, 2001; Francisco *et al.*, 2002; Kennedy, 2016). Generally speaking, a concept map requires students to reflect on their learning and the specific learning objectives by having them identify key concepts that have been covered in the course and using words or short phrases to link multiple concepts where appropriate. It is important for students to draw links between the concepts to establish the relationship between them and solidify their deeper understanding of the foundational course concepts. The process of making concept maps engages students in a form of active learning, and the continual cycle of reflection required to update the conceptual links makes concept mapping a potentially effective form of metacognitive engagement (Chevron, 2014).

The theoretical framework underpinning the impact of concept mapping on student conceptual understanding can be traced to Ausubel's assimilation theory of learning (Ausubel, 1968). Though this is reviewed in detail by Novak (1984), we highlight here that Ausubel proposes meaningful learning takes place when the learner relates new knowledge to concepts he/she already knows, and when the learner can identify the key concepts in the new knowledge and how to relate these to other concepts in other contexts. Ausubel points out if this connection of concepts from new knowledge to other concepts does not occur, verbatim/non-substantive learning can still occur, but this type of learning has less value and can actually interfere with subsequent learning. It is quite clear concept mapping can play an important role in promoting the type of meaningful learning described here (Nesbit and Adesope, 2006; Turan-Oluk and Ekmekci, 2018), and should certainly be more strongly considered by STEM instructors to be part of their instructional arsenal.

As alluded to above, not only does a concept map assignment provide an opportunity for students to develop more completely developed mental models, it can be a valuable tool for assessing the students' conceptual understanding (Novak and Gowin, 1984). Assessment of the effectiveness of student conceptual understanding might be a daunting task for many instructors, but fortunately a variety of rubrics have been developed for carrying out the evaluation of student conceptual thinking. Typical rubrics focus on the scoring of hierarchies presented by the students and/or assess at the validity of the proposed links in traditional concept map assignments (Novak and Gowin, 1984; Francisco *et al.*, 2002), whereas more recent studies have proposed creating new types of concept linking activities that can be more easily evaluated (Ye *et al.*, 2015).

Though concept maps are clearly a valuable intervention in developing and assessing student conceptual understanding, their use in first-year college level chemistry courses has not been broadly demonstrated. Previous studies on the use of concept maps in general chemistry courses are generally limited to small enrollment implementations ($n < 100$; Regis *et al.*, 1996; Markow and Lonning, 1998; Besterfield-Sacre *et al.*, 2013; Luxford and Bretz, 2014), and are often used only to map concepts within specific learning units of the course. Furthermore, for studies that have been carried out in large enrollment courses the analysis of student data has generally been limited to a small subset of the class population (Burrows and Mooring, 2015). Perhaps the most noteworthy implementation of a concept map intervention in a large enrollment general chemistry course comes from Francisco and coworkers (2002). In this study, the implementation focused not only on the student aspect of completing the concept map assignments, but also incorporated graduate student teaching assistants (TAs) into the grading of concept maps and delivery of feedback to the students. It was shown concept maps in large enrollment classes do positively impact the development of student conceptual knowledge, in particular demonstrating the concept map intervention appeared to improve student performance on complex multi-step algorithmic problems. However, Francisco and coworkers also reported significant resistance from the students to include the concept maps as graded activities and the implementation required significant training of the TAs in order to achieve an observable positive impact.

Hence, the goal was to create a concept map implementation for a large enrollment general chemistry course that achieved the following objectives: (1) balance the need to create a significant incentive for students to complete the concept maps with the desire to avoid student resistance to being graded on a subjective measure; (2) provide a template for a concept map assignment that can be more easily adopted by other instructors of large enrollment courses; and (3) allow for the design of a quasi-experiment that adds to the limited pool of data describing the impact of concept map assignments on student learning outcomes in large enrollment general chemistry courses. The experimental hypotheses that drove the research design were: (1) students who completed a quarter-long concept map intervention would achieve more significant learning gains compared to a control group of students who did not use concept mapping; and (2) higher proficiency in creating valid and well-developed concept maps would correlate to gains in conceptual understanding. Herein, we will describe how a concept map assignment was administered in a streamlined fashion using TA-led discussion group recitation sessions. We will also describe results from a quasi-experiment in which student performance on a concept inventory and student survey responses were compared between a concept map treatment group, and a control group in which students completed weekly journal entries.

Implementation design and research methods

Course description

The concept map implementation was carried out in a third quarter general chemistry course (CHEM 001C). This is the third course in the same three-quarter general chemistry sequence offered at UCR. The topics covered in this course include: chemical equilibrium; acid–base chemistry; buffers and titrations; electrochemistry; coordination chemistry; and nuclear chemistry. This course was taught for an "on sequence" cohort of first-year students in the spring of 2018 (S18), but also included second year students and upper-class students who may have needed the course as a general college science requirement or a pre-requisite for medical and other health-related professional schools.

Quasi-experimental design

The S18 course consisted of a large enrollment lecture, which met two times per week for 80 minutes, and associated recitation sections that met once per week for 50 minutes (30–40 students each). Approximately one-third of the lecture meetings used flipped classroom modules (Eichler and Peeples, 2016), while the remaining class periods included a mixture of lecture, peer-to-peer discussion, and interactive clicker questions. The quasi-experimental design assigned recitation sections taught by a TA who had prepared in advance to implement the concept map intervention as the treatment group (n = 115). The remaining recitation sections, taught by a different TA, were assigned as the control group and required the students to complete a weekly journal entry in lieu of a concept map (n = 123). This design also allowed the instructors to assign similar workloads and use the same grading scheme in both groups. To minimize potential "treatment" effects in the journal entry control group, these students wrote weekly journals that summarized what they had learned in each week's lectures and no guidance regarding how to structure the journal entries was provided. Students were simply instructed to describe what they learned in lecture each week. Conversely, the concept map treatment group was instructed to build on the map from the previous week, and to emphasize important concepts from each chapter and show the relationship between various chapters.

Each concept map or journal was awarded a single point for nine of the ten weeks, and four points for the final week. Though all assignments were graded simply for completion and these points were awarded as extra credit toward the overall lecture grade, students were informed that no points would be awarded if they simply submitted their previous concept map. During the course of the term the TA found no instances in which students handed in unchanged concept maps from previous weeks. This grading design was chosen in an effort to promote student compliance in completing the weekly concept map assignment without creating the type of resistance one might expect to arise when subjective assignments such as this are graded more rigorously.

To increase the convenience of collecting and evaluating the concept maps, students were required to use a concept mapping program. Cmap is a freely available software program that students were able to download to their computers for the development of the concept maps, and students were able to save and submit their maps as PDF files (Cañas *et al.*, 2005; Novak and Cañas, 2008).‡ The TA was familiarized in the use of the Cmap program and was able to answer any questions students had with the program. To help ensure students were able to navigate the Cmap program and properly save their concept maps, the TA created a video tutorial that the students could view anytime during the term on the online course management system.

Questions regarding the Cmap program of the creation of concept maps were covered in the first TA-facilitated recitation meetings. Students were instructed to develop their concept maps freely and were only provided an outline of a concept map containing the chapters which would be covered during the course (see Appendix 1, ESI†). The TA spent approximately 30 minutes of the first discussion group session providing an overview of how to create a concept map and why the concept maps were being included as an ongoing assignment. Aside from being given the preliminary concept outline, students had complete control over their concept maps and were instructed to build their concept maps using the material they had learned each week. This enabled students to have full control over how they would build the concept map and in what way they would connect key concepts to one another. Guidance in developing the concept maps was provided weekly through class-wide verbal feedback, which focused on highlighting key misconceptions identified by the TA. Approximately 10 minutes were devoted each week for this weekly feedback. Following each midterm exam, an instructor-completed concept map was shown to the students in their discussion sections and the TA led a dialogue to help elucidate for the students how proper connecting terms should be constructed. These post-exam debriefs took up approximately 10–15 minutes of discussion group time, and helped model for the students the types of conceptual thinking that should be used for the remainder of the quarter as they continued to build their maps. In the last discussion group session prior to the final exam, the TA spent approximately 15–20 minutes discussing with the students how they should use their final concept map as a study tool for the final exam and reviewed the traits of a well-developed concept map.

Grading rubric for concept maps

Concept maps were scored using an adapted version of the concept map rubric developed by Besterfield-Sacre and co-workers (2013). Concept maps were scored on comprehensiveness of the covered material, organization and linking between chapters, the correctness of the material in each chapter, and the correctness of the links between concepts (see Table 1). To ensure there was consistency in how the scoring rubric was used to allocate points, two coders discussed how points were to be awarded and each independently scored a random sample of 10 concept

‡ Cmap concept mapping software: https://cmap.ihmc.us/.

maps for comprehensiveness and organization/links (the graders consisted of a graduate student teaching assistant and an undergraduate research assistant). The course instructor then met with the two coders to discuss how the rubrics were applied in the grading and clarified how the student responses should be evaluated. The coders then independently evaluated another random sample of 10 concept maps and it was found that they agreed on the scoring on greater than 80% of the items. The remaining concept maps were scored for the comprehensiveness and organization/links categories. To ensure consistency and accuracy for scoring in the correctness category two experienced chemistry instructors evaluated a random sample of concept maps, and scores were compared in the same manner as described above to calibrate the application of the rubric. An additional random sample of concept maps were evaluated by both instructors to ensure over 80% of the items were coded in an equivalent fashion, and the remaining concept maps were coded for correctness.

Fig. 1 and 2 illustrate examples of a well-developed and poorly developed concept map, respectively. In Fig. 1, the student used a color-coded system to connect sub topics to broader concepts and the concept map has several branches linking the broader concepts to each other. The connecting phrases between the various key concepts are also well developed and almost universally accurate. The concept map shown in Fig. 2 lacks many of the key concepts covered in each chapter and does not include many of the sub-topics observed in the more developed concept maps. Additionally, many of the connections that should have been made between the broader key concepts were not observed (*e.g.*, other than a single connection between ICE table and salt solutions, there are no connections drawn between any of the other chapters).

If readers are interested in implementing the concept map intervention as described here an instructor guide is included in the Appendices (see Appendix 2). This includes instructor notes that provide suggestions for preparing TA's to facilitate the weekly assignments, how the TA included the concept map assignment into the weekly recitation discussions, and how periodic feedback on the concept maps was provided to the students. Instructions for using the Cmap program and the concept map inventory used in this study are also provided (see Appendix 2).

Concept inventory, SALG survey, and data collection

To test the first quasi-experimental hypotheses a concept inventory and student self-assessment of leaning survey (SALG) was used in a pre-test/post-test alternative treatment/control group design as described by Barbera and co-workers (Mack, 2019a), and all student data was collected under an approved human subjects protocol (UCR Institutional Review Board protocol HS-10-135). Though to our knowledge the efficacy of a concept map intervention has not been previously evaluated using this type of quasi-experimental design in a chemistry course, we would like to highlight the report from Burdo and O'Dwyer (2015) that assessed the impact of a concept map treatment in an undergraduate physiology course. In this prior study the concept map treatment group was compared to a negative control group and a second treatment group in which retrieval practice was chosen as the independent variable of interest. The experimental design described by Burdo and O'Dwyer acts as a good model for the quasi-experimental reported herein, though it is noted this previous study assessed the impact of the concept map treatment on more traditional course exams as opposed to a measure of conceptual understanding.

The concept inventory consisted of a total of 16 multiple choice questions and was graded out of a total score of 16 (1 point for each item). This concept inventory was administered as a pre-test for students in the treatment and control groups in the first week of recitation. The pre-test was administered using the online quiz function in the course management system. Students were not able to access the quiz once it was completed and they were not allowed to view the questions or answers at any point during the quarter. The same questions from the concept pre-test were placed into the final exam for students in both the treatment and control groups, and this post-test was used to measure the improvement in conceptual understanding throughout the ten-week quarter. Any subsequent references to the concept inventory "post-test" are restricted to the 16 concept inventory questions that were embedded in the final exam. Any references to the "final exam" include the entire exam, which consisted of the 16 concept inventory questions, 14 additional algorithmic multiple-choice questions, and five free response questions. Finally, the authors note it would have been ideal to use a previously validated concept inventory to probe student gains in conceptual understanding. However, the unique combination of topics covered in this course dictated the use of a customized concept inventory that better matched the 12 respective course learning objectives. Because the concept inventory contained multiple dimensions (*i.e.*, five different categories of concepts) the stratified alpha (α_s) reliability coefficient was calculated as described by Widhiarso and Ravand (2014), and an item-analysis was carried out for the 16 test items (see Appendices 3–5, ESI†).

The SALG survey was also administer in a pre/post format using a modified version of the freely available instrument.

Table 1 Scoring rubric for concept maps. Each category is scored 1–3 with the total score for each concept map being scored out of 9

Points	Comprehensiveness	Organization/links	Correctness
3	Contains 1 or more key concepts in each chapter.	Map is clear and easy to follow. Contains links between 4 or more discussed chapters.	Map contains 1 or less errors.
2	Contains 1 or more key concepts in 2–4 chapters.	Map is slightly cluttered. Contains links between 3 chapters.	Map contains 2–3 errors.
1	Contains one or more key concepts in 0–1 chapters.	Map is not easy to follow. Contains links between only 2 chapters.	Contains 4 or more errors.

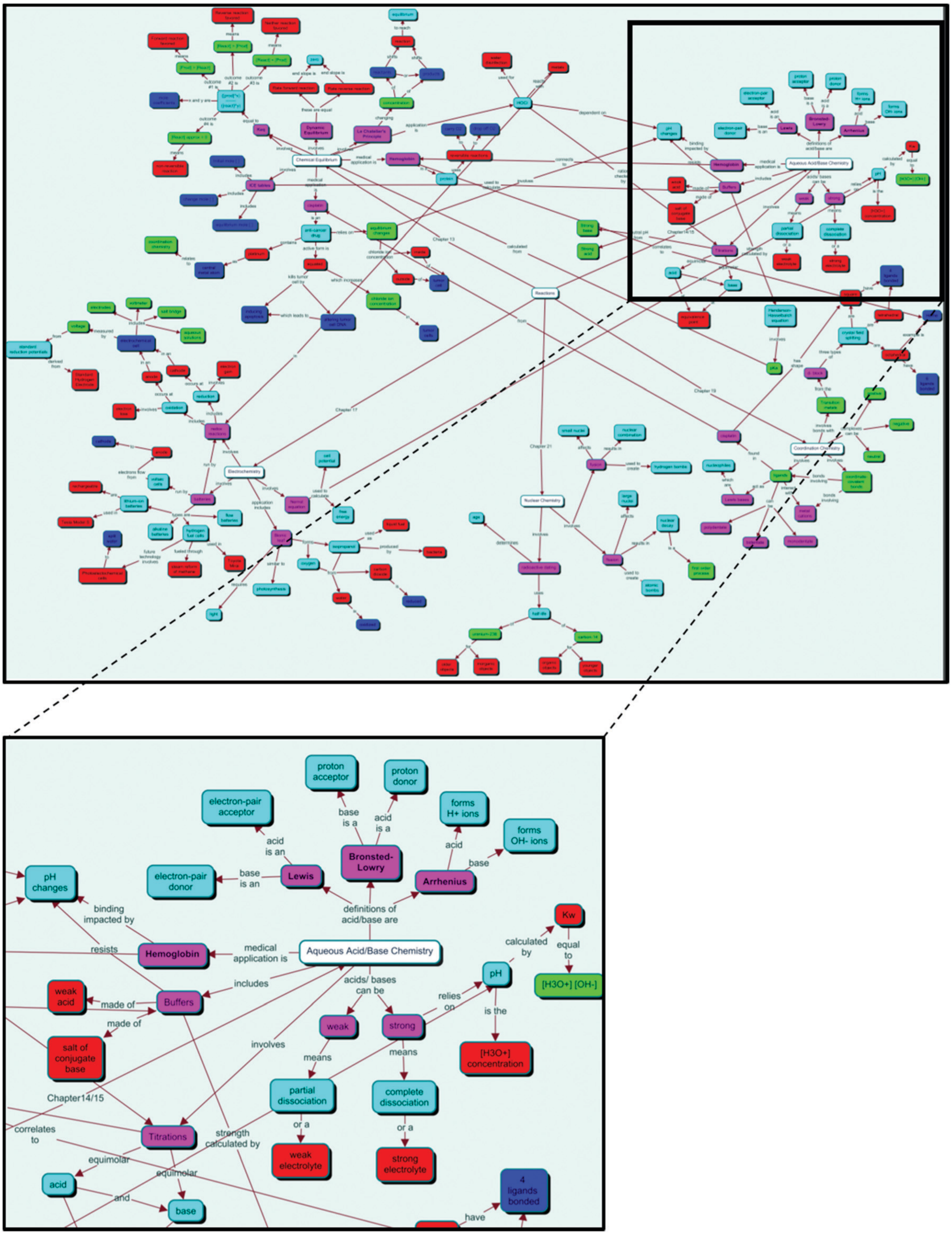


Fig. 1 Example of a well-developed concept map with correct connections and connecting phrases (comprehensiveness = 2, organization/links = 3, correctness = 3, total score = 8/9).

Questions that focused on conceptual learning gains, learning gains related to specific course topics, and learning gains related to applying course concepts in other contexts were chosen for the purpose of this study (see Table 7). The pre-SALG survey was available to students during a seven-day period starting two days before the first class and the post-SALG was available to students during a seven-day period starting two days before the final exam, and both surveys were completed using the online SALG interface. Students were informed if they completed both the pre- and post-SALG they would receive a small amount of extra

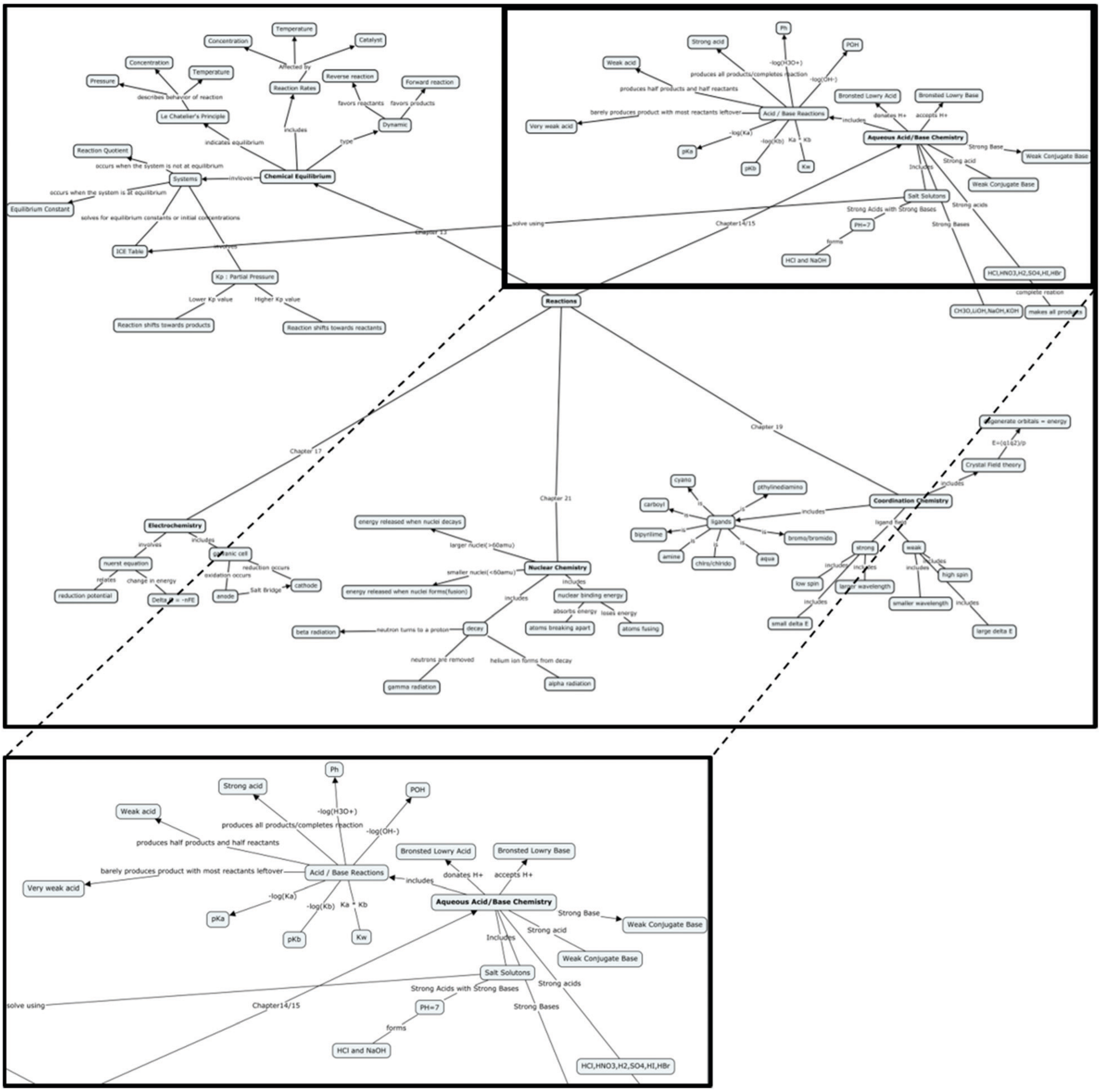


Fig. 2 Example of a less well-developed concept map with significant numbers of incorrect connections and connecting phrases (comprehensiveness = 1, organization/links = 1, correctness = 1, total score = 3/9).

credit on their final course point total (5 points out of 1000 total course points).

Statistical analyses

In the process of testing the quasi-experimental hypotheses, student performance on the pre- and post-test concept inventory was compared between the treatment and control groups. The incoming academic preparation for the students in both study populations were also compared. All statistical analyses were carried out using the SPSS Statistics 24 software package.§

§ IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

The incoming academic preparation of the students in the concept map treatment and journal entry control groups was compared using an analysis of variance (ANOVA). Concept inventory pre/post-test gains within the treatment and control groups were analyzed using paired *t*-tests, post-test scores were compared between the treatment and control group using an independent samples *t*-test, and a comparison of the performance on the post-test concept inventory between the treatment and control group was carried out using a multiple linear regression model. In order to impart more rigorous statistical control of the student incoming academic preparation, multiple regression analyses were carried out in which the concept inventory post-test scores were compared between treatment and control groups, while holding constant the concept inventory pre-test scores, high school overall grade point average (GPAs), and math SAT scores (Leech *et al.*, 2003). The relationship between the ability to construct high quality concept maps and performance on the concept inventory post-test was also evaluated by determining the Pearson correlation coefficient, and a multiple linear regression model was used to determine if the correlation between concept map rubric scores and concept inventory post-test scores was significant while holding constant the incoming student academic preparation. Analyses were carried out as described by Cohen (1988) in order to estimate the power to detect a significant regression coefficient in the multiple linear regression models. Finally, in an effort to obtain a preliminary overview of the student perception of the concept map treatment the descriptive responses on the SALG survey were compared between the concept map treatment and journal control groups.

Results and discussion

Table 2 summarizes the descriptive statistics for the control and treatment groups participating in the quasi-experiment. The academic preparation of students in the treatment and control groups appeared to be roughly equivalent when considering the distribution of pre-test scores and high school GPAs within each group, yet the concept map treatment group did appear to have slightly higher math SAT scores compared to the journal control group. The concept inventory pre-test scores, math SAT scores, and overall high school GPAs were compared using an ANOVA, and these results appear to corroborate the notion that there is no significant difference in the incoming academic abilities of the students between the two populations as measured by high school GPA and concept inventory pre-test (the null hypothesis stating there is no difference between the mean scores for all three of these independent variables cannot be rejected at the p = 0.05 level; see Appendix 6, ESI†). Conversely, the null hypothesis stating the math SAT scores are equivalent between groups can be rejected (F = 6.681, $p < 0.05$; see Appendix 5, ESI†).

The comparison of the concept inventory pre- and post-test suggests students in the treatment and control groups made gains in conceptual understanding during the course of the term. A paired t-test in which the average pre- and post-test concept inventory scores were compared indicates a statistically significant increase in average score for the post-test within both the treatment and control groups (journal control: mean difference = 4.3889; $t = -15.058$; $p < 0.001$; concept map treatment: mean difference = 4.61017; $t = -16.302$; $p < 0.001$; see Table 3). These results are important to note because, although the concept inventory questions were not selected from a previously validated question set, this instrument appears to measure gains in conceptual understanding with statistically significant results for both groups. The general utility of the concept inventory was confirmed by an item analysis and a single-administration internal consistency analysis. The item analysis indicates all of the questions on the inventory exhibited an item difficulty between 0.30–0.85 (see Appendix 3, ESI†), the item discrimination analysis revealed 12 of the 16 questions possess a point-biserial correlation greater than 0.20 (see Appendix 4, ESI†), and the stratified alpha reliability coefficient (α_s) was found to be 0.661, suggesting moderate reliability for the concept inventory (see Appendix 5, ESI†).

To begin evaluating the first research hypothesis, in which it was stated the concept map treatment would yield greater gains in conceptual learning relative to the journal control group, an independent samples t-test was used to compare the mean scores on the concept inventory post-test between the two groups (see Table 4). Based on this independent samples t-test, it appears the concept map treatment group performed better on the post-test relative to the journal control group (mean difference = 0.844, p = 0.020). The mean difference observed between groups translates to an approximately 5% increase in exam performance for the concept map treatment group, and the Cohen's d suggests the concept map treatment had a moderate effect on the concept inventory performance. However, because the study groups in this quasi-experiment were not randomly assigned caution must be taken in over-interpreting these results. More specifically, because the assignment of the study groups was not randomized, relying solely on the comparison of means in which no background academic traits were statistically controlled was not prudent. Therefore, it was desired to carry out further analysis in which the academic preparation characteristics among the treatment and control groups could be statistically controlled.

Multiple regression analyses

Even though the quasi-experimental design is not able to control for the myriad of independent variables that likely impact student performance, including incoming academic preparation, a multiple regression analysis can be utilized to statistically control for these independent variables. Given the fact the math SAT scores appear to be slightly higher for the concept map treatment group, it is especially pertinent to hold this variable constant when comparing the concept inventory post-test scores between the treatment and control groups. The model employed here included three different independent variables: the concept inventory pre-test scores, overall high school GPAs, and math SAT scores. Because the quasi-experimental hypotheses are interested in determining the impact of the concept map treatment on conceptual understanding, the linear regression model included the concept inventory post-test as the dependent variable. The final exam contained a significant number of algorithmic problems that are likely less correlated to conceptual thinking, therefore this was not included in any of the analyses.

Since it was hypothesized that high school GPA, math SAT scores, and the concept inventory pre-test scores would likely correlate with performance on the concept inventory post-test, these co-variates were held constant when comparing the post-test score dependent variable between the treatment and control groups. Generally speaking, high school GPA is known to be a strong predictor of student success in college-level courses

Table 2 Descriptive statistics for the recitation section treatment and control groups (control = journal group; treatment = concept map)

	Control (n = 115)	Treatment (n = 123)
High School GPA Avg.	3.74 ± 0.26	3.77 ± 0.27
SAT Math Avg.	582 ± 80	612 ± 85
Pretest Avg. (out of 16)	5.57 ± 2.07	6.09 ± 2.60
Post test Avg. (out of 16)	9.91 ± 2.81	10.8 ± 2.7
Exam 1 Avg. (out of 100)	64.0 ± 15.7	66.0 ± 15.3
Exam 2 Avg. (out of 100)	75.1 ± 15.9	78.4 ± 14.5
Final exam total Avg. (out of 400)	244 ± 60	253 ± 59
Avg. rubric score on final concept/journal entry (out of 9)[a]	7.00 ± 1.23 (n = 100)	6.17 ± 1.55 (n = 97)
Asian	n = 46	n = 58
Black or African American	n = 3	n = 0
Hispanic or Latino	n = 45	n = 36
Multi-race/unknown/non-resident Alien	n = 11	n = 13
White	n = 9	n = 16
Male	n = 55	n = 58
Female	n = 58	n = 61
Gender not reported	n = 2	n = 2

[a] Not all students submitted a final concept map/journal. The n value designates the number of students that completed the final concept map/journal assignment.

Table 3 Paired t-test comparing concept inventory pre-test and post-test scores within the treatment and control groups for the concept map implementation

	n	Mean difference	Std. deviation	Std. error mean of difference	p
Journal control pre- *vs.* post-test mean	108	4.39	3.03	0.29	<0.001
Concept map treatment pre- *vs.* post-test mean	118	4.61	3.07	0.28	<0.001

Table 4 Independent samples t-test comparing concept inventory post-test between the concept map treatment and journal control groups

	t	df	Mean difference	Std. error mean of difference	p	Cohen's d effect size
Journal control (mean = 9.91) *vs.* concept map treatment (mean = 10.8)	2.34	235	0.84	0.36	<0.05	0.308

(Zwick and Sklar, 2005), and in fact has been shown to be a stronger predictor of success than standardized tests such as the ACT and SAT (Geiger and Santelices, 2007). Recent reports also suggest math SAT scores can be a strong predictor of student performance in general chemistry (Vincent-Ruz *et al.*, 2018; Mack *et al.*, 2019b), which suggests this variable is likely to be positively correlated to the dependent variable. These previous studies therefore suggest including both high school GPA and math SAT scores as co-variates in the regression model should help isolate the impact of the group participation (treatment *vs.* control) on the post-test scores. Including the concept inventory pre-test scores as a covariate fits within the creation of the explanatory model, as this assessment more directly measures existing conceptual understanding. Holding this variable constant is also expected to aid in isolating the impact of the concept map treatment on the final concept inventory performance.

The results of the final multiple regression analysis are shown in Table 5. The concept inventory post-test dependent variable output was compared between all participants in the treatment and control groups, while keeping constant the independent variables related to student academic preparation (high school GPA, SAT math scores, and the concept inventory pre-test scores). Though participation in the treatment group was positively related to performance on the concept inventory post-test, this result was not statistically significant (unstandardized b = 0.222, p = 0.540). The independent variables of high school GPA, math SAT, and concept inventory pre-test were also positively related to performance on the concept inventory post-test (unstandardized b = 0.174, 0.0150, and 0.235, respectively), though interestingly the impact of high school GPA was not statistically significant. Unsurprisingly, the concept inventory pre-test appeared to have the strongest relationship to performance on the post-test measure. Because the high school GPA, math SAT, and concept inventory pre-test covariates may be redundant in terms predicting student performance on the post-test, the correlations between these co-variates were calculated to determine if collinearity within the regression model might be present (see Appendix 7, ESI†). None of the pairs of independent variables was found to have a correlation constant greater than 0.282, therefore no further changes were made to the regression model.

A *post hoc* power analysis was carried out to estimate the model's power to detect a significant regression coefficient as described by Cohen (1988). This power estimate was carried out by comparing the R^2 of the full model to the R^2 of the model in which the group participation independent variable was not included (estimated power $\approx$0.05; see Table 5; see Appendix 8 (ESI†) for the

Table 5 Multiple regression analysis. Includes full class; dependent variable = concept inventory post-test. Group indicates coded treatment/control (journal control group = 0; concept map treatment group = 1)[a]

	Unstandardized coefficients		Standardized coefficients		
	b	Std. error	Beta	t	p
Constant	−1.03	2.92		−0.355	0.723
Group	0.222	0.362	0.0390	0.614	0.540
High School GPA	0.174	0.674	0.0160	0.258	0.796
SAT Math	0.015	0.002	0.439	6.71	<0.001
Concept inventory pre-test	0.235	0.080	0.439	2.94	<0.05

[a] $R = 0.534$; $R^2 = 0.285$; adjusted $R^2 = 0.270$; standard error of the estimate = 2.46; estimated power to detect a significant regression coefficient ≈0.05 (see Appendix 8, ESI for description of the power estimation).

regression model without the group participation independent variable). Though the multiple regression model suggests there was not a significant correlation between study group participation and concept inventory post-test, the power analysis suggests there is a relatively strong likelihood this model yields a false negative conclusion (*i.e.*, the erroneous retention of the null hypothesis). The *post hoc* power analysis included an estimate of the effect size index, f^2, for the change in R^2 of the full regression model compared to the regression model in which the class treatment independent variable was removed. Cohen identifies f^2 effect size indexes of 0.02, 0.15, and 0.35 as small, medium, and large, respectively (1988). The effect size index for the regression model reported here was estimated to be 0.0014 (see Appendix 8, ESI†). This suggests the concept map treatment indeed had a minimal effect on the concept inventory post-test, and provides some explanation as to why the sample size used in this study resulted in low statistical power. Cohen also describes how to estimate sample size required for a desired level power (1988), and a model that includes the same number of independent variables used in the current study would require approximately 618 participants to achieve a power of 0.80 with an f^2 effect size index of 0.02.

To evaluate the second research hypothesis, which stated more well-developed concept mapping skills will correlate to gains in conceptual understanding, the correlation between concept inventory post-test scores and final concept map rubric scores was determined. The Pearson's correlation coefficient (ρ) suggests there is indeed correlation between concept mapping skills and performance on the final concept inventory post-test ($\rho = 0.295, p < 0.05$; see Fig. 3). Because this correlation did not account for the potential impact of other confounding variables on student performance on the concept inventory, a multiple regression model was also used to estimate the relationship between concept mapping skills and performance on the concept inventory post-test while holding constant the students' incoming academic preparation. The model indicates there is a positive relationship between concept mapping skills and concept inventory post-test scores, yet it was not found to be statistically significant (see Table 6; unstandardized $b = 0.210$, $p = 0.147$). A *post hoc* power analysis was used to determine the power of the model to detect a significant regression coefficient as described above, and it is estimated this model might very well yield a false negative result (estimated power ≈0.24; see Table 6; see Appendix 9, ESI,† for the regression model without the concept map rubric independent variable, and calculation of the f^2 effect size index and a detailed description of how the power was estimated; it is noted this was a separate power

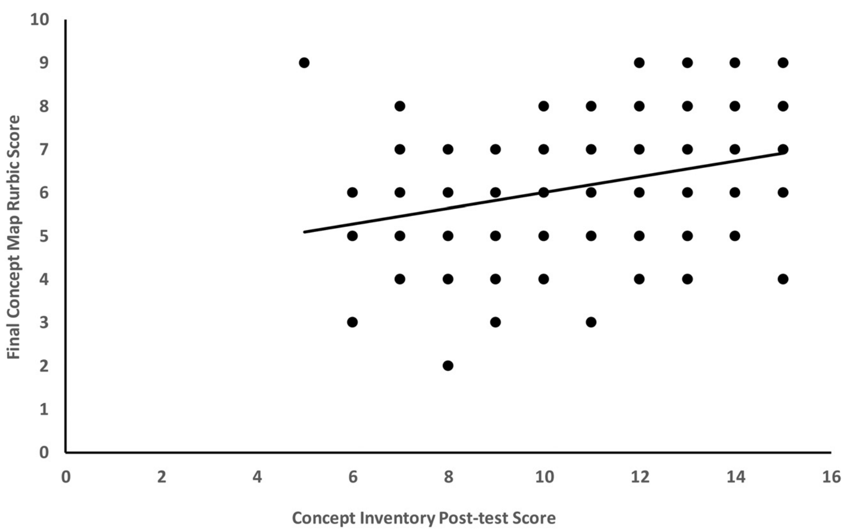


Fig. 3 Correlation of final concept map/journal entry rubric score with concept inventory post-test score (concept map rubric scores *vs.* concept inventory post-test scores; two-tailed Pearson correlation = 0.295; $p < 0.05$).

Table 6 Multiple regression of concept map rubric scores correlated to concept map post-test scores, holding constant math SAT, HS GPA, and concept inventory pre-test[a]

	Unstandardized coefficients		Standardized coefficients		
	b	Std. error	Beta	t	p
Constant	−0.881	4.300		−0.205	0.839
Final concept map rubric score	0.210	0.144	0.147	1.47	0.147
High School GPA	0.274	1.010	0.0260	0.271	0.787
SAT Math	0.0130	0.0030	0.424	4.00	<0.001
Concept inventory pre-test	0.225	0.106	0.217	2.11	<0.05

[a] $R = 0.586$; $R^2 = 0.344$; adjusted $R^2 = 0.307$; standard error of the estimate = 2.21; estimated power to detect a significant regression coefficient ≈0.24 (see Appendix 9, ESI for a description of the power estimation).

analysis from that described above for the regression model summarized in Table 5, and the two power analyses were carried out for the estimated effect sizes of the two regression models). In short, the analyses described above make it difficult to arrive at a definitive conclusion in regards to whether students found to create more well-developed concept maps may have performed better on the concept inventory post-test.

Affective perceptions of students – SALG survey

The quasi-experimental design employed in this study relied on delivering the concept map treatment during co-curricular recitation sections taught by graduate TAs. Because this treatment was a relatively small part of the larger course structure, its impact on the students' ability to make gains in conceptual understanding of the course content may have been limited relative to the learning interventions carried out in the main lecture during the course of the academic term. Therefore, the Student Assessment of Learning Gains (SALG)¶ was used to survey the student's perceptions of affective outcomes. Students were asked to gauge the gains made in connecting chemistry concepts to real world issues, their ability to connect chemistry concepts to concepts covered in other STEM disciplines, and their general interest in the course content. These types of affective learning outcomes cannot be measured in a content-based concept inventory, but are certainly connected to long-term classroom success (Middlecamp *et al.*, 2006). The SALG was administered in this study in a pre-post format in which students estimated their existing knowledge state and affective engagement in the course content prior to the beginning of the course, and then reported on their final learning outcomes at the end of the term.

The SALG was administered to the entire concept map treatment and journal entry control groups, and the survey respondents represent a sub-population of each group due to compliance issues related to survey completion (post-SALG respondents; $n = 49$ for the treatment group; $n = 30$ for the control group). The pre- and post-survey responses were paired, providing an opportunity to track student perceptions of gains in self-reported affective outcomes. Unfortunately, compliance issues further limited the number of students who completed both the pre- and post-survey instruments (pre/post paired respondents: $n = 36$ for treatment; $n = 18$ for control). The average Likert scale responses for all of the pre- and post-SALG questions among both study groups are summarized in Appendix 10 (ESI†), and the distributions of Likert scale responses on the nine post-SALG survey questions are illustrated in Fig. 4 (questions 3–6) and Appendix 11 (ESI†) (questions 1–2, 7–9). These descriptive results indicate the concept map treatment group had higher proportions of positive responses (5 = strongly agree; 4 = agree) than the control group on all but one of the post-SALG survey questions. However, any interpretation of these survey responses must be tempered given the fact the response rate on the post-SALG survey was quite low and the pre-SALG baseline responses could not be matched for a large number of these respondents. In an effort to compare how many students from each study group made gains on the various survey questions, the percentage of students who made gains was plotted against the number of questions for which an improvement was reported (see Appendix 12, ESI†). This analysis indicates that 78% of the concept map treatment students reported improvement in the pre- to post-SALG responses for five or more of the nine questions, whereas 72% of the journal control group students reported such gains. Because of the limited number of students who responded to both the pre- and post-SALG survey these results do not provide unambiguous evidence as to whether the concept map treatment led to greater impact on student perceptions of conceptual thinking or affective outcomes relative to the journal control group.

Though the SALG survey data did not explicitly reveal the concept map students were impacted in their affective learning outcomes, using this type of data to supplement more quantitative exam or concept inventory data could provide useful insights in future studies. Students who feel they are better prepared may have additional confidence, which in turn helps them overcome test anxiety and perform better on high stakes assessments (Hackathorn *et al.*, 2012). One could take the view that even if a concept map treatment were to yield equivalent exam scores compared to traditional assignments, if it resulted in gains related to measures of self-confidence, self-efficacy, and the ability to connect chemistry to real world ideas it is likely to positively impact interest and student success (Middlecamp *et al.*, 2006; Lindstrom and Middlecamp, 2017).

Limitations of the study

Quasi-experimental studies conducted within an entire academic term are often limited due to history and maturation

¶ SALG survey: https://salgsite.net/about.

Table 7 SALG survey questions that were included in the quasi-experimental analysis

Survey questions
Pre-SALG:[a] **presently, I understand**. . .
Post-SALG:[b] **as a result of your work in this class, what gains did you make in your understanding of**. . .
1. Chemical equilibrium.
2. Acid base chemistry.
3. How ideas we will explore in this class relate to ideas I have encountered in other classes within this subject area.
4. How ideas we will explore in this class relate to ideas I have encountered in classes outside of this subject area.
5. How studying this subject helps people address real world issues.
Pre-SALG:[a] **presently, I am in the habit of**. . .
Post-SALG:[b] **as a result of your work in this class, what gains did you make in the following skills**. . .
6. Connecting key ideas I learn in my classes with other knowledge.
7. Applying what I learn in classes to other situations.
8. Using systematic reasoning in my approach to problems.
Pre-SALG:[a] **presently, I**. . .
Post-SALG:[b] **as a result of your work in this class, what gains did you make in the following skills**. . .
9. Feel(ing) comfortable working with complex ideas.

[a] Likert scale: 1 = not at all, 2 = just a little, 3 = somewhat, 4 = a lot, 5 = a great deal. [b] Likert scale: 1 = no gains, 2 = a little gain, 3 = moderate gain, 4 = good gain, 5 = great gain.

effects within the study groups, and by the non-random distribution of background traits among the participants in the study (Mack *et al.*, 2019a). In the current study, the history and maturation internal validity threats were minimized by the use of treatment and control groups within the same class within the same academic term. The validity threat associated with the non-random distribution of incoming academic preparation traits was also reduced by the multiple linear regression model that was used to compare the treatment and control group concept inventory post-test scores. Perhaps the most obvious limitation of this study was the fact the treatment and control groups were taught by two different graduate TAs, therefore instructor effects could not be controlled. Despite possible biases in instructional effectiveness that might have been inherent between the two study groups, the two TAs did have similar experience teaching the discussion group sections and had collaborated in preparing classroom activities for previous offerings of the discussion group sections. These factors likely minimized the potential differences in student learning outcomes usually linked to instructor effects. Finally, it is noted the fact this was a single-institution study might lead to potential external validity threats. This study was conducted on an extremely diverse campus within a large enrollment course, suggesting the success of implementing the concept map intervention should translate to many higher education settings. With that said, future studies should ultimately aim to replicate the experimental implementation across a diverse set of institutions.

The comparison of means using the independent samples *t*-test, and the Pearson's correlation of concept inventory scores and concept map rubric scores appear to suggest the two research hypotheses should not be rejected. However, when multiple linear regression analyses were used to control for the participants' incoming academic preparation, the difference in concept inventory post-test scores between the study groups was no longer significant, nor was the relationship between concept mapping skills and performance on the concept inventory. Arriving at a definitive judgement regarding the impact of the concept map intervention is further complicated by the *post hoc* power analyses that estimated the power to detect a difference in the regression coefficients. These results suggest there is a chance the research hypotheses are being erroneously rejected based on the multiple linear regression analyses. Future studies should therefore attempt to more effectively isolate the impact of the concept map intervention and/or increase the sample size in order to increase the effect size of the treatment. This would aid in arriving at a more decisive conclusion regarding the impact of the concept map intervention on student conceptual learning gains.

Even if there is some correlation between improved concept mapping and conceptual understanding, this finding would be limited due to the fact it could not be determined if the ability to create well-developed concept maps resulted from the concept map intervention itself, or if students had a pre-existing ability to create high level concept maps upon entering the study. Because concept mapping skills were not measured for incoming students it is simply not possible to reconcile this question in the current study. The data seem to suggest there is some chance improved concept mapping leads to gains in conceptual understanding, therefore finding a way to help students become more proficient in developing higher quality concept maps appears to be a worthwhile instructional goal. Future studies could attempt to measure incoming concept mapping skills, though this would still pose problems from an experimental standpoint. Initial skills in the concept mapping of a pre-existing knowledge state may not translate to concept mapping skills related to the learning objectives covered in the study, and the preliminary concept mapping evaluation itself could impact the performance on the final concept map created by students. Nevertheless, creating an experimental design that addresses this limitation should be a priority.

Though the SALG instrument was used in an attempt to gain insight about student perceptions of affective learning outcomes, compliance issues resulted in a limited subset of study participants who provided feedback. Caution is therefore warranted in interpreting any general trends in the student responses to the SALG survey provided herein. Future studies should aim to increase the participant response rate, and might consider

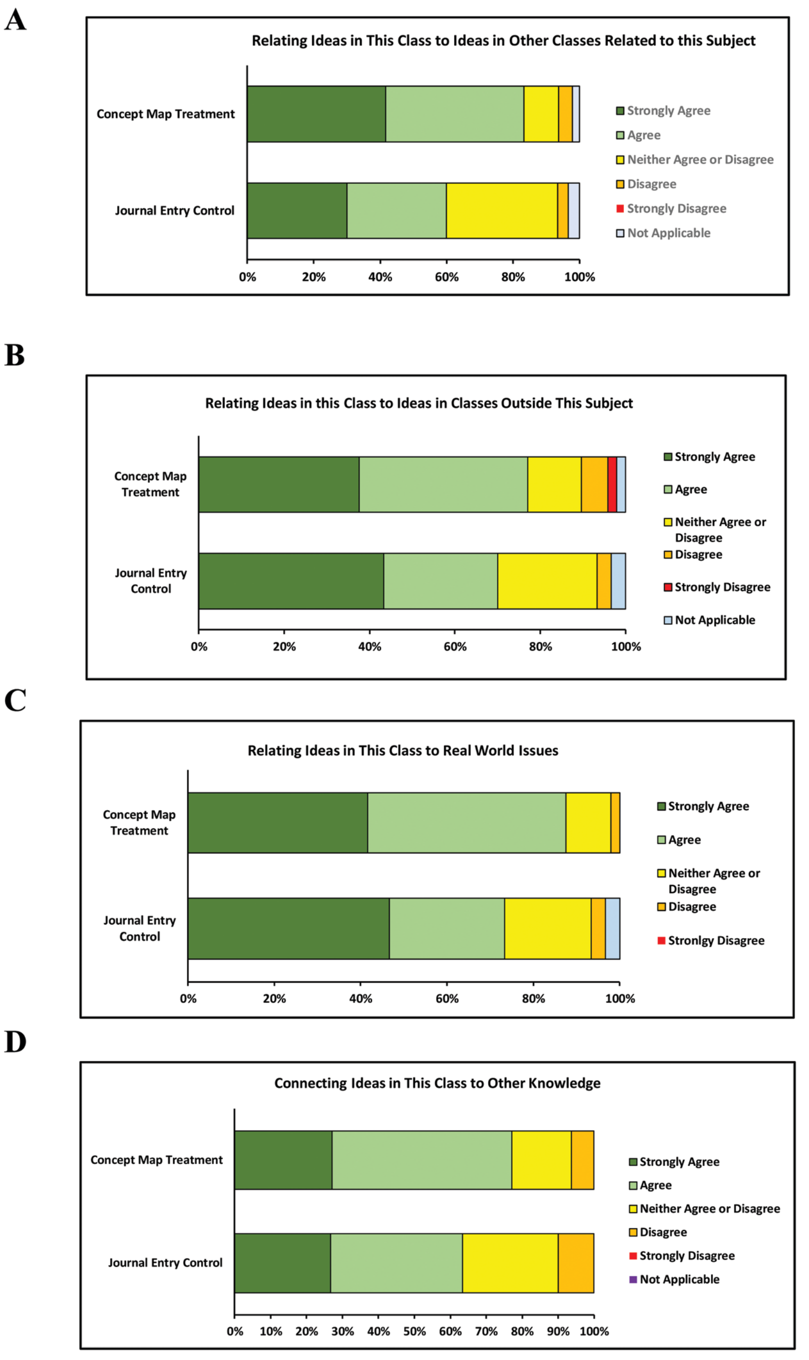


Fig. 4 Post-SALG responses to SALG questions: (A) #3; (B) #4; (C) #5; (D) #6. Post-SALG sample sizes: treatment = 49, control = 30.

including focus group interviews to gather additional data on students' perceptions of affective learning outcomes that could help corroborate data collected in the SALG.

Conclusion

In summary, the purpose of this study was to create a streamlined concept map implementation and generate new evidence about the impact of concept map assignments on large enrollment general chemistry courses. Though no unambiguous conclusion can be made regarding the impact of the concept map treatment on conceptual understanding under these quasi-experimental conditions, the results presented here suggest the ability to create well-developed concept maps might correlate to improved learning gains in conceptual understanding. Future research should therefore aim to ascertain if there is indeed a causal relationship between conceptual learning and creating well-developed concept maps and how students can be coached into creating more advanced concept maps.

It is noted the concept map assignment was structured in a way that would allow easy integration into a traditional large enrollment course. More specifically, the goal was to balance making the concept map assignment significant enough to impact student performance without making the administration and grading of concept maps a burden in terms of TA/instructor workload. Additionally, it was desired to avoid creating student resistance to the concept map assignments, which can arise if the subjective grading rubric is used to assign grades for the concept maps. The concept map implementation described in this study can act as a template for instructors who wish adopt this type of learning intervention and assessment, however future implementations might be re-designed to improve student engagement and comfort using concept maps while maintaining the ease of implementation. Increasing student engagement with concept mapping without relying on a subjective grading rubric could be accomplished by coupling the concept map assignment with a close-ended task that assesses students' ability to properly connect concepts. For instance, linking the concept map assignments described here with the Measure of Linked Concepts (MLC's) described by Lewis and coworkers (Ye *et al.*, 2015) or the more recently reported Creative Exercises (CE's) designed by Ye and coworkers (Gilewski *et al.*, 2019) would provide a less subjective method of evaluating the students. This would likely reduce student anxiety about being judged on their concept maps while simultaneously increasing the incentive for students to take the concept map assignment seriously. Including the MLC's or CE's in a quasi-experimental research design might also provide a means to determine if concept mapping skills can be taught to selected student populations and if true gains in concept mapping lead to improved conceptual understanding. Regardless of how instructors might adapt the use of concept maps into their course, making this tool a more prominent component of the instructional tool box can help students attain the type of meaningful learning described in Ausubel's assimilation theory of learning.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors gratefully acknowledge the assistant editor and anonymous referees for providing extensive feedback during the review process. Their suggestions and advice greatly improved the overall quality of the final manuscript.

Notes and references

Ausubel D. P., (1968), *Educational psychology: A cognitive view*, New York: Holt, Rinehart, & Winston.

Besterfield-Sacre M., Gerchak J., Lyons M. R., Shuman L. J. and Wolfe H., (2013), Scoring Concept Maps: An Integrated Rubric for Assessing Engineering Education, *J. Eng. Educ.*, **93**(2), 105–115.

Burdo J. and O'Dwyer L., (2015), The effectiveness of concept mapping and retrieval practice as learning strategies in an undergraduate physiology course, *Adv. Physiol. Educ.*, **39**, 335–340.

Burrows N. L. and Mooring S. R., (2015), Using concept mapping to uncover students' knowledge structures of chemical bonding concepts, *Chem. Educ. Res. Pract.*, **16**(1), 53–66.

Cañas A. J., Carff R., Hill G., Carvalho M., Arguedas M., Eskridge T. C., Carvajal R., (2005), Concept Maps: Integrating Knowledge and Information Visualization BT, in Tergan S.-O. and Keller T. (ed.), *Knowledge and Information Visualization: Searching for Synergies*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 205–219.

Chevron M.-P., (2014), A metacognitive tool: theoretical and operational analysis of skills exercised in structured concept maps, *Perspect. Sci.*, **2**(1), 46–54.

Cohen J., (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn, New York, NY: Lawrence Erlbaum Associates, ch. 9.

Cook E., Kennedy E. and McGuire S. Y., (2013), Effect of Teaching Metacognitive Learning Strategies on Performance in General Chemistry Courses, *J. Chem. Educ.*, **90**(8), 961–967.

Cooper M. M., Corley L. M. and Underwood S. M., (2013), An investigation of college chemistry students' understanding of structure–property relationships, *J. Res. Sci. Teach.*, **50**(6), 699–721.

Cros D.; Maurin M.; Amouroux R.; Chastrette, M.; Leber, J.; Fayol, M., (1986), Conceptions of First-year University Students of the Constituents of Matter and the Notions of Acids and Bases, *Eur. J. Sci. Educ.*, **8**(3), 305–313.

Eichler J. F. and Peeples J., (2016), Flipped classroom modules for large enrollment general chemistry courses: a low barrier approach to increase active learning and improve student grades, *Chem. Educ. Res. Pract.*, **17**(1), 197–208.

Francisco J. S., Nakhleh M. B., Nurrenbern S. C. and Miller M. L., (2002), Assessing Student Understanding of General Chemistry with Concept Mapping, *J. Chem. Educ.*, **79**(2), 248.

Gabel D., (1999), Improving Teaching and Learning through Chemistry Education Research: A Look to the Future, *J. Chem. Educ.*, **76**(4), 548.

Galloway K. R. and Bretz S. L., (2015), Measuring Meaningful Learning in the Undergraduate General Chemistry and Organic Chemistry Laboratories: A Longitudinal Study, *J. Chem. Educ.*, **92**(12), 2019–2030.

Geiger S. and Santelices M. V., (2007), Validity of High School Grades in Predicting Student Success Beyond the Freshman Year: High School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes, University of California, Berkeley Center for Studies in Higher Education, https://escholarship.org/uc/item/7306z0zf.

Gilewski A., Mallory E., Sandoval M., Litvak M., Ye L. (2019), Does linking help? Effects and student perceptions of a learner-centered assessment implemented in introductory chemistry, *Chem. Educ. Res. Pract.*, **20**, 399–411.

Hackathorn J., Cornell K., Garczynski A., Solomon E., Blankmeyer K. and Tennial R., (2012), Examining exam reviews: a comparison of exam scores and attitudes, *J. Scholarship Teach. Learn.*, **12**(3), 78–87.

Harrison A. G. and Treagust D. F., (2018), Secondary students' mental models of atoms and molecules: implications for teaching chemistry, *Sci. Educ.*, **80**(5), 509–534.

Johnstone A. H., (1993), The development of chemistry teaching: a changing response to changing demand, *J. Chem. Educ.*, **70**(9), 701.

Kennedy S. A., (2016), Design of a Dynamic Undergraduate Green Chemistry Course, *J. Chem. Educ.*, **93**(4), 645–649.

Leech N. L., Gliner J. A., Morgan G. A., Harmon R. J. and Harmon R. J., (2003), Use and Interpretation of Multiple Regression, *J. Am. Acad. Child Adolesc. Psychiatry*, **42**(6), 738–740.

Lindstrom T.; Middlecamp C., (2017), Campus as a Living Laboratory for Sustainability: The Chemistry Connection, *J. Chem. Educ.*, **94** (8), 1036–1042.

Luxford C. J. and Bretz S. L., (2014), Development of the Bonding Representations Inventory To Identify Student Misconceptions about Covalent and Ionic Bonding Representations, *J. Chem. Educ.*, **91**(3), 312–320.

Mack M. R., Hensen C. and Barbera J., (2019a), Metrics and Methods Used To Compare Student Performance Data in Chemistry Education Research Articles, *J. Chem. Educ.*, **96**, 401–413.

Mack M.R., Stanich C.A., Goldman L.M., (2019b), Math Self-Beliefs Relate to Achievement in Introductory Chemistry Courses. Chapter in *It's Just Math: Research on Students' Understanding of Chemistry and Mathematics*, *ACS Symp. Ser.*, **1316**, 81–104.

Markow P. G. and Lonning R. A., (1998), Usefulness of concept maps in college chemistry laboratories: Students' perceptions and effects on achievement, *J. Res. Sci. Teach.*, **35**(9), 1015–1029.

Middlecamp C. H.; Jordan T.; Shachter A. M.; Kashmanian Oates K.; Lottridge S., (2006), Chemistry, Society, and Civic Engagement (Part 1): The SENCER Project, *J. Chem. Educ.*, **83** (9), 1301.

Mulford D. R. and Robinson W. R., (2002), An Inventory for Alternate Conceptions among First-Semester General Chemistry Students, *J. Chem. Educ.*, **79**(6), 739 , DOI: 10.1021/ed079p739.

Nesbit J. C. and Adesope O. O., (2006), Learning with concept and knowledge maps: a meta-analysis, *Rev. Educ. Res.*, **76**, 413–448.

Nicoll G., Francisco J. S. and Nakhleh M., (2001), An Investigation of the Value of Using Concept Maps in General Chemistry, *J. Chem. Educ.*, **78**(8), 1111.

Novak J. D., (1984), Application of advances in learning theory and philosophy of science to the improvement of chemistry teaching, *J. Chem. Educ.*, **61**(7), 607.

Novak J. D., (1990), Concept maps and Vee diagrams: two metacognitive tools to facilitate meaningful learning, *Instr. Sci.*, **19**(1), 29–52.

Novak J. D. and Cañas A. J., (2008), The Theory Underlying Concept Maps and How to Construct and Use Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition.

Novak J. D.; Gowin B. B., (1984), *Learning How to Learn*, Cambridge: Cambridge University Press.

Rickey D. and Stacy A. M., (2000), The Role of Metacognition in Learning Chemistry, *J. Chem. Educ.*, **77**(7), 915.

Regis A., Albertazzi P. G. and Roletto E., (1996), Concept Maps in Chemistry Education, *J. Chem. Educ.*, **73**(11), 1084.

Turan-Oluk N. and Ekmekci G., (2018), The effect of concept maps, as an individual learning tool, on the success of learning the concepts related to gravimetric analysis, *Chem. Educ. Res. Pract.*, **19**(3), 819–833.

Vincent-Ruz P., Binning K., Schunn C.D., Grabowski J., (2018), The Effect of Math SAT on Women's Chemistry Competency Beliefs, *Chem. Educ. Res. Pract.*, **19**, 342–351.

Widhiarso W. and Ravand H., (2014), Estimating reliability coefficient for multidimensional measures: a pedagogical illustration, *Rev. Psychol.*, **21**(2), 111–121.

Ye L., Oueini R. and Lewis S. E., (2015), Developing and Implementing an Assessment Technique To Measure Linked Concepts, *J. Chem. Educ.*, **92**(11), 1807–1812.

Zwick R. and Sklar J.C., (2005), Predicting College Grades and Degree Completion Using High School Grades and SAT Scores: The Role of Student Ethnicity and First Language, *Am. Educ. Res. J.*, **42**, 439–464.

JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Article

Efficacy of an Asynchronous Online Preparatory Chemistry Course: An Observational Study

Grace Henbest, Kiana Mortezaei, Teresa Alvelais, Courtney Murphy, and Jack F. Eichler*

Cite This: J. Chem. Educ. 2020, 97, 4287–4296 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In an ongoing effort to increase student retention and success in the undergraduate general chemistry course sequence, a fully online preparatory chemistry course was developed and implemented at a large public research university. To gain insight about the efficacy of the online course, an observational study was carried out in which student performance on final exams and performance in the subsequent general chemistry course were compared between the online cohort and a previous student cohort which completed the preparatory chemistry course in a traditional lecture format. Multiple linear regression analyses were used to compare final exam scores and general chemistry course grades between the online and in-person student cohorts, while statistically controlling for incoming student academic achievement. Results from these analyses suggest the fully online course resulted in statistically significant increases in both the preparatory chemistry final exam scores and course grades in the subsequent general chemistry course. Because the retention of less academically prepared students in STEM majors is a historical problem at the institution in which the online preparatory chemistry course was implemented, the analyses also aimed to determine if this at-risk group demonstrated similar achievement relative to the population at large. Notably, it was determined that students with the lowest incoming Math SAT scores appeared to perform better in the online course relative to the analogous group of students in the in-person course. Though the observational nature of this study does not allow for the determination of causality, these results suggest a fully online course can result in improved performance for large populations of students, without resulting in a negative achievement gap for less academically prepared students. The structure and implementation of the online course and the results from the statistical analyses will be described herein.

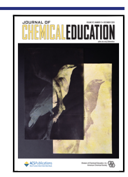
KEYWORDS: *First-Year Undergraduate/General, Computer-Based Learning, Curriculum, Internet/Web-Based Learning, Enrichment/Review Materials*

■ INTRODUCTION

In response to the persistent need to improve retention and interest in STEM disciplines, there is an ongoing effort at undergraduate institutions to ensure students are adequately prepared for college-level science and math.[1] Because the year-long general chemistry sequence is a common barrier for student success in STEM fields,[2] preparatory courses or summer bridge programs designed to teach the fundamental skills required for general chemistry are common in undergraduate chemistry programs.[3] At the University of California—Riverside (UCR), all students in the College of Natural and Agricultural Sciences (CNAS) are required to take general chemistry, and the use of general chemistry as a prerequisite for other science courses is also common across the University of California (UC) system. The increasing need to ensure incoming first-year students are adequately prepared for this course is partly evidenced by the fact that, within the UC system, six out of the nine undergraduate campuses offer an in-person preparatory chemistry course.[4] At UCR, the preparatory chemistry course has been in the university course catalog for more than a decade but has been offered intermittently depending on personnel and resource availability.

Justification for offering preparatory courses or summer bridge programs is rooted in the chemical education literature, which suggests that students who participate in these types of programs perform better, on average, in subsequent chemistry courses than students who do not receive this additional preparation.[1,5,6] Eitemüller and Habig report that a two week in-person bridge course led to short-term improvements in content knowledge for students with low prior chemistry knowledge, whereas students with high prior content knowledge appeared to make longer-term performance gains in the subsequent general chemistry course.[5] Botch and co-workers report that an optional 20 h online bridge course helped self-selected high achieving students achieve improved performance in general chemistry relative to nonparticipants; however, it is noted less-prepared students did not opt into the program at significant levels.[1] Though the positive impact to performance in general chemistry for specific types of students appears to be clearly evidenced, the benefits of these preparatory experiences are counterbalanced by the fact they require additional departmental resources, and academic-year prep

Received: March 30, 2020
Revised: October 29, 2020
Published: November 24, 2020



courses can lead to "off sequence" populations of students who are less likely to persist in their STEM major if these courses are offered in the fall term.[7]

Due to the constraints in resources for our traditional in-person preparatory chemistry course and a desire to create a course that can be offered both in the summer (as a bridge-type of experience for incoming first-year students) and during the regular academic term, efforts were initiated to create an online version of our traditional in-person preparatory chemistry course. The use of online instruction is often viewed by faculty with a skeptical eye, especially among older and more experienced instructors.[8] However, there is a growing body of evidence suggesting online instruction, when implemented with full-time residential college students enrolled in four-year degree programs, performs as well and often better than equivalent traditional instruction.[9,10] The impact of online instruction is likely mediated by the way in which it is structured, evidenced by the meta-analysis conducted by Means and co-workers.[11] This study, in which effect sizes were extracted from a pool of 50 courses that included online instruction in a variety of disciplines and instructional levels, found that online learning performed better on average than equivalent in-person learning, but that advantage was significant only for online learning delivered in a blended learning environment. Conversely, the improved performance from purely online courses was not found to be significantly greater than in-person instruction.

The body of research looking at the efficacy of online instruction in undergraduate chemistry is less well developed, though the recent report by Faulconer and co-workers suggests an online implementation of a general chemistry course including both lecture and lab compares favorably to the equivalent in-person course.[12] Regarding online preparatory chemistry courses, the findings from Botch and co-workers described above provide some evidence that an online bridge course can positively impact student performance in general chemisry.[1] Additionally, the study reported by Docktor and co-workers suggests an online summer bridge intervention positively impacted student performance in the subsequent fall-term general chemistry course. More specifically, it was found that students who completed self-paced modules within the adaptive−responsive ALEKS learning system performed on par with students who placed directly into the general chemistry course and performed significantly better on the general chemistry course final exam compared to students who took the traditional in-person preparatory chemistry course.[3]

The positive outcomes in online learning environments described above can be framed in Mayer's cognitive theory of multimedia learning.[13] The cognitive theory of multimedia learning purports that gaining new knowledge through online instruction can effectively take place if one receives information from two sensory platforms: auditory and visual. This information is then organized and works in conjunction with prior knowledge that is retrieved from one's long-term memory to foster meaningful learning as opposed to rote learning and helps overcome the limitations of one's working memory capacity. Mayer also outlines how the research on multimedia learning can inform instructional design, and lists the reduction of extraneous processing, managing essential processing, and fostering generative processing as guiding principles.[13] When properly designed, online instruction has the potential to avoid the cognitive overload that is generally associated with student learning in traditional in-person lecture environments. Despite the potential benefits of asynchronous online learning (asynchronous is defined as online learning in which students work individually at their own pace), there are potential gaps in the research that should be considered by both educational researchers and practitioners of online instruction. In particular, Mayer notes there may be boundary conditions for the impact of online instruction. Foundational research needs to be broadened to help determine how different types of multimedia instructional methods impact different types of learners (i.e., low versus high knowledge learners), how online instruction might be tailored for different types of learning objectives (i.e., conceptual learning versus algorithmic problem solving), and how online learning is able to affect transfer of knowledge rather than retention of knowledge.[13] Instructors should be cognizant of these boundary conditions when designing and implementing online courses. Additionally, the findings from Means and co-workers suggest the use of online learning in concert with other instructional interventions may ultimately explain the apparent advantage observed with online learning.[11] For instance, utilizing online learning within a blended learning structure can lead to increased interactions among learners, since the blended format still incorporates some in-person instruction in the course. Providing students these opportunities for sense-making and articulation of conceptual ideas with other students is generally more difficult to achieve in purely online asynchronous courses.

In the present report, the implementation of a fully online, asynchronous preparatory chemistry course will be described. To our knowledge, this is the first report in the chemistry education literature on the efficacy of a fully online, for-credit preparatory chemistry course offered in a regular academic term. The course described herein was initially offered during the fall term analogous to previous in-person offerings of the course. It is also noted the course was not designed and implemented within the framework of a research study in which specific instructional interventions were compared to a contemporaneous "teaching as usual" control. The goal was to structure the online learning environment in a way that was most likely to avoid logistical and technical roadblocks, while hopefully leading to similar student performance outcomes as previous offerings of the in-person version of the course. Because the corresponding author was the instructor of record for both the new online course and the previous in-person version of the course, it was possible to carry out an observational study in which performance outcomes were compared between the two student cohorts. The results of this analysis will be described and discussed within the context of the previous research that investigated the efficacy of preparatory chemistry courses and within the context of the cognitive theory of multimedia learning.

■ METHODS

Course Structure

The analysis described herein was conducted to determine the impact a new asynchronous preparatory chemistry (prep chem) course had on student performance, both on the final exam score in the prep chem course and on the final course grade in the subsequent general chemistry course (CHEM 001A). The online prep chem course was taught in the fall of 2018 (this student cohort will be designated "Online") and was compared to a student cohort who took the traditional in-

https://dx.doi.org/10.1021/acs.jchemed.0c00294
J. Chem. Educ. 2020, 97, 4287−4296

Table 1. Online Modules for Preparatory Chemistry Course and Lecture Schedule for In-Person Course

Unit for Online Course	Dates for Online Course	Topics	Activities	Lecture Schedule for In-Person Course[a]
1	Oct. 1–14	Matter, measurements, and dimensional analysis	1. Online lectures/iLearn quiz	Oct. 2, 4
2	Oct. 15–28	Atomic structure	1. Online lectures/iLearn quiz	Oct. 9, 11
3	Oct. 29–Nov. 11	Electronic structure	1. Online lectures/iLearn quiz 2. Online iLearn midterm exam	Oct. 16, 18
Midterm exam	Due Nov. 16	Units 1–3		Oct. 25, Nov. 15
4	Nov. 12–25	Chemical reactions	1. Online lectures/iLearn questions	Oct. 30, Nov. 1, 6, 8, 13
5	Nov. 25–Dec. 10	Solutions and miscellaneous math topics	1. Online lectures/iLearn questions	Nov. 27, 29
Final exam	Due Dec. 13	Units 1–5		Dec. 14

[a]Each date for the In-person course corresponds to a 50 min lecture.

person prep chem course the previous fall (this cohort/class will be designated "In-person"). The In-person course met twice a week for 50 min and included weekly 120 min recitation sessions facilitated by graduate student teaching assistants (TAs). The assignments in the traditional lecture consisted of two midterms, a final exam, and weekly online quizzes. The final exam for the In-person course was a 3 h exam that consisted of 40 questions worth a total of 400 points, and no additional graded homework was assigned aside from the weekly online quizzes in the course management system. Students were given extra credit for attending the weekly discussion group sections, in which group problem solving worksheets were completed under the guidance of a graduate student TA, but these activities were not graded or collected. Overall, the course was graded on a pass–fail basis in which students were awarded three units of free elective course credit. The schedule of topics and full description of the course are provided in the attached Supporting Information.

In lieu of attending the in-person lectures, students in the Online course were directed to complete online video tutorials in a self-paced manner and complete the associated online quizzes within the course management system, and the in-person discussion group sections were eliminated from the Online course. Though the individual modules were self-paced, these online activities had deadlines that were scheduled approximately every 2 weeks to ensure students did not attempt to complete all of the course content in a condensed time frame at the end of the term. The weekly quizzes from the In-person course were adapted for use in the Online course and quizzes within each unit were combined into one quiz that was due at the end of each module (see Table 1 for the schedule of topics and module deadlines). The online video tutorials were produced using a Learning Glass system that is capable of overlaying other graphics and presentation slides within the video.[14] A screenshot of a sample Learning Glass video is shown in Figure 1.

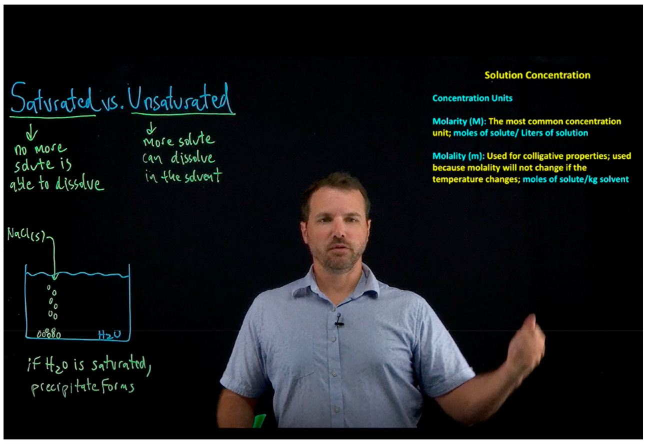


Figure 1. Screenshot of learning glass video used in the online prep chem course (ONLINE). All video tutorials were produced using the Learning Glass system.[14]

In addition to the biweekly quizzes, the Online course had one midterm and a final exam. The midterm exam for the Online course was not completely analogous to the In-person midterms since it covered a different set of topics; however, midterms from both the In-person and Online courses consisted of multiple-choice questions that assessed computational skills and basic conceptual knowledge retention. In order to mitigate unapproved collaboration among students on the exams, the ProctorU online exam proctoring service was used for the midterm and final exams.[15] Students could arrange to take the exams at home or in an on-campus study room, but the proctoring service verified students identities and monitored the students' work flow through the computer web cam and prevented students from opening up any programs aside from the online test interface in the course management system. The campus library at UCR provides free access to laptop computers, which ensured students had access to computers that were compatible with the ProctorU service. The proctoring service did have an associated cost that increased on the basis of the length of the exam; therefore, the final exam for the Online course was scheduled as a 2 h exam that consisted of 30 multiple-choice questions, worth a total of 400 points. Though the Online final exam had fewer questions, this exam was created using a subset of questions from the In-person final (the In-person final was not distributed to students; therefore, students in the Online course did not have access to the prior exam). The In-person final exam allowed 4.5 min per question and included 16 calculation-based questions and 24 conceptual questions. Conversely, the Online exam allowed 4 min per question and included 14 calculation-based questions and 16 conceptual questions. Item analysis data could not be retrieved for the fall 2017 In-person course; however, the these data were available for the Online course through the online course management system (see Supporting Information Appendix S1A,B for the Online and In-person final exams, and Supporting Information Appendix S1C for the Online final exam item analysis). The item analysis data indicate there is a good distribution of question difficulty and the majority of the exam items appeared to correlate to overall

https://dx.doi.org/10.1021/acs.jchemed.0c00294
J. Chem. Educ. 2020, 97, 4287–4296

exam performance (eight test items had means < 0.70, 17 test items had means between 0.70 and 0.90, and five test items had means between 0.90 and 1.0; 26 of the 30 exam items had item discrimination indexes greater than +0.20). It is also noted that though item analysis data could not be compared between the two final exams, the fact the Online final exam allowed less time per question and included a higher ratio of calculation-based questions suggests the Online was not inherently less difficult than the In-person exam.

Analogous to the In-person course, the Online prep chem course was also graded on a pass−fail basis but was worth only two units of free elective course credit since it did not include the additional weekly discussion group sessions. The full course description is provided in the attached Supporting Information. It is noted that both the In-person and Online courses used the freely available OpenStax textbook,[16] and both courses were focused on skill development rather than higher order reasoning skills. In particular, considerable time was devoted to strengthening skills related to unit conversions, dimensional analysis, and the introduction to lower order learning objectives associated with atomic theory and electronic structure.

In both the In-person and Online courses, the prep chem course was part of the CNAS Scholars Learning Communities (LCs). Students who joined the LC and were eligible to take the prep chem course would register for the class as part of their LC block of classes. Approximately 50% of incoming CNAS first-year students who did not place directly into the regular general chemistry course were enrolled in the prep chem course through the LC in both the In-person and Online courses, and over 90% of students in both cohorts were enrolled in a first-year learning community. The students who were placed into the prep chem course but did not take the course in the fall term could not be tracked for this analysis.

The analysis described below includes tracking the final student grades in the subsequent winter-term CHEM 001A courses (winter-term 2018 = "In-person W18"; winter-term 2019 = "Online W19"). Unfortunately, it was not possible to track the matriculation of the In-person and Online student cohorts into the specific CHEM 001A sections. There were three CHEM 001A sections of approximately 280−300 students in both the W18 and W19 terms, with three instructors teaching the three sections in W18 and two instructors teaching the three sections in W19 (one instructor taught two sections in W19; one instructor who taught one section of CHEM 001A in W18 and also taught one section W19). Though there was some inherent instructor variability across the W18 and W19 cohorts, all sections were taught using the same textbook, all instructors used the same online homework system, and all instructors taught using predominantly traditional lecture techniques (i.e., there was no significant use of collaborative group learning or active learning). Three of the four instructors from W18 and W19 (Instructors A−C) based their final grades on the online homework, two midterm exams, and a comprehensive final exam (all midterm exams and final exams were multiple choice, but did vary between instructors); the fourth instructor (Instructor D from W19) assigned the online homework as optional practice and assigned slightly higher weights to the midterm exams and final exam. The percentages of students who did not achieve a grade of C− or higher ranged from 10 to 20% for all sections across the W18 and W19 terms (the success rates were not reported for each instructor, but the department reported these data for all sections in an anonymous fashion). The descriptions for how each CHEM 001A instructor assigned grades are provided in the Supporting Information Appendix S2A,B.

Statistical Analyses

The observational comparison of student performance outcomes between the In-person and Online cohorts was approved by the UCR Institutional Review Board (IRB) as an exempt study under protocol HS-19-296. An analysis of variance (ANOVA) was first used to determine if there were statistically significant differences in the quantitative variables between the two study groups (prep chem final exam scores, CHEM 001A course grade, high school GPA, and Math SAT scores). Because the online course was not implemented within a typical experimental or quasi-experimental research design, multiple regression analyses were used to model the impact of the online course on performance outcomes relative to the previous In-person course, while statistically controlling for incoming academic preparation and student demographic identity. Student populations from the Online course were considered the "treatment" group, whereas populations from the In-person course were considered the "control" group. These independent variables were dummy coded, in which 1 = "online course participant" and 0 = "in-person participant," allowing for the determination of whether the online course significantly related to the prep chem final exam or CHEM 001A course grade dependent variables. Linear regression models that included ethnicity and gender as independent variables were also dummy coded (1 = female, 0 = male; 1 = specific racial identity; 0 = other). As described by Murnane and Willett, a regression equation was used to measure the impact of participating in the online prep chem course.[17] This model accounts for the various independent variables, and calculates their effect on the dependent variable (see eq 1).

$$y = a + b_1x_1 + b_2x_2 + ... + b_kx_k + \varepsilon \qquad (1)$$

All ANOVA and multiple linear regression analyses were carried out using the SPSS Statistics 24 software package.[18] For the analyses in which the prep chem final exam score was included as the dependent variable, the scores for the Online and In-person cohorts were reported as a score out of 100 points (i.e., exam score percentage). For the analyses in which the CHEM 001A course grade was included as the dependent variable, the course grade was reported as a numerical GPA quality point (on a scale from 0 to 4; this scale includes different point values for ± letter grades; see Supporting Information Appendix S2C for summary of letter grade-quality point conversions). It is noted the quality point GPA scale is more ordinal in nature (there are 11 discrete quality point values possible on the 0−4 scale); however, the robustness of linear regression and other parametric tests in adequately handling ordinal data has been previously described.[19] Because the quality point scale had 11 possible outcomes and the GPA quality point dependent variable approximated a normal distribution in the study populations (see Supporting Information Appendix S3B,C), this variable was treated analogously to a quantitative discrete variable in the multiple linear regression analyses.

The retention and four-year graduation rates of students in the UCR College of Natural and Agricultural Sciences continues to lag behind those observed in the College of Humanities and Social Sciences; therefore, the multiple linear regression analyses included a treatment group−math SAT interaction term as described by Aiken and West.[20] This was included to determine if student academic preparation, as measured by math SAT scores, moderated the impact of the class treatment. To gain insight about how math SAT scores moderated class treatment, scatter plots were created in which classroom performance (prep chem final exam score or CHEM 001A course grade) was plotted against classroom treatment at three levels of math SAT performance. The three tertiles of performance were labeled "High SAT," "Medium SAT," and "Low SAT". The math SAT scores were divided into tertiles to ensure the sample size of each grouping was similar, and it is noted the math SAT scores were not skewed toward the high end of the score range (low math SAT score range = 360−540, medium math SAT score range = 550−590, high math SAT score range = 600−790). Finally, the statistical power and effect sizes for the multiple linear regression analyses were calculated as described by Cohen.[21]

All analyses were performed under the working null hypothesis that states "there is no relationship between participation in the different type of prep chem course and the performance-based dependent variable" (i.e., prep chem final exam score or CHEM 001A course grade). The analyses described above were carried out to address the following research questions:

1. Does converting a traditional in-person preparatory chemistry course into an asynchronous fully online course negatively impact student performance in the prep chem course?
2. Does converting a traditional in-person preparatory chemistry course into an asynchronous fully online course negatively impact student performance in the subsequent general chemistry course?
3. Does converting a traditional in-person preparatory chemistry course into an asynchronous fully online course negatively impact academically less prepared students?

■ RESULTS

Overall Performance of Online Course

The descriptive statistics for the In-person and Online student cohorts are summarized in Table 2. The gender and demographic makeup of the two cohorts were similar and were representative of the campus-wide undergraduate population.[22] The Online prep chem cohort appeared to have higher final exam scores in the prep chem course and higher grades in CHEM 001A than the In-person cohort, though it appeared the high school GPA and math SAT scores were also higher for the Online cohort. To determine if the differences in performance-based variables and quantitative variables related to incoming academic performance were statistically significant, preliminary ANOVA tests were used to compare the Online and In-person cohorts. These analyses confirmed there appears to be a significant difference between the two populations, specifically that the Online cohort had statistically higher high school GPAs, math SAT scores, prep chem final exam scores, and CHEM 001A course grades relative to the In-person cohort (see Tables S1 and S2).

Because the incoming academic preparation of the Online cohort appeared to be higher than the In-person cohort, high school GPA and math SAT scores were statistically controlled in the multiple linear regression models as described above.

Table 2. Descriptive Statistics Comparing Prep Chem Student Cohorts In-Person and Online Courses

Statistic	In-Person (n = 408)	Online (n = 463)
No. of students in first-year learning community (%)	396 (97)	435 (94)
Mean prep chem final exam score (out of 400 points)	276 ± 51	314 ± 51
Mean prep chem final exam (%)	69.0 ± 12.0	78.6 ± 11.4
Total no. of students who failed course (%)	23 (5.6)	21 (4.6)
Total no. of students who withdrew from course (%)	2 (0.49)	2 (0.43)
Mean general chemistry (CHEM 001A) grade[a]	2.58 ± 0.83	2.93 ± 0.63
Mean high school GPA	3.72 ± 0.45	3.80 ± 0.24
Mean math SAT score[b]	554 ± 57	572 ± 62
Male students, n (%)	134 (33)	151 (33)
Female students, n (%)	267 (65)	307 (66)
Sex of students not reported, n (%)	7 (2)	5 (1)
Asian students, n (%)	132 (32)	160 (35)
White students, n (%)	51 (13)	49 (10)
Hispanic or LatinX students, n (%)	191 (47)	204 (44)
Multiracial students, n (%)	21 (5)	30 (6.5)
Black or African American students, n (%)	8 (2)	13 (3)
Native Hawaiian/Pacific Islander students, n (%)	1 (0.2)	1 (0.2)
Unknown ethnicity, n (%)	3 (0.6)	4 (0.9)
Nonresident of U.S., n (%)	1 (0.2)	2 (0.4)

[a]GPA quality point on 0−4 scale. [b]SAT scores for the mathematics section have a range of 200−800.

The analyses comparing the performance of the entire class cohorts are summarized in Tables 3 and 4, respectively, and the tests for the various assumptions required to carry out multiple linear regression are summarized in the Supporting Information Appendix S3. The regression analysis in which classroom participation was included as the independent variable (coded 0 for In-person, 1 for Online) and prep chem final exam scores were included as the dependent variable indicates the Online cohort had significantly higher scores than the In-person cohort (Table 3). Because the class participation independent variables were dummy coded (1 = Online course; 0 = In-person course), the unstandardized B coefficient represents the change statistic for the dependent variable (i.e., the average change in the dependent variable for the Online population relative to the In-person population). Therefore, it was found students in the Online course performed, on average, 8.6 points higher on the prep chem final than students in the In-person course, while holding constant the independent variables related to incoming academic preparation. This relationship between class treatment group and prep chem final exam scores was significant to the p = 0.001 level (unstandardized B = 8.593, p < 0.001; see Table 3). The multiple linear regression also confirms previous findings that indicate high school GPA and math SAT scores are positively related to student performance on the prep chem final exam,[23,24] as these two independent variables had positive unstandardized B coefficients that were significant to the p = 0.05 level and p = 0.001 level, respectively (see Table 3; it is noted the unstandardized B coefficients for these continuous variables represent the unit increase in the dependent variable for every unit increase in the independent variable). Though the high school GPA and math SAT covariates were found to

Table 3. Multiple Linear Regression Analysis Determining the Impact of Course Participation on Prep Chem Final Exam Score for the Total Student Population

Independent Variable[a]	Unstandardized B	Standard Error	Standardized β	t	p	95% Confidence Interval Results for B	
						Lower Bound	Upper Bound
(Constant)	30.832	9.209		3.348	0.001	12.756	48.907
Treatment group[b] (1 = Online; 0 = In-person)	8.593	0.812	0.341	10.585	<0.001	7.000	10.186
High school GPA	3.555	1.668	0.070	2.131	0.033	0.281	6.829
Math SAT	0.046	0.010	0.221	4.464	<0.001	0.026	0.066
Treatment group *SAT	0.027	0.013	0.100	2.061	0.040	0.001	0.054

[a]Independent variable ("group") = class participation (1 = online, 0 = in-person); dependent variable = prep chem final exam score (final exam scores were reported as scores out 100 points); online prep chem course, n = 463; in-person course, n = 408. An analysis including gender and ethnic identity as independent variables was carried out. Because the explanatory power of the model was not improved by including these covariates, gender and ethnic identity were not included in the final analysis (see Supporting Information Table S3). [b]R = 0.405; R^2 = 0.164; adjusted R^2 = 0.160; standard error of the estimate = 11.54; f^2 effect size = 0.130. See the Supporting Information, Table S5 for effect size and estimated power calculations.

Table 4. Multiple Linear Regression Analysis Determining the Impact of Course Participation on General Chemistry[a] Course Final Grade for the Total Student Population

Independent Variable[b]	Unstandardized B	Standard Error	Standardized β	t	p	95% Confidence Interval Results for B	
						Lower Bound	Upper Bound
(Constant)	−0.961	0.569		−1.690	0.091	−2.001	0.238
Treatment group[c] (1 = Online; 0 = In-person)	0.267	0.050	0.177	5.319	<0.001	0.167	0.364
High school GPA	0.433	0.103	0.143	4.208	<0.001	0.272	0.680
Math SAT	0.003	0.001	0.280	5.463	<0.001	0.002	0.004
Treatment group *SAT	0.001	0.001	0.080	1.611	0.108	0.000	0.003

[a]General chemistry was taken in the subsequent quarter after the prep chem course. [b]Independent variable ("group") = class participation (1 = online, 0 = in-person); dependent variable = CHEM 001A course grade (expressed as 0−4 GPA quality point). An analysis including gender and ethnic identity as independent variables was carried out. Because the explanatory power of the model was not improved by including these covariates, gender and ethnic identity were not included in the final analysis (see Supporting Information Table S4). [c]Online prep chem course, n = 463; in-person course, n = 408. R = 0.325; R^2 = 0.106; adjusted R^2= 0.102; standard error of the estimate = 0.712; f^2 effect size = 0.0336. See the Supporting Information, Table S6 for effect size and estimated power calculations.

be significant predictors of student performance on the final exam, the standardized β coefficient for the class treatment group independent variable suggests that classroom participation was the strongest predictor of final exam performance in this model (see Table 3).

The regression analysis in which classroom participation was included as the independent variable, and CHEM 001A course grades were included as the dependent variable, indicates the Online cohort also had significantly higher general chemistry grades than the In-person cohort (Table 4). It was found students in the Online course earned course grades that were, on average, 0.27 GPA quality points higher than students in the In-person course, while holding constant the independent variables related to incoming academic preparation. The relationship between class treatment group and CHEM 001A course grade difference was significant to the p = 0.001 level (unstandardized B = 0.267, p < 0.001; see Table 4). High school GPA and math SAT also appeared to have a positive relationship with the CHEM 001A course grade, with both independent variables having positive unstandardized B coefficients that were significant to the p = 0.001 level (p < 0.001 for both variables; see Table 4). Though the improved general chemistry course grades for the Online cohort were statistically significant, the standardized β coefficients suggest that math SAT scores are the strongest predictor of CHEM 001A course grade.

The demographic breakdowns of the two study populations were representative of the entire undergraduate population at the institution in which the courses were offered. However, since it has been previously reported that gender and ethnicity can be negatively associated with persistence and success in chemistry/STEM majors,[25,26] it was desired to determine if gender or ethnic identity might be associated with lower performance in the Online prep chem course. Thus, linear regression models were also created in which gender and ethnic identity were included as independent variables to confirm if these student characteristics might be predictive of student performance (see Tables S3 and S4). These covariates did not appear to have a significant relationship to the prep chem final exam scores or CHEM 001A course grades, as none of these independent variables had unstandardized B coefficients that were significant at the p = 0.05 level. Additionally, the explanatory power of neither model improved significantly when the gender and ethnic identity variables were included (the change in adjusted R^2 increased by less than 1% and the change in the ANOVA F statistic upon adding the gender and ethnic identity covariates was not statistically significant; see Tables S3A and S4A). Because the explanatory power of the regression models was not improved, and because the overall statistical power of the models would be reduced upon adding these additional independent variables, the models summarized in Tables 3 and 4 were used in the final analysis.

The multiple linear regression analyses summarized in Tables 3 and 4 suggest the null hypothesis stating "there is no relationship between participation in the different type of

https://dx.doi.org/10.1021/acs.jchemed.0c00294
J. Chem. Educ. 2020, 97, 4287−4296

prep chem course and the performance-based dependent variable" can be rejected for both dependent variables. Despite the apparent significance of the improved performance for the online prep chem student cohort, it was prudent to determine the effect size and statistical power of these analyses. The f^2 effect size and statistical power for multiple regression analyses were calculated as previously described by Cohen (see Tables S5 and S6 for descriptions of these calculations, and Chapter 9 in *Statistical Power Analysis for the Behavioral Sciences*).[21] The f^2 effect size, which indexes the degree of departure from the null hypothesis (i.e., the degree of departure from no treatment effect), was found to be 0.130 and 0.0336, respectively for the prep chem final exam score and CHEM 001A course grade analyses. Cohen designates an f^2 effect size of 0.02 as small, 0.15 as medium, and 0.35 as large.[21] The statistical power, the probability the statistical test can lead one to reject the null hypothesis, was determined to be greater than 0.995 and approximately 0.91, respectively, for the prep chem final exam score and CHEM 001A course grade analyses (the generally accepted value for adequate statistical power is 0.80). The statistical power analyses suggest the rejection of the null hypothesis was likely not erroneous, but the effect size suggests the effect on prep chem final exam scores was moderate and the effect on CHEM 001A course grade was low.

Performance of Less Academically Prepared Students

Because the prep chem course was populated with students who scored lower on the campus-wide math placement exam, it was desired to determine if the online prep chem course might not be as effective for the least academically prepared students. To gain insight about whether less-prepared students might be negatively impacted by the online course treatment, a treatment group/math SAT interaction term was included in the multiple linear regression models summarized in Tables 3 and 4 (this interaction term is labeled TreatmentGroup*SAT). Using this interaction term in the multiple linear regression provided the opportunity to determine if the relationship between class treatment (Online versus In-person) is moderated by student math SAT performance.[20] Math SAT scores were chosen to create this interaction term due to the fact that enrollment in the prep chem course at this institution is based on a math placement exam, and since it has been previously shown that math SAT scores are strong predictors of performance in chemistry, the math SAT score was expected to moderate the class treatment/performance relationship.[23,24]

The multiple linear regression models summarized in Tables 3 and 4 indicate the class treatment/math SAT interaction term is a statistically significant predictor of performance on the prep chem final exam (standardized $\beta = 0.100$, $p = 0.040$), whereas it was not found to be a significant predictor of the CHEM 001A course grade (standardized $\beta = 0.080$, $p = 0.108$). These results suggest the effect of the class treatment on prep chem final exam performance is different at different levels of math SAT performance, but the effect of class treatment on CHEM 001A course grade is not significantly moderated by math SAT scores. To visualize how students at different levels of academic preparation performed in the Online course relative to the In-person course, regression plots of student performance versus class treatment participation (coded 0 and 1 for In-person and Online, respectively) were created at three levels of math SAT performance (see Figure S1A,B). To create these grouped plots, the entire study population was divided into tertiles based on math SAT scores, with the lowest third of scores being designated "Low SAT," the middle third of scores being designated "Medium SAT," and the highest third of scores being designated "High SAT." The regression plots for the final exam score dependent variable appear to confirm the notion that the relationship between class participation and final exam score is different at different levels of math SAT performance (see Figure S1A). More importantly, it appears students in the lowest third of math SAT preparation performed better in the Online course relative to the In-person course. This is evidenced by the fact that final exam scores were more positively correlated to class treatment for the Low SAT tertile (Low SAT, $R = 0.411$; Medium SAT, $R = 0.311$; High SAT, $R = 0.344$; see Figure S1A). Though the class treatment/math SAT interaction term was not statistically significant for the CHEM 001A general chemistry course grade analysis, the regression plot of CHEM 001A course grade versus treatment group suggests the least academically prepared students from the Online cohort may have performed better in the subsequent general chemistry course relative to the In-person cohort (Low SAT, $R = 0.279$; Medium SAT, $R = 0.189$; High SAT, $R = 0.145$; see Figure S1B). Since math SAT scores have been shown to be good predictors of student performance in general chemistry, the fact better performance on the math SAT appears to moderate the impact of the class treatment is not unexpected.

■ DISCUSSION

Performance in the Asynchronous Online Prep Chem Course and Research Limitations

Overall, the asynchronous online prep chem course appears to not only perform on par with the analogous in-person course but also result in higher prep chem final exam scores and higher course grades in the subsequent general chemistry course. The improved student performance that appeared to be associated with the Online prep chem course corroborates findings from previous studies of prep chem interventions. As mentioned in the introduction, Botch and co-workers reported that a 20 h self-paced series of online modules improved student success rates and course grades in the subsequent general chemistry course,[1] and the report from Dockter and co-workers suggested a self-paced prep chem experience based on the ALEKS online learning system also led to greater student performance in the subsequent general chemistry course relative to the traditional in-person prep chem course.[3] With regard to the specific research questions for the present study, it appears the Online course treatment (1) did not negatively affect student performance in the prep chem course, (2) did not negatively impact student performance in the subsequent general chemistry course, and (3) did not negatively impact the least academically prepared students. Because instructor-generated final exams were used to measure student performance in the prep chem courses, one limitation lies in the fact these assessments were not systematically vetted for reliability or validity. With that said, it is noted the exam questions were either modeled from typical end-of-chapter problems from a commonly used general chemistry textbook or were created by an instructor with over 20 years of experience teaching general chemistry in a higher education setting. Additionally, the item analysis for the Online final exam suggests the item difficulties and item-total correlations were in acceptable ranges, and because the Online course's final exam included questions taken directly from the previous

https://dx.doi.org/10.1021/acs.jchemed.0c00294
J. Chem. Educ. 2020, 97, 4287−4296

In-person course, the comparison of student performance on these two assessments is likely to be meaningful.

Because this was an observational study in which randomization of the student populations was not possible, and the Online treatment was not implemented with a rigorously controlled "teaching as usual" treatment group, a causal relationship between the Online course treatment and the improved student performance metrics used in this study cannot be determined. In particular, the multiple linear regression analysis in which grade performance in the subsequent general chemistry course was modeled could not account for instructor effects or other motivational factors among the students that might have arisen after the prep chem course treatment. The diminished explanatory power of the model in which the general chemistry course grade was included as the dependent variable is evidenced by the lower adjusted R^2 for this analysis, which indicates only approximately 10% of the variance in CHEM 001A course grades was explained by the included set of predictor variables (adjusted R^2 = 0.106; see Table S4A). Since the Online course and previous In-person course were taught by the same instructor, and the same set of lecture content from the In-person course was converted into the asynchronous online video lectures, the multiple linear regression analysis of prep chem final exam scores should have been less susceptible to the impact of uncontrolled confounding variables. Despite the strong similarities of the Online and In-person courses, the set of predictor variables included in the prep chem final exam score multiple linear regression model only accounted for approximately 16% of the variance in final exam scores (adjusted R^2 = 0.160; see Table S3A). This suggests there were other unaccounted for factors that may have influenced student performance in the Online course. One obvious source of variation in student performance in the prep chem course may have been the different length and format of the final exam for the Online course. However, because the final exam for the Online course allowed less time per question item and included a higher proportion of calculation-based questions than the In-person final (calculation-based questions typically take more time for students to complete), one might expect these factors would have negatively impacted the performance on the Online final exam. It was also noted in Methods that students in the Online course were monitored by proctors during the final exam; thus, improved student performance due to unauthorized collaboration or access to resources is likely ruled out. In total, the fact the Online cohort had higher scores, on average, than students in the In-person course is therefore likely not explained by the subtle differences in the final exam structure or implementation logistics. The low explanatory power of the regression model including the prep chem final exam score as the dependent variable could be explained by differences in individual student course loads, differences in the quality or quantity of learning community support, or personal events that might moderate academic performance. Unfortunately, these factors and/or other confounding variables could not be statistically controlled in the comparison of the Online and In-person cohorts.

Though a direct causal link between improved performance and participation in the Online course treatment cannot be made, a few comments on the potential impact of the asynchronous online course are warranted. The improved student performance on the prep chem final exam and course grades in the subsequent general chemistry course are particularly noteworthy given the fact students in the online prep chem did not have the benefit of the additional 120 min weekly discussion group sections. As highlighted in the Introduction, the meta-analysis performed by Means et al. found that the impact of online courses was generally higher for courses in which a blended learning format was adopted.[11] This likely results from the fact that including some face-to-face classroom activities provides greater opportunities for student discussion and sense-making and/or increased interactions with the instructor. The previous In-person prep chem course included a significant amount of group problem solving in the weekly discussion group sections, and the graduate student teaching assistants who facilitated these sessions also routinely encouraged the students to ask questions for concept clarification. The fact the Online course did not offer these opportunities for student group work or routine teaching assistant feedback suggests there was no intrinsic advantage to the Online course beyond the asynchronous online video content and periodic online assessments. This might suggest the general success of the online self-paced prep chem experience was possibly linked to a reduction in cognitive load, which would help mitigate the inherent limitations in working memory capacity.[13] However, future experimental or quasi-experimental research designs would be needed to definitively make this conclusion.

In addition to the inherent limitations of the observational research design, in which randomization of the study populations and creation of well controlled study groups was not possible, it is prudent to note other factors that might limit the scope of interpretation for the results described herein. First, as mentioned in Methods, students in the Online and In-person cohorts participated in first-year learning communities. Though students from both cohorts were enrolled in these learning communities in roughly the same proportion, the first-year learning communities at UCR have been shown to have positive impact on overall student retention and success. Therefore, not being able to include this experience as a covariate in the multiple linear regression might explain the relatively low explanatory power for these models. Second, though both cohorts of students were generally representative of the broader undergraduate population, the various academic motivations and/or career aspirations of the students could not be ascertained in this analysis. If trends in individual motivation and/or goals to matriculate to different post-baccalaureate pathways differed between the Online and In-person cohorts (e.g., desire to attend allied health professional school, and graduate school, etc.), this could have influenced student performance on the prep chem final and/or subsequent general chemistry course grade. Finally, there is an obvious external validity limitation in this study. If one were to conclude the Online course treatment had a significant impact on the student performance metrics, it would be necessary to carry out future experiments to determine if this effect would translate to different types of student populations at different types of undergraduate institutions.

Suggested Best Practices of Online Instruction

Despite the fact it is not possible to make a direct causal link between the Online course treatment and observed increase in student performance described herein, we provide a brief discussion for how the asynchronous online modules used in the Online course adhered to the principles laid out in Mayer's cognitive theory of multimedia learning. This is especially

https://dx.doi.org/10.1021/acs.jchemed.0c00294
J. Chem. Educ. 2020, 97, 4287−4296

important if instructors adopt online instruction that does not have any associated in-person discussion built into the course (i.e., blended learning). As stated in the Introduction, Mayer suggests the online learning environment should aim to reduce extraneous processing, manage essential processing, and foster generative processing.[13] Upon review of the Online learning materials, the learning glass lecture videos and online quizzes adhered to the guidelines proposed by Mayer reasonably well. In particular, the videos were segmented into several 10−15 min videos that could be viewed by the students at their own pace, rather than one 50−60 min live synchronous online lecture. By segmenting the essential course content and allowing students to self-pace their work, this likely managed essential processing and reduced cognitive load. Second, most of the videos included practice problems for which students were encouraged to stop the video and complete on their own before proceeding through the rest of the tutorial. This structure provided students an opportunity to make sense of the material being taught in the tutorial, and likely fostered generative processing. Conversely, the Learning Glass videos employed in the Online prep chem course likely had a design flaw that may have led to an increase in extraneous processing. Mayer posits that students can simultaneously engage an auditory and visual channel in a multimedia learning environment; however, if one of these channels is overloaded, it can result in cognitive overload and adversely affect the instructional objective.[13] The snapshot depicted in Figure 1 illustrates the Learning Glass picture-in-picture feature which allowed the opportunity to embed figures and computer-generated text into the video. This might seem to be a feature that enhances the production value of the video; however, providing the overlaying text along with the live drawing/sketching on the Learning Glass likely overloaded the visual channel for students to some extent. This is certainly a design feature that needs to be more carefully thought out in future implementations of the online prep chem course.

In addition to the general online instructional design features, we also propose that the types of learning outcomes associated with the course might also play a role in the potential improvement in student performance. For instance, the Online prep chem course described in the present study included learning objectives that would be considered zero- or one-dimensional skill-based learning objectives, as defined by the three-dimensional learning assessment protcol.[27] Because these types of learning objectives do not require more sophisticated conceptual understanding or linking of multiple concepts, it is possible these skill-based learning objectives are more readily achieved in an asynchronous online learning environment. Upon reviewing the prep chem implementations of Botch et al. and Dockter et al., it seems these previous interventions also focused on skill-based learning objectives such as math skills, algorithmic problem-solving, measurement and dimensional analysis, stoichiometric calculations, basic atomic theory, and naming chemical compounds.[1,3] If two- or three-dimensional learning objectives were to be included in an online prep chem course, it is possible designing an online learning environment that leads to positive outcomes might be more challenging for instructors. For instance, one could imagine the difficulty in fostering generative processing in an asynchronous online learning environment for learning objectives that involve cross-cutting concepts and/or deeper conceptual understanding of core content knowledge. How would students articulate this type of understanding and how could instructors provide feedback on this type of thinking in an asynchronous online environment? This might be a case in which a blended learning structure, which combines elements of asynchronous online learning with in-person discussion, would be necessary to achieve the desired student learning outcomes. Mayer notes the research on multimedia learning needs to be expanded to gain insight about what type of interventions will work best to promote conceptual learning versus algorithmic problem solving, and longer-term transfer of knowledge versus knowledge retention.[13] Perhaps this will be an area of growing interest to the chemistry education research community.

CONCLUSIONS

The implementation of online instruction is often driven by external pressures such as limits in classroom space, lack of instructional personnel, and/or the desire of colleges and universities to expand their enrollments beyond the typical residential full-time student. The results presented herein add to the existing literature that suggests a fully online course can result in student performance that is observed to be significantly greater than the student performance in the equivalent in-person course, and therefore should be offered for this more intrinsic reason. Because it was found the fully online asynchronous prep chem course resulted in improved student performance outcomes, both in the prep chem course itself and in the subsequent general chemistry course, more chemistry departments at institutions of higher learning should strongly consider developing similar online prep chem courses. Most importantly, the analysis presented here also revealed the least academically prepared students in the course were not adversely affected by a fully online asynchronous learning environment. This suggests using an asynchronous online prep chem course has the potential to help less-prepared students succeed in general chemistry, ultimately improving retention in STEM disciplines.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.0c00294.

Data Tables S1−S6; Figures S1A, S1B; Appendix 1 listing final exams for online and in-person courses with item analysis; Appendix 2 showing grading systems for CHEM 001A courses; Appendix 3 listing tests for assumptions for multiple linear regression analyses (PDF)

Syllabus for online prep chem course (PDF)

Syllabus for in-person prep chem course (PDF)

AUTHOR INFORMATION

Corresponding Author

Jack F. Eichler − *Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States;* orcid.org/0000-0003-2755-3636; Email: jack.eichler@ucr.edu

Authors

Grace Henbest − *Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States*

https://dx.doi.org/10.1021/acs.jchemed.0c00294
J. Chem. Educ. 2020, 97, 4287−4296

Kiana Mortezaei − *Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States*
Teresa Alvelais − *Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States*
Courtney Murphy − *Department of Chemistry, University of California, Riverside, Riverside, California 92521, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jchemed.0c00294

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge the University of California Office of the President Innovative Learning Technology Initiative (ILTI) for funding the development of the online preparatory chemistry course. We also gratefully acknowledge the anonymous manuscript reviewers, whose feedback and suggestions helped improve the overall quality of the final manuscript.

■ REFERENCES

(1) Botch, B.; Day, R.; Vining, W.; Stewart, B.; Hart, D.; Rath, K.; Peterfreund, A. Effects on student achievement in general chemistry following participation in an online preparatory course: Chem Prep, a voluntary, self-paced, online introduction to chemistry. *J. Chem. Educ.* **2007**, *84* (3), 547.
(2) Bayer Corp.. Bayer facts of science education XV: A view from the gatekeepers-STEM department chairs at America's top 200 research universities on female and underrepresented minority undergraduate STEM students. *J. Sci. Educ. Technol.* **2012**, *21*, 317−324.
(3) Dockter, D.; Uvarov, C.; Guzman-Alvarez, A.; Molinaro, M. Improving preparation and persistence in undergraduate STEM: Why an online summer preparatory chemistry course makes sense. In *Online Approaches to Chemical Education*; Sörensen, P. M., Canelas, D. A., Eds.; ACS Symposium Series, Vol. *1261*; American Chemical Society: Washington, DC, USA, 2017; pp 7−33, DOI: 10.1021/bk-2017-1261.ch002.
(4) View All Courses. *Catalog*, Cross-Campus Enrollment, University of California; http://crossenrollcourses.universityofcalifornia.edu/catalog?subject_area=7&with_preview=&pageSize=10 (last accessed 2020/10/29).
(5) Eitemüller, C.; Habig, S. Enhancing the Transition? Effects of a Tertiary Bridging Course in Chemistry. *Chem. Educ. Res. Pract.* **2020**, *21*, 561−569.
(6) Schmid, S.; Youl, D. J.; George, A. V.; Read, J. R. Effectiveness of a short, intensive bridging course for scaffolding students commencing university-level study of chemistry. *Int. J. Sci. Educ.* **2012**, *34* (8), 1211−1234.
(7) Smith Falk, A.; Guzman-Alvarez, A.; Molinaro, M. Understanding the curve: Implications of norm-referenced grading in large introductory science courses (Poster). *35th Annual Conference, Society for Teaching and Learning in Higher Education (STLHE)*, Jun. 16−19, 2015, Vancouver, BC, Canada.
(8) Myers, C. B.; Bennett, D.; Brown, G.; Henderson, T. Emerging online learning environments and student learning: An analysis of faculty perceptions. *J. Educ. Technol. Soc.* **2004**, *7* (1), 78−86.
(9) Williams, S. L. The effectiveness of distance education in allied health science programs: A meta-analysis of outcomes. *Am. J. Dist. Educ.* **2006**, *20* (3), 127−141.
(10) Shachar, M.; Neumann, Y. Differences between traditional and distance education academic performances: A meta-analytic approach. *Int. Rev. Res. Open Distrib. Learn.* **2003**, *4* (2), 153.
(11) Means, B.; Toyama, Y.; Murphy, R.; Baki, M. The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teach. Coll. Rec.* **2013**, *115* (3), 030302.
(12) Faulconer, E. K.; Griffith, J. C.; Wood, B. L.; Acharyya, S.; Roberts, D. L. A comparison of online and traditional chemistry lecture and lab. *Chem. Educ. Res. Pract.* **2018**, *19* (1), 392−397.
(13) Mayer, R. E. Applying the Science of Learning to Multimedia Instruction. In *The Psychology of Leaning and Motivation: Cognition in Education*; Mestre, J. P., Ross, B. H., Eds.; Academic Press: San Diego, CA, USA, 2011; Vol. *55*, DOI: 10.1016/B978-0-12-387691-1.00003-X.
(14) *Learning Glass*, Learning Glass Solutions; https://www.learning.glass/product-line/?gclid=EAIaIQobChMIwrzjzMiY6AIVZxitBh3vUg-nEAAYASAAEgJ2fvD_BwE (last accessed 2020/10/29).
(15) *ProctorU*, https://www.proctoru.com/ (last accessed 2020/10/29).
(16) General Chemistry. *Open Stax*, Rice University, Houston, TX, USA; https://openstax.org/details/books/chemistry-2e (last accessed 2020/10/29).
(17) Murnane, R. J.; Willett, J. B. In *Methods Matter: Improving Causal Inference Educational and Social Science Research*; Oxford University Press: New York, NY, USA, 2011; Vol. *1*, 286−332.
(18) *IBM SPSS Statistics for Windows*, Version 24.0; IBM: Armonk, NY, USA, Released 2016.
(19) Norman, G. Likert scales, levels of measurement and the "laws" of statistics. *Adv. in Health Sci. Educ.* **2010**, *15*, 625−632.
(20) Aiken, L. S.; West, S. G. *Multiple regression: Testing and Interpreting Interactions*; Sage: Newbury Park, London, U.K., 1991.
(21) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: New York, NY, USA, 1988; Chapter 9.
(22) Enrollments: Demographic. *Institutional Research*, University of California, Riverside;https://ir.ucr.edu/stats/enroll/demographic (last accessed 2020/10/29).
(23) Vincent-Ruz, P.; Binning, K.; Schunn, C. D.; Grabowski, J. The Effect of Math SAT on Women's Chemistry Competency Beliefs. *Chem. Educ. Res. Pract.* **2018**, *19*, 342−351.
(24) Zwick, R.; Sklar, J. C. Predicting College Grades and Degree Completion Using High School Grades and SAT Scores: The Role of Student Ethnicity and First Language. *Am. Educ. Res. J.* **2005**, *42*, 439−464.
(25) Bancroft, S. F.; Fowler, S. R.; Jalaeian, M.; Patterson, K. Leveling the Field: Flipped Instruction as a Tool for Promoting Equity in General Chemistry. *J. Chem. Educ.* **2020**, 97 (1), 36−47.
(26) Li, L. Gender Equality in Science—Who Cares? *J. Chem. Educ.* **2002**, *79* (4), 418.
(27) 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*, No. e0162333.

Chemistry Education Research and Practice

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: Chem. Educ. Res. Pract., 2021, 22, 842

Incorporating concept development activities into a flipped classroom structure: using PhET simulations to put a twist on the flip†

Hoi-Ting Wu,[a] Kiana Mortezaei,[a] Teresa Alvelais,[a] Grace Henbest,[a] Courtney Murphy,[a] Ellen J. Yezierski *[b] and Jack F. Eichler *[a]

Implementation of the flipped classroom approach into STEM courses has been popularized in the last decade and has generally been reported to improve student performance outcomes. In a flipped classroom setting, students typically first encounter course content in the online format and subsequently engage in some form of active learning during the in-person class meetings. Although the flipped classroom approach can promote increased student engagement and provide an opportunity to apply content encountered in the classroom, this structure does not generally give students opportunities for discrete concept development prior to the application phase of learning. In an effort to build concept development activities into a flipped classroom structure, five learning cycle activities were implemented in a large enrollment first-term general chemistry course that has previously implemented the flipped classroom design. Four of these learning cycle activities incorporated PhET simulations as part of the exploration phase of learning, and all five activities were facilitated during the in-person class meetings to initiate the learning cycle. The activities were designed to help students explore models and engage in concept development. The application phase of the learning cycle was facilitated by flipped classroom modules or in-person classroom activities that included whole-class questioning coupled with collaborative think-pair-share discussion. Performance gains in conceptual understanding were evaluated by employing a one-group, pre-post-post research design. Non-parametric Friedman's tests indicate a significant main effect across time for each concept development activity, and *post hoc* Wilcoxon signed rank tests indicate the post-test and final exam scores are significantly higher than the pre-test scores for each activity ($p < 0.001$ for each pre-post and pre-final pairwise comparison). The findings reported herein demonstrate that concept development activities can be successfully integrated with flipped classroom modules and the combination of the introductory learning cycle activities and flipped classroom application activities led to knowledge gains that persisted through the end of the course. In total, creating this type of blended learning environment appears to help students achieve understanding of core general chemistry concepts.

Received 23rd March 2021,
Accepted 20th May 2021

DOI: 10.1039/d1rp00086a

rsc.li/cerp

Introduction

Studies on the efficacy of the flipped classroom course structure have grown steadily over the last decade and have generally shown that this classroom intervention leads to an improvement of student learning outcomes relative to traditional didactic lecture (Jensen *et al.*, 2015; Casselman *et al.*, 2019; Naibert *et al.*, 2020). Implementations of flipped classrooms vary, but typically some portion of the traditional in-class lecture is moved to an online learning space followed by the facilitation of active learning in the subsequent live classroom period. The type of classroom activities that are facilitated during the in-person instruction are wide ranging, but generally include some form of collaborative group learning (Weaver and Sturtevant, 2015), formal peer-led learning (Liu *et al.*, 2018), or whole-class questioning with think-pair-share discussion (Flynn, 2015). Regardless of the exact nature of the face-to-face learning activity, the in-class time is commonly structured to help students build upon the content encountered in the asynchronous pre-class learning environment (Naibert *et al.*, 2020).

Although the flipped classroom structure has been identified as a possible way to reduce cognitive load relative to the

[a] *Department of Chemistry, University of California, 501 Big Spring Road, Riverside, California 92521, USA. E-mail: jack.eichler@ucr.edu; Tel: +(951) 827-3794*
[b] *Department of Chemistry and Biochemistry, Miami University, 360B Hughes Laboratories, Oxford, Ohio 45056, USA. E-mail: yeziere@MiamiOH.edu; Tel: +(513) 529-2819*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1rp00086a

delivery of content in a traditional lecture format (Seery, 2015), the initial steps of concept development are not well-understood or deliberately supported by the usual pre-class content. This is likely because students typically only have a listening and/or note-taking role in the online learning space. Thus, there is an opportunity and a need to build concept development into the flipped classroom design. This could be achieved by including an in-class activity prior to the asynchronous online learning module, in which exploration and concept development are facilitated through sense-making of phenomena, models, and data. This modification is aligned with a pedagogical method known as the learning cycle (Lawson and Karplus, 2002), which includes exploration, term introduction or concept development, and application phases of instruction. As mentioned above, the flipped classroom already has a strong application component in which the in-class activities build upon the pre-class online content. However, the exploration and concept development phases cannot be readily integrated into pre-class asynchronous learning activities, and to our knowledge incorporating preliminary concept development into a flipped classroom structure has not been reported. There are reports of incorporating POGIL (Process Oriented Guided Inquiry Learning) activities in the in-person portion of the flipped classroom (Hibbard *et al.*, 2016; Canelas *et al.*, 2017). Though POGIL involves collaborative group learning that aims to facilitate conceptual understanding through exploration, explanation, and sense-making processes (Moog and Spencer, 2008), the studies reported by Hibbard *et al.* and Canelas *et al.* were built on flipped classroom structures in which the POGIL activities were completed after the pre-class learning assignments.

In the present study, learning cycle activities that promoted exploration and concept development were implemented as introductory in-class activities prior to standard flipped classroom modules, which were then followed by in-class activities centered on whole-class questioning and think-pair-share group learning. The learning cycle activities were developed for five foundational learning objectives in the first-term general chemistry curriculum, four of which employed PhET simulations to engage students in the exploration phase of the learning cycle. PhET interactive simulations‡ were integrated in the concept development activities, because they can provide immediate feedback to users and are designed to support student connection-making and focus on active knowledge construction (D'Angelo *et al.*, 2014).

Theoretical frameworks for learning

The flipped classroom approach is built upon theoretical frameworks that support student learning in both the pre-class online and in-person classroom environments. With respect to the pre-class online learning, Sweller's cognitive load theory has been previously identified to be one of the foundational frameworks linked to the impact of the flipped classroom (Seery, 2015). Cognitive load theory states that because a learner has limited working memory capacity and short-term memory, excessive levels of cognitive load can interfere with meaningful learning (Sweller, 1988). The rigid structure of traditional lecture settings can limit students' ability to process and understand new information at their own pace. In the flipped classroom, some of the traditional lectures are replaced with pre-class online activities that can be completed by students with fewer time constraints and opportunities to reflect on their learning, ultimately leading to reduced cognitive load. Numerous studies on the efficacy of flipped classroom have been published in the chemistry education literature in recent years. A *meta*-analysis of evidence-based instructional practices in undergraduate chemistry courses concluded flipped classroom implementations generally lead to positive outcomes with respect to student performance (positive effect sizes ranging from small to medium were observed in the 15 studies included in the analysis), and therefore it should be considered for use as an evidence-based instructional practice that promotes student success (Rahman and Lewis, 2019).

Because the in-person classroom activities associated with flipped classroom implementations are often built on some form of collaborative group learning, this aspect of the flipped classroom is broadly grounded in Vygotsky's social constructivism framework (Vygotsky, 1978). Social constructivism emphasizes the importance of the collaborative nature of learning, and posits that cognitive functions are closely linked to social interactions. Classroom exercises such as peer-led team learning, POGIL, and even less formal think-pair-share activities are partially rooted in social constructivism due to the dependency upon collaborative work that supports learning in these interventions. The impact of social constructivism on the flipped classroom assumes collaborative group learning is embedded in the in-person classroom activities, and it should be noted that the opportunities for group interactions may indeed vary across different flipped classroom implementations.

Von Glasersfeld has proposed that constructivism is a framework in which knowledge is constructed in the mind of the learner from experience, and the function of cognition is to organize such experiences (Von Glasersfeld, 2001). This view of constructivism can be explicitly connected to classroom instruction through the learning cycle instructional model (Lawson and Karplus, 2002). The learning cycle engages students in iterative progressions of exploration of phenomena, concept invention/term introduction, and application. The learning cycle is particularly germane to our thinking about the flipped classrooms, because as noted above, the in-class active learning that follows the pre-class modules is typically an application activity. The learning cycle is notable for its student-centeredness and the opportunities it affords for students to build knowledge from experiencing phenomena in the exploration phase and making sense of the findings from the exploration in the concept invention/term introduction phases. It is important to consider that a teacher-centered lecture is antithetical to the experienced-focused exploration phase of the learning cycle. Applying the constructivist framework underlying the learning cycle, the current study aims to help students develop conceptual

‡ PhET Interactive Simulations. From https://phet.colorado.edu/.

knowledge through the exploration and concept invention activities that take place in class prior to more typical flipped classroom structures.

Methodological framework

To better integrate the constructivist approach into the in-class activities, it is necessary to engineer a classroom experience in which students engage in concept development within an existing flipped classroom framework. Such engineering requires an iterative process of activity design, evaluation, and redesign until desired outcomes are achieved. Design-based research (Cobb *et al.*, 2003; Wang and Hannafin, 2005) was employed by Minshall and Yezierski to iteratively design and revise a learning cycle activity for general chemistry on bond making and breaking (Minshall and Yezierski, 2021). This work served as proof-of-concept for the learning efficacy of the intervention and informed how analogous activities could be designed and implemented for the current study. With respect to evaluating the efficacy of newly designed activities, a one-group pre-post-post test design was employed. This quasi-experimental approach allowed for the effects of the activities to be detected while controlling for prior knowledge (pretest).

Four of the activities described here incorporated PhET simulations as part of the exploration phase of the learning cycle. The use of PhET simulations as a way to promote investigative inquiry and foster sense-making in chemistry instruction has been previously reported (Lancaster *et al.*, 2013), and the use of the PhET Atomic Interactions simulation§ within the exploration phase in a concept development activity was described by Minahall and Yezierski (Minshall and Yezierski, 2021). The implementation of the Atomic Interactions activity followed a quasi-experimental design in which each new version of the activity was tested to evaluate the quality of the material in term of students' learning outcomes. The present study builds upon this work, with the aim to further extend the learning benefits of interactive simulations to other foundational concepts in the general chemistry curriculum.

Three new activities that incorporated PhET simulations were used in addition to the previously reported Atomic Interactions activity, and one additional activity used a static POGIL model to engage students in concept exploration (Moog and Farrell, 2002). Students worked through the learning cycles in small and collaborative groups in class before interacting with the flipped classroom modules or in-person instruction that incorporated whole-class questioning with think-pair-share collaborative learning. The learning cycle activities provided students the opportunity to engage in concept development by making observations within the simulation/model, respond to concept development questions, engage in group discussion, verbally articulate their thinking, and receive feedback on that thinking from peers. The students then built on that conceptual foundation as they progressed through the subsequent flipped classroom modules and whole-class questioning activities. Five learning objectives addressed in most first-term general chemistry courses were chosen and targeted for the learning cycle activities, because the authors have routinely observed that students often struggle with core concepts related to these five learning objectives (see Table 1). These five topic areas are included in the General Chemistry Anchoring Concepts Content Map (Holme and Murphy, 2012), suggesting these are pertinent to most undergraduate general chemistry courses.

The theoretical and methodological frameworks described above were integrated within a first quarter general chemistry course to address the following research questions:

1. How can the use of introductory learning cycle activities foster conceptual learning for five foundational concepts in a first-term undergraduate general chemistry course?

2. How will knowledge gains (if realized) persist over the course of the term?

3. How can analyses of incorrect student responses provide insight about lingering inaccurate chemical ideas?

Methods

Setting and sample

The learning cycle activities were implemented in a first-quarter general chemistry course (CHEM 001A) taken by all science majors at a large research-intensive public university that is designated as a Hispanic-serving institution (HSI). The topics addressed in this course included: scientific practices and measurement; atomic structure; electronic structure of the atom; chemical bonding and molecular structure; compounds and the mole, chemical reactions, and stoichiometry. The course met over 10 weeks, twice each week for 80 minutes (total enrollment = 231 students), and associated recitation sections met once each week for 50 minutes (30–40 students per recitation section). The details of the course structure are provided in the course syllabus, which is provided in Appendix 1 (ESI†).

Students were informed through a verbal consent process that their performance on the pre- and post-test assessments would be evaluated for research purposes (under approved IRB protocol HS-10-135). They were also informed that all pre/post/post test data would be reported in aggregate form (*i.e.*, all data would be reported in a way that would not reveal individual student identities) and that they could request to have their test data excluded at any time. During the course of the study no students requested to have their data excluded from the analysis.

Learning cycle activities, flipped modules, and in-class activities

Five important concept areas in first-term general chemistry were chosen for learning cycle activities. The Atomic Interactions activity by Minshall and Yezierski was used as previously described (Minshall and Yezierski, 2021), and authors EJY and JFE co-wrote the other four learning cycle activities. Four out of five activities were designed based on the available PhET

§ PhET Atomic Interactions simulation. From https://phet.colorado.edu/en/simulation/atomic-interactions.

Table 1 List of learning cycle activities and PHET simulations used in present study

Learning cycle activity title	Version and year of PhET simulation	Learning cycle activity sheet and teaching notes[a]	Learning objectives students will be able to:
Nuclear Atom	N/A	Nuclear atom POGIL activity (Moog and Farrell, 2002).	(1) Identify the atomic number and mass of atoms for different elements; (2) Identify the importance of atomic number in identifying atoms of different elements; (3) Identify different isotopes of the same element; and (4) Use the total number of protons and electrons to determine if ions are present and determine the charge of the ion.
Coulomb's Law	1.0.9, 2019‡‡	https://phet.colorado.edu/en/contributions/view/5909	(1) Predict how electrostatic attraction/repulsion changes as a function of distance between charged particles; (2) Predict how electrostatic attraction/repulsion changes as a function of the magnitude of charge between charged particles; and (3) Describe the concept of electrostatic attraction/repulsion in preparation for applying it to atoms.
Atomic Interactions	1.1.0, 2019§	https://phet.colorado.edu/en/contributions/view/5860	(1) Predict how the potential energy of a bond changes with changing interatomic distance; (2) explain how attractive and repulsive Coulombic forces impact the potential energy of a chemical bond; and (3) Predict if bond making and bond breaking events are exothermic or endothermic.
Molecular Shape	1.2.8, 2019§§	https://phet.colorado.edu/en/contributions/view/5910	(1) Use Lewis structures to determine the number of bonded atoms and number of non-bonding electron pairs on the central atom of a molecular structure; (2) Predict the molecular geometry and molecular shape of structures using the VSEPR model; and (3) Predict if the real bond angle is approximately equal to or less than the ideal bond angle for different types of molecular shapes.
Molecular Polarity	1.0.15, 2019¶¶	https://phet.colorado.edu/en/contributions/view/5911	(1) Explain the difference between a bond dipole and a molecular dipole; (2) Use electronegativity and knowledge of molecular shape to predict if molecules possess a net molecular dipole or not; and (3) Use electrostatic potential maps to explain if a net molecular dipole exists or not.

[a] The activity sheets, pre/post/post test items, and detailed teaching notes are available to download from the PhET website.

simulations (Table 1), and each was completed by students in an 80 minute class period. A previous study revealed some students encountered difficulty relating electrostatic forces and potential energy (Minshall and Yezierski, 2021). Thus, the Coulomb's Law activity was implemented prior to the Atomic Interactions activity with the aim to build that prerequisite knowledge.

To help students develop the preliminary concepts independently while engaging with the activities, they were divided into ad-hoc groups of 3–4 students to explore the PhET simulations (Table 1) and complete the activity sheets. The students were instructed to write their answers on the worksheet. The course instructor and two graduate student teaching assistants were available to assist and answer questions as needed, and the instructor had the simulation running on the classroom projector. If points of confusion arose regarding the using the simulation, the instructor would demonstrate parts of the simulation for the class; however, students needed to make their own observations and document results. The activity sheets and detailed teaching notes for the activities that included the PhET simulations, along with the activities sheet and teaching notes are available online (see Table 1). The Nuclear Atom activity is a POGIL worksheet that includes a static drawing in lieu of a PhET simulation (Moog and Farrell, 2002).

After the introductory learning cycle activities, students engaged in either a flipped classroom module or an in-person classroom structure that employed whole-class questioning with think-pair-share collaborative group learning (see Fig. 1). A detailed description of the flipped classroom modules and whole-class questioning activities, and how these exercises were timed throughout the term is provided in Appendix 2 (ESI†). For the flipped classroom modules, students were given 3–4 days to watch instructor-created videos and complete the associated online quiz that was due prior to the in-class application exercises. The online videos also included embedded questions that students completed in the Playposit system.¶ For the flipped module in-class application exercises, the instructor briefly reviewed the material from the online videos and introduced the topic to be discussed that day in the in-class worksheet. Students then worked in collaborative groups of 3 or 4 while completing the worksheet, and they had access to the instructor and two graduate student teaching assistants if clarification was needed. Student work was evaluated by converting the free response questions to multiple choice format, and students submitted answers using the classroom polling system (PollEverywhere‖). The instructor then used the last 10–15 minutes of the class period to review any questions in which a large proportion of students submitted incorrect responses and concluded by summarizing the main learning objectives. For some topics, it was determined that students should engage with in-class application exercises in the classroom period immediately after the learning cycle activity

¶ Playposit is available at: https://go.playposit.com/.
‖ PollEverywhere is available at: https://www.polleverywhere.com/.

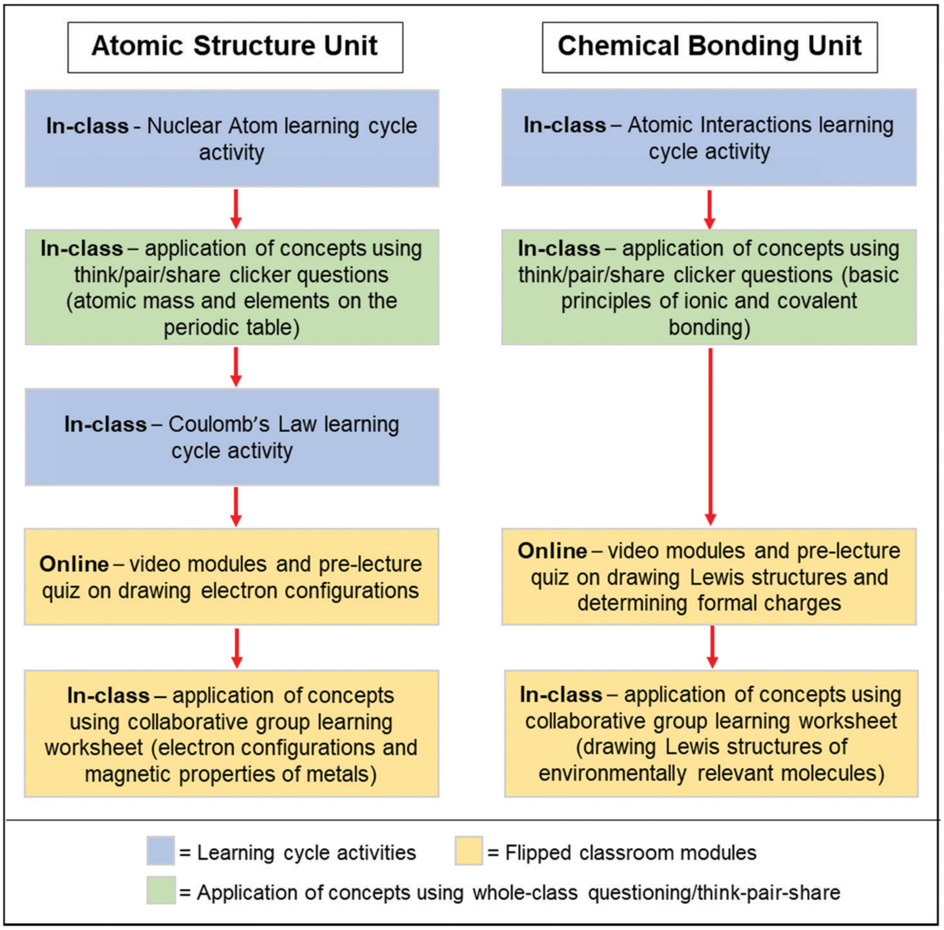


Fig. 1 Representative examples for how the concept development activities and flipped/hybrid modules were administered. The full schedule of all concept development activities and flipped/hybrid classroom modules is provided in Appendix 1 (ESI†), and the detailed description of the classroom and online learning activities is provided in Appendix 2 (ESI†) (● blue boxes denote learning cycle activities; ● green boxes denote application of concepts using whole-class questioning/think-pair-share; ● and yellow boxes denote flipped classroom modules).

(see Appendix 2, ESI†). No online pre-class learning modules were included in these instances, and students applied concepts developed during the learning cycle activities in the subsequent class meeting *via* instructor-led discussion and whole-class questioning that included think-pair-share collaborative group learning. Analogous to the learning cycle activities, student responses to the whole-class questions were submitted using the classroom polling system. The course syllabus provided in Appendix 1 (ESI†) describes how the performance in the in-class polling factored into the overall course grade.

Evaluation and test instruments

Evaluation of the efficacy of the learning cycle activities was carried out using a one-group pre-post-post quasi-experimental design. Students were instructed by email and in-person during the class meeting 2–3 days prior to the activity to bring a laptop or other electronic device capable of running the iLearn/Blackboard quiz system** on the day of the activity. Students were also informed that free laptops and tablets were available from the campus library. On the day of the activity, students took the pre-test individually and without using any online resources. Students were informed that the pretest would not count toward their course grade, but they were encouraged to do their best work. Based on the instructor's in-class observations, no obvious collaboration took place. The pre-test in iLearn was made unavailable to the students immediately after the pre-test submission. At the end of the activity, worksheet questions that addressed the overall learning objectives of the activity were converted to clicker questions, and the instructor facilitated the input of these questions (*e.g.*, scaffolded activity questions 6b and 6c from the Atomic Interactions activity were used as in-class clicker questions§). The instructor then facilitated a class discussion in which students reported out answers to the activity questions, and the most important concepts or learning objectives were highlighted. To evaluate changes in knowledge, students completed an activity post-test by the end of the day subsequently to completing the activity in class. Students were informed that the post-test was counted for extra credit and instructed not to collaborate with other students (note: the extra points available for all post-test

** iLearn/Backboard quiz system. From https://www.blackboard.com/teaching-learning/learning-management.

questions constituted less than 1% of the total available course points). The post-test in iLearn was left open for viewing to students for the remainder of the quarter, but students could not alter their answers after the test submission. All the pre- or post-test questions from the iLearn quiz administrations were embedded into the final exam. The final exam review guide encouraged students to review all the in-class activities for the final exam, but the students were not informed that the pre-/post-test questions would be included on the final exam.

Authors EJY and JFE co-wrote all the test items and came to consensus in confirming the content validity of the final set of test items for each activity. The single administration reliability of the test items was evaluated by calculating Cronbach's alpha for the set of test items for each activity, and the ability of the individual items to discriminate between high and low performers was estimated by calculating the item-total correlation values for the items (see Appendix 3, Tables 1 and 2, ESI†). These analyses suggest the scores on items across administrations for each activity were correlated to one another and generally appeared to discriminate between high- and low-performing students. However, the length of the tests only ranged from 2–5 items, making it difficult to arrive at definitive conclusions regarding the reliability of the tests (Tavakol and Dennick, 2011). To gain additional insight about how well the test items measured changes in student knowledge, Sankey diagrams were used to track responses at each time point of test, for each individual item (see Appendix 5, Fig. 1 and 2, ESI†). These data indicate the tests detected clear changes in student knowledge for all the activities except the Coulomb's Law activity, for which the students appeared to demonstrate a high level of *a priori* knowledge. In total, the item analyses and reliability data suggest the tests were acceptable measures of student knowledge for these five conceptual domains. A detailed summary and explanation of the item analysis and single administration test reliability are provided in Appendix 3 (ESI†).

Statistical analyses

To evaluate the impact of the learning cycle activities on student learning, inferential statistical analyses were conducted on student scores on the tests across three time points. All statistical analyses were conducted on SPSS (Statistics Package for the Social Sciences) Statistics 26 software package.†† Because of the non-parametric nature of the test data, a Friedman's test was employed to determine if student performance changed significantly over the three testing points for each activity assessment. A Kendall's concordance test was used to estimate the effect size of the test scores for each activity. To determine at which testing points significant changes in performance occurred, a *post hoc* Wilcoxon signed ranked test was used to evaluate pair-wise comparisons among scores (pre-test, post-test, and final exam) for each activity. The level of significance for the pair-wise comparisons was adjusted using a Bonferroni correction to control the family-wise error rate.

Table 2 Friedman test statistics and effect size measurements. A non-parametric Friedman Test was conducted to measure the learning gains across three testing time points for each activity, and Kendall's concordance coefficient (W) was calculated to determine the effect size

	Friedman test statistics			
Activity	χ^2	df	p	W^a
Nuclear atoms	216.515	2	<0.001	0.558
Coulomb's law	72.460	2	<0.001	0.187
Atomic interaction	249.213	2	<0.001	0.642
Molecular shape	156.071	2	<0.001	0.402
Molecular dipole	134.978	2	<0.001	0.348

[a] $W = 0.1$ indicates a small effect, $W = 0.3$ indicates a medium effect, and $W > 0.5$ indicates a large effect.

Results and discussion

Conceptual learning outcomes

To answer the first two research questions that aimed to determine how the learning cycle activities foster conceptual learning and if knowledge gains persisted over time, student performance was evaluated for each of the five activities at three time points. The general performance gains for the five activities in pre-test, post-test, and the final exam are shown in Fig. 2. Fig. 2a shows the distribution of the number of correct responses (or score) in all activities across the three time points. Out of the 19 test questions, the average score improved from the pre-test (11) to both the post-test (16) and final exam (17). The results suggest that knowledge gains were observed right after the learning cycle activity, and this knowledge was sustained through the final exam. Fig. 2b–f show the distribution of correct responses for each activity across the three time points. A cursory examination of the distributions of scores shows an immediate overall improvement for all five learning cycle activities as evidenced by the increase in number of students who answered all the test items correctly on the post-test (see Fig. 2b–f). The number of students who answered all the test items correctly on the final exam was equal to or greater than that observed on the post-test for all the activities except the Molecular Shapes learning cycle activity, in which there was a decrease in the total number of students who answered all the test items correctly on the final exam relative to the post-test.

Because the Atomic Interactions was previously developed and evaluated by Minshall and Yezierski (Minshall and Yezierski, 2021), it was of interest to compare the results of the current study to this previous report. Additionally, the learning objectives for the Atomic Interactions activity include having students correctly identify the energy changes associated with chemical bond forming and breaking processes. Since students often incorrectly associate bond breaking with an exothermic energy change (Boo, 1998), it was also desirable to determine if the efficacy of this activity would translate to a large enrollment course. As shown in Fig. 2d, approximately 73% of the students incorrectly answered the pre-test questions

†† IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.

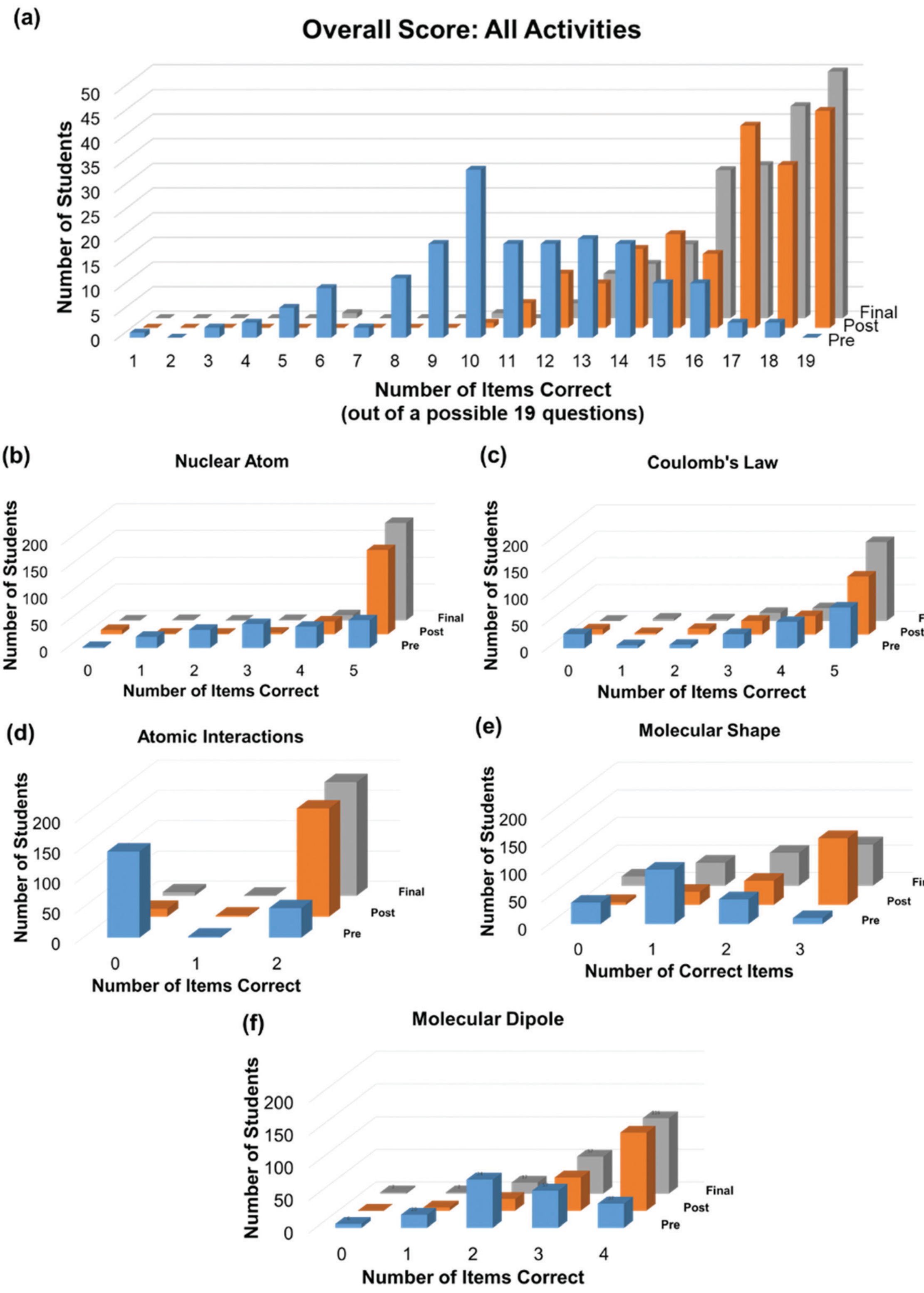


Fig. 2 The total number of students who correctly answered the different number of test items for pretest (blue), posttest (orange), and final exam (grey). These data are shown for: (a) the combined test items for all activities; (b) Nuclear Atom test items; (c) Coulomb's Law test items; (d) Atomic Interactions test items; (e) Molecular Shape test items; and (f) Molecular Dipole test items.

regarding whether bond breaking and bond making events are exothermic or endothermic. However, most students successfully demonstrated improved performance after the post-test (94% of students answered both test items correctly; see Fig. 2d) and retained this conceptual knowledge on the final exam (97% of students answered both test items correctly; see Fig. 2d). It is noteworthy that a higher percentage of students correctly answered the questions on the final exam using the same activities than the previous study carried out by Minshall and Yezierski, in which 82% of the 34 students were able to able to identify and distinguish between exothermic and endothermic processes (Minshall and Yezierski, 2021). The results from this current study are highly encouraging, as the sample is almost 6 times the size of previous study (231 verse 34 students). It suggests that the activities can be implemented in a large enrollment course without sacrificing student performance gains. This observation might be explained by the implementation of the Coulomb's Law activity‡‡ prior to the Atomic Interactions activity in the present study. Because the knowledge of attractive and repulsive electrostatic forces is required to explain how the potential energy of a chemical bond changes as a function of interatomic distance, we speculate that including the Coulomb's Law activity might have primed students to make better sense of their observations during the Atomic Interactions activity.

To determine if the changes in student performance over time were statistically significant, a non-parametric Friedman test was conducted to evaluate differences among the total number of correct test items at the three time points for each activity. The Friedman's Chi-square value (χ^2) for all activities was found to be significant ($p < 0.001$), indicating the learning cycle activities had a significant effect on student performance (Table 2). The Kendall's Coefficient of Concordance (W) was used to measure the effect size of each activity. A W value of 0.5 or above indicates large differences among the three time points. Most activities have medium or large effect sizes; only the Coulomb's Law activity had a small effect size with this group of students ($W = 0.187$). The smaller effect size observed for the Coulomb's Law activity is possibly explained by the fact that the concepts addressed in this activity ended up being less challenging than originally expected (see Appendix 5, Fig. 1a–d, ESI†), as the students appeared to have a high level of *a priori* knowledge on electrostatic attraction/repulsion.

Post hoc pairwise comparisons of the non-normally distributed test scores were conducted using a Wilcoxon-signed rank test (Table 3) to determine if the student performance was significantly different at the three testing points. The effect size (r) for each pairwise comparison was calculated using previous published method (Tomczak and Tomczak, 2014), and the cutoffs for r are as follows: 0.1 (small effect), 0.3 (medium effect), and above 0.5 (large effect) (Cohen, 1988). Students' test scores improved from the pre-test to both the post-test and final exam, for all five activities (for all post-test/pre-test and final exam/pre-test pairwise comparisons; $p < 0.001$). The learning cycle activities yield large effect sizes for test performance for the Atomic Interactions and Molecular Shape activities, medium effect sizes for the Nuclear Atom and Molecular Dipole activities, and a small effect size for the Coulomb's Law activity (see post-test/pre-test pairwise comparisons in Table 3). The combination of the learning cycle activities and subsequent flipped modules/in-class exercises yield a medium to large effect for all five conceptual domains (see final exam/pre-test pairwise comparisons in Table 3).

‡‡ PhET Coulomb's Law simulation. From https://phet.colorado.edu/en/simulation/coulombs-law.

Table 3 A Wilcoxon signed-rank test (Z) was conducted to evaluate the average differences of two test timepoints for each learning cycle activity. The effect size (r) for each pairwise comparison was also calculated to determine the effect of the activity between the two timepoints

Activity	Pair-wise score comparison	Z	p (2-tailed)[a]	r[b]
Nuclear Atom	Post-test/pre-test	8.597	<0.001	0.44
	Final exam/pre-test	10.296	<0.001	0.52
	Final exam/post-test	4.101	<0.001	0.21
Coulomb's Law	Post-test/pre-test	3.488	<0.001	0.18
	Final exam/pre-test	7.804	<0.001	0.40
	Final exam/post-test	4.767	<0.001	0.24
Atomic Interaction	Post-test/pre-test	11.402	<0.001	0.58
	Final exam/pre-test	11.708	<0.001	0.59
	Final exam/post-test	1.762	0.078	0.09
Molecular Shape	Post-test/pre-test	10.222	<0.001	0.52
	Final exam/pre-test	8.131	<0.001	0.41
	Final exam/post-test	5.269	<0.001	0.27
Molecular Dipole	Post-test/pre-test	8.972	<0.001	0.46
	Final exam/pre-test	8.932	<0.001	0.45
	Final exam/post-test	0.340	0.734	0.02

[a] Significance level adjusted using a Bonferroni correction (α: 0.05/3 = 0.017). [b] The interpretation guidelines for r are as follows: 0.1 (small effect), 0.3 (medium effect), and above 0.5 (large effect).

Elucidating persistent inaccurate thinking

To address the third research question that asked how analyses of the student responses might inform changes for future implementations of the activities, Sankey diagrams were used to identify the common inaccurate ideas that appeared to persist after students engaged in the learning cycle interventions. Sankey diagrams that depict which responses were selected for each time point of testing were also created for the test items in which more than 10% of the students selected incorrect answers on the post-test and/or final exam (see Fig. 3 and Appendix 5 Fig. 1, ESI†). It should be noted that the Sankey diagrams only include the students who completed all three tests (90% or more of the study sample, depending on the test). Tracking which item distractors were selected at each phase of testing provided an opportunity to identify which specific inaccurate ideas generally persisted among the participants. Four test items from the Molecular Shape and Molecular Dipole activities were observed to reveal the most notable areas of inaccurate thinking (see Fig. 3 and Table 4).

In Question 2 of the Molecular Shape test, the desired conceptual learning outcome is that an unbonded electron pair gives rise to the strongest repulsive force to a nearby bonded electron pair. For the question, "*Which would you predict to have the strongest repulsive force to a nearby electron pair?*", the most common distractor in the pre-test was the answer "a bonding electron pair." This was likely because students recognized electrostatic repulsion occurs between bonded atoms, but they neglected to consider the effect of non-bonding electrons on the bond angle between the bonded atoms. Though the number of students who chose this answer decreased from 69 to 12 from

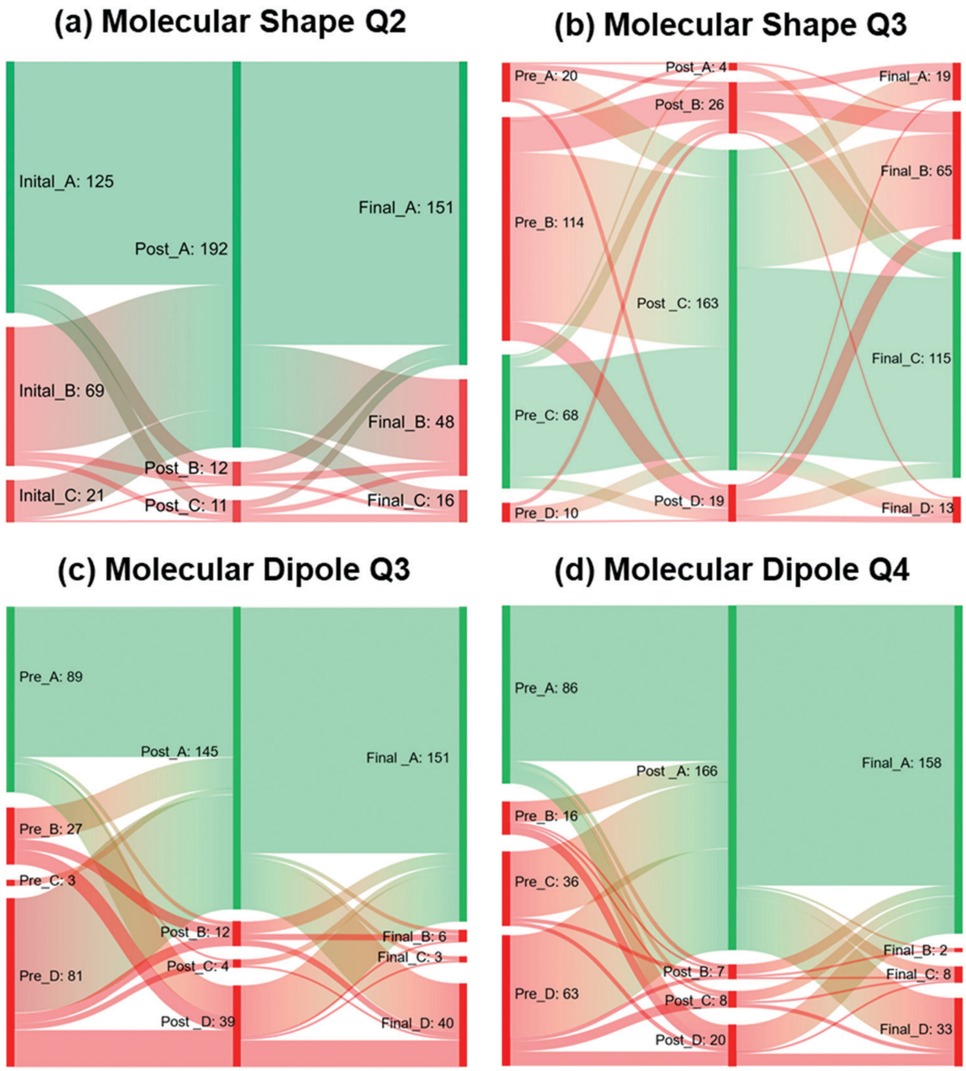


Fig. 3 Three-node Sankey diagrams display the distributions of students correctly (green) or incorrectly (red) answer the question along with the answer choices at three time-points: pre-pre-test (left node), post-post-test (middle node), and the final Exam (right node).

pre- to post-test administration, 48 students reverted back to the "a bonding electron pair" distractor on the final exam (see Fig. 3a). Question 3 of the Molecular Shape activity also evaluated students' understanding of the impact of non-bonding electron pairs on molecular shape. The most common distractor in the pre-test for this item was the answer "CH_4," likely because students associated smaller bond angles with molecules in which more atoms were bonded to the central atom. The number of students who chose this answer decreased from 114 to 26 from pre- to post-test administration, yet 65 students again chose this common distractor on the final exam (Fig. 3b). Overall, most students appeared to make knowledge gains with respect to identifying the impact of non-bonding electron pairs on the electron pair geometries and shapes of molecules. However, given the fact over 10% of the students reverted back to the inaccurate thinking for these two test items as described above, practitioners might consider refining the Molecular Shapes activity to mitigate this reversion in future implementations. Because some students appear to neglect the impact of the non-bonding electron pairs, the activity could guide students to draw out the molecules with all bonding atoms and non-bonding electro pairs (this could be added into Table 2 of the Molecular Shapes activity§§). If students were asked to draw the molecules in this way, and also to use arrows to identify where electrostatic repulsion occurs within the molecule, this might lead to longer lasting knowledge gains for this particular concept.

In Question 3 of the Molecular Dipole activity, the desired learning outcome was for students to understand the relationship between molecular structure and molecular dipoles. In this particular question, students were required to choose the response that explains why water molecules are polar and carbon dioxide molecules are non-polar. Although the number of students who chose the most common distractor decreased

§§ PhET Molecular Shapes simulation. From https://phet.colorado.edu/en/simulation/molecule-shapes.

Table 4 Four questions from Molecular Shape (MS) and Molecular Dipole (MD) activities had the highest incorrect responses illuminating the common areas of inaccurate thinking among students

MS Q2. Which would you predict to have the strongest repulsive force to a nearby electron pair?
A. A non-bonding electron pair
B. A bonding electron pair
C. These have the same repulsive force

MS Q3. Which would you expect to have the smallest bond angle?
A. NH_3
B. CH_4
C. H_2O
D. These would all have the same bond angle.

MD Q3. Which explains why water molecules are polar and carbon dioxide molecules are non-polar?
A. Water molecules have a non-linear molecular shape
B. Carbon dioxide has non-polar bonds
C. Both molecules have polar bonds
D. Hydrogen and oxygen have a large electronegativity difference

MD Q4. Which molecule has a net molecular dipole?
A. OCl_2
B. Cl_2
C. CO_2
D. BH_3

from 81 to 39 from pre- to post-test administration, 40 students still answered incorrectly on the final, and over 10% of students reverted from the correct answer on the post-test to the most common distractor on the final exam. The most common distractor stated "hydrogen and oxygen have a large electronegativity difference." Although this statement is accurate, it does not explain why water is polar molecule and carbon dioxide is not. The students who chose this answer likely only considered the electronegativity difference between the atoms in the bonds in identifying molecular dipoles, without considering how the three-dimensional structure plays a role in giving rise to molecular polarity. For Question 4 in the Molecular Dipoles activity, the number of students who chose the most common distractor decreased from 63 to 20 from the pre- to post-test administration. Though only 33 students chose this distractor on the final exam, over 10% of students who chose this answer on the final reverted back to this common distractor after answering correctly on the post test. Because the most common distractor for this question was "BH_3," we speculate students who chose this answer did so based on comparing its structural formula to the other answer choices, rather than carefully evaluating differences in electronegativity and/or molecular geometry (BH_3 was unique among the answer choices in having three atoms bonded to a central atom; see Appendix 4, ESI†). The critical thinking questions in the Molecular Dipole activity explicitly ask students to identify what properties are required for a molecule to possess a net dipole.¶¶ Therefore the observation that over 10% of students continued to incorrectly identify which molecules have net dipoles and/or why net dipoles are present was surprising. In future implementations of the activity, instructors might consider including a post-activity discussion highlighting that the molecular formula should not be used as a heuristic to identify the presence of net molecular dipoles, and that simply determining the presence of polar bonds is not sufficient in identifying molecular dipoles. Instructors might also consider having students draw out the net dipole vectors in addition to the electron density maps, which could help reinforce how both electronegativity differences and molecular shape determine if a net molecular dipole is present.

Discussion

Knowledge gains and the constructivist theory

The pre-post-post test results suggest the incorporation of introductory, in-person activities built upon the constructivist theory of learning within a flipped classroom structure can be successfully implemented in an existing flipped classroom structure. The one-group quasi-experimental design does not allow for the explicit determination of cause and effect between the specific learning interventions and observed knowledge gains, however we can speculate about how the classroom activities described herein may have impacted the performance on the post-test and final exam test items. Because the post-test was administered after the in-class learning cycle concept development activities, the preliminary knowledge gains observed on the post-test for all five conceptual domains are likely attributed to the introductory concept development activities (see post-test/pre-test pairwise comparisons in Table 3). After the introductory learning cycle activities, students completed flipped modules and/or engaged in whole-class questioning/think-pair-share exercises that were intended to fulfill the application phase of the learning cycle. Thus, the observed gains on the final exam test items are possibly linked to the overall combination of introductory concept development activities and flipped module/in-class application exercises (see final exam/pre-test pairwise comparisons in Table 3). Even though the post-test/pre-test and final exam/pre-test pairwise comparisons were statistically significant for all conceptual domains, the effect sizes and final exam/post-test pairwise comparisons suggest additional growth may have occurred after the introductory concept development learning cycle activity for three of the five conceptual domains (see Table 3). The time between completing the learning cycle activities and the final exam ranged from nine weeks to three weeks. This difference in time lag may have resulted in different levels of long-term knowledge retention for the different activities. However, the fact at least three weeks passed between the learning cycle activities and final exam suggests the application of the concepts in the flipped modules and in-class exercises led to notable retention of knowledge gains.

These observed knowledge gains can be viewed within the context of the constructivist theory of learning. As described in the introduction, Vygotsky's theory of constructivism posits that understanding comes from making meaningful connections

¶¶ PhET Molecular Dipole simulation. From https://phet.colorado.edu/en/simulation/molecule-polarity.

through experience between prior knowledge and new knowledge (Vygotsky, 1978). The in-class learning cycle activities allowed students to explore the simulations in a guided fashion, and they constructed informal and general ideas about the subject individually built upon the foundation of previous learning. Vygotsky's theory states that language plays a fundamental role in shaping meanings. Because students wrote down their observations and personal interpretation of the simulations, and then subsequently engaged in group discussions to further extend their concept development, learning was supported by social constructivist principles. Additionally, a cycle of exploration, concept development, and concept application was created by combining the introductory learning cycle activities with the flipped modules/in-class exercises. This suggests the overall knowledge gains observed for the five conceptual domains can also be rooted in Von Glasersfeld's view of constructivism (Von Glasersfeld, 2001). Students were given the opportunity to construct new conceptual knowledge from experience (the exploration and concept development phases in the introductory learning cycle activity), then apply this new knowledge in the online flipped module exercises and in-class collaborative group learning exercises. The totality of this learning experience appears to have to led to both short term and longer-term conceptual understanding for students.

Limitations of the study

Because this design-based study did not include a ''teaching as usual'' control group comparison, this might be viewed as a research limitation. However, previous research has provided compelling evidence about the positive impact of learning cycle activities on student performance gains and demonstrated the general efficacy of learning cycle activities (Musheno and Lawson, 1999; Escalada *et al.*, 2004; Rodriguez *et al.*, 2020), therefore, creating a teaching as usual quasi-experimental control group that purposefully excluded the learning cycle activities was not necessary or prudent.

The short duration of the tests for each activity and the fact these measures of student knowledge were only administered to one sample may have limited the rigor of the reliability analysis; however, the item-total correlations and the apparent ability of the tests to detect gains in knowledge suggest the tests were adequate measures of student knowledge. It is also noted the pre-test performance did not impact the students' grades, whereas the post-test questions counted as extra credit and the assessment questions were integrated as normal questions within the final exam. Because the students did not have the same grade incentive on the pre-test, this could have biased the improved performance on the post-test and final exam. Unfortunately, this limitation is difficult to overcome in a class-based quasi-experimental design, as it is generally not desired to have the pre-test scores impact student grades. The same set of items of were administered three times. Although this lends reliability to the results, it could limit the validity in measuring student knowledge (*i.e.*, students could just be remembering the question and answer). However, this validity limitation is minimized because students were not provided with the answers nor were alerted that the same items would be on subsequent tests. Finally, the test questions were designed by the authors to evaluate conceptual knowledge and determine if students continued to adopt previously observed inaccurate chemical ideas. Despite being designed to probe conceptual understanding, the multiple-choice format of the questions may have limited the ability to fully capture the products of student sense-making and conceptual thinking. Future studies could involve creating items that include some free response questions and/or other methods that more effectively elicit and document student sense-making and conceptual understanding.

It is uncertain that the success of the learning outcomes in this study can be translated to other institutional settings. With the challenges of implementing the activities in a large enrollment course that had a high percentage of first-generation students, significant and positive learning outcomes were observed for each learning cycle activity in the present study. This suggests the inferential statistical claims made regarding performance gains in the present study should generalize to other undergraduate introductory/general chemistry courses. Though the activities need to be implemented in other institutions with different student populations to confirm this notion, the fact the performance gains reported here for the Atomic Interactions activity appear to be comparable to those reported by Minshall and Yezierski (Minshall and Yezierski, 2021) further suggest the learning cycle activities can be successfully implemented in institutional settings with disparate student populations.

With respect to the implementation of the PhET-based learning cycle activities, access to devices could be a potential limitation at some institutions. If laptop or tablet loan programs are not available, instructors could modify the implementation to allow students to share devices. If enough students have access to devices such that at least one student per group can run the PhET simulation, this should allow instructors to facilitate the learning cycle activities. If access to laptops or tablets is even more limited, instructors might consider running the simulation as a demonstration on the classroom projection system, and students could then make the necessary observations and progress through the scaffolded questions.

Conclusions and implications

Adopting the learning cycle activities allowed for students to actively engage in concept development and participate in the skill development within a flipped classroom environment. The learning cycle activities were seamlessly integrated into an existing course that included flipped classroom modules. The present study demonstrated flipped classroom modules coupled with in-class learning cycle activities that incorporated PhET simulations have increased the immediate and longer-term performance gains in a large classroom setting. More specifically, it appears the introductory learning cycle activities led to immediate increases in conceptual knowledge for all five categories of learning objectives, and the application of these concepts in subsequent flipped modules and/or in-class activities with whole-class questioning and think-pair-share collaborative group learning seems to have allowed students to retain or build upon that conceptual knowledge through the end of the course. Although more students appeared to struggle with selected concepts in the Molecular Shape and Molecular Dipole learning objectives, future implementations of the activities can be modified to more explicitly highlight the concepts in these two activities for which students persisted to demonstrate inaccurate thinking. This design research approach should not only yield improved performance outcomes for the activities described herein, but it also holds promise to act as a model for a scholarly approach to high-quality materials development more broadly in chemistry education.

Using simulations as a means for students to generate observations in learning cycle activities within a non-laboratory classroom setting is a promising approach to integrate concept development activities in higher education introductory chemistry courses. It is hoped the current study, in conjunction with the previous report from Minshall & Yezierski (Minshall and Yezierski, 2021), demonstrates how students can improve conceptual understanding when the learning cycle approach is used with simulation-based activities and will inspire instructors from a variety of institutional settings to engage their students in this approach. With respect to the use of the flipped classroom in higher education chemistry, educational studies generally provide only a vague description of coupling online video tutorials with some form of in-class active learning. To our knowledge no previous reports describe including an in-class concept development activity prior to the cycle of online/in-person learning. We hope the current study gives classroom practitioners a roadmap for how concept development activities can be incorporated into a flipped classroom structure, as well as a more complete picture of what is actually involved in the pre-class online learning and in-class active learning environments that produced the performance gains described herein. If instructors were to put a twist on the flip as described here, they would be creating a blended learning environment that actively engages students in a new way and likely helps students achieve understanding of core general chemistry concepts.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge the University of Colorado-Boulder PhET Interactive Simulations resource site for creating these valuable open-access instructional tools. The authors also acknowledge YuChen Xiao for helping with the facilitation of the in-class activities and the general chemistry students who participated in our study.

References

Boo H. K., (1998), Students' understandings of chemical bonds and the energetics of chemical reactions, *J. Res. Sci. Teach.*, **35**(5), 569–581.

Canelas D. A., Hill J. L. and Novicki A., (2017), Cooperative learning in organic chemistry increases student assessment of learning gains in key transferable skills, *Chem. Educ. Res. Pract.*, **18**(3), 441–456.

Casselman M. D., Atit K., Henbest G., Guregyan, C., Mortezaei K. and Eichler J. F., (2019), Dissecting the flipped classroom: Using a randomized controlled trial experiment to determine when student learning occurs, *J. Chem. Educ.*, **97**(1), 27–35.

Cobb P., Confrey J., DiSessa A., Lehrer R. and Schauble L., (2003), Design experiments in educational research, *Educational Researcher*, **32**(1), 9–13.

Cohen J., (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn, New York, NY: Lawrence Erlbaum Associates, chap. 3.

D'Angelo C., Rutstein D., Harris C., *et al.*, *Simulations for STEM learning: systematic review and meta-analysis*, Menlo Park, CA: SRI International; 2014.

Escalada L. T., Rebello N. S. and Zollman D. A., (2004), Student explorations of quantum effects in LEDs and luminescent devices, *Physics Teacher*, **42**(3), 173–179.

Flynn A. B., (2015), Structure and evaluation of flipped chemistry courses: organic & spectroscopy, large and small, first to third year, English and French, *Chem. Educ. Res. Pract.*, **16** (2), 198–211.

Hibbard L., Sung S. and Wells B., (2016), Examining the effectiveness of a semi-self-paced flipped learning format in a college general chemistry sequence, *J. Chem. Education*, **93**(1), 24–30.

Holme T. and Murphy K., (2012), The ACS Exams Institute undergraduate chemistry anchoring concepts content map I: General Chemistry, *J. Chem. Education*, **89**(6), 721–723.

Jensen J. L., Kummer T. A. and Godoy P. D. D. M., (2015), Improvements from a flipped classroom may simply be the fruits of active learning, *CBE—Life Sciences Education*, **14**, 1–14.

Lancaster K., Moore E. B., Parson R. and Perkins K. K., (2013), Insights from using PhET's design principles for interactive chemistry simulations, *ACS Symposium Series*, **1142**, 97–126.

Lawson A. E. and Karplus R., (2002), The learning cycle, *A love of discovery*, Dordrecht: Springer, pp. 51–76.

Liu Y.; Raker J. R. and Lewis J. E., (2018), Evaluating student motivation in organic chemistry courses: moving from a lecture-based to a flipped approach with peer-led team learning, *Chem. Educ. Res. Pract.*, **19**(1), 251–264.

Minshall B. L. and Yezierski E. J., (2021), Data-driven activity reform: employing design research to improve scaffolding and concept development, *Chem. Educ. Res. Pract.*, **22**(1), 136–145.

Moog R. and Farrell J., (2002), *Chemistry: a guided inquiry*, NY: John Wiley & Sons.

Moog R. S. and Spencer J. N. (ed.), (2008), *Process oriented guided inquiry learning*, Washington, DC: American Chemical Society, **vol. 994**.

Musheno B. V. and Lawson A. E., (1999), Effects of learning cycle and traditional text on comprehension of science concepts by students at differing reasoning levels, *J. Res. Sci. Teaching*, **36**(1), 23–37.

Naibert N., Geye E., Phillips M. M. and Barbera J., (2020), Multicourse Comparative Study of the Core Aspects for Flipped Learning: Investigating In-Class Structure and Student Use of Video Resources, *J. Chem. Educ.*, **97**(10), 3490–3505.

Rahman M. T. and Lewis S. E., (2019), Evaluating the evidence based instructional practices in chemistry through *meta*-analysis, *J. Res. Sci. Teach.*, **57**, 765–793.

Rodriguez J. M. G., Hunter K. H., Scharlott L. J. and Becker N. M., (2020), A review of research on process oriented guided inquiry learning: Implications for research and practice, *J. Chem. Educ.*, **97**(10), 3506–3520.

Seery M. K., (2015), Flipped learning in higher education chemistry: emerging trends and potential directions, *Chem. Educ. Res. Pract.*, **16**(4), 758–768.

Sweller J., (1988), Cognitive load during problem solving: Effects on learning, *Cognitive science*, **12**(2), 257–285.

Tavakol M. and Dennick R., (2011), Making sense of Cronbach's alpha, *JIAMSE*, **2**, 53–55.

Tomczak M. and Tomczak E., (2014), The need to report effect size estimates revisited. An overview of some recommended measures of effect size, *Trends in Sport Sciences*, **21**(1), 19–25.

Von Glasersfeld E., (2001), The radical constructivist view of science, *Foundations of Science*, **6**(1), 31–43.

Vygotsky L., (1978), *Mind in Society*, London: Harvard University Press.

Wang F. and Hannafin M. J., (2005), Design-based research and technology-enhanced learning environments, *Educ. Technol. Res. Develop.*, **53**(4), 5–23.

Weaver G. C. and Sturtevant H. G., (2015), Design, implementation, and evaluation of a flipped format general chemistry course, *J. Chem. Educ.*, **92**(9), 1437–1448.

(No existing text)		**Chemistry Education Designated Emphasis** College of Natural and Agricultural Sciences Jack Eichler (Chemistry), Director jack.eichler@ucr.edu **Advisory Committee & Participating Faculty** Ana Bahamonde (Chemistry) Christopher J. Bardeen (Chemistry) Ludwig Bartels (Chemistry) Gregory J.O. Beran (Chemistry) Matthew Casselman (Chemistry) Chia-En A. Chang (Chemistry) Quan “Jason” Cheng (Chemistry) Matthew Conley (Chemistry) James Davies (Chemistry) Pingyun Feng (Chemistry) Boniface Fokwa (Chemistry) Joseph Genereux (Chemistry) W. Hill Harman (Chemistry) Richard Hooley (Chemistry) De-en Jiang (Chemistry) Ryan Julian (Chemistry) Kevin Kou (Chemistry) Catharine Larsen (Chemistry) Vincent Lavallo (Chemistry) Leonard J. Mueller (Chemistry) Michael Pirrung (Chemistry) Richard Schrock (Chemistry) Christopher Y. Switzer (Chemistry) Timothy Su (Chemistry) Kathryn Uhrich (Chemistry) Yinsheng Wang (Chemistry) Min Xue (Chemistry) Yadong Yin (Chemistry) Francisco Zaera (Chemistry) Haofei Zhang (Chemistry) Jingsong Zhang (Chemistry) Wenwan Zhong (Chemistry) Linlin Zhao (Chemistry) **Designated Emphasis Requirements** The Designated Emphasis in Chemistry Education is a course of study intended to give students a background in designing and carrying out and educational research study in the context of higher education chemistry teaching and learning. The program is optional and the courses used for the DE may not be counted toward	

		MS or PhD requirements. The Designated Emphasis in Chemistry Education has two core requirements: 1. Completion of three (3) graduate courses (12 units), with one (1) graduate course being in Chemistry Education and two (2) graduate courses in Psychology: a. *CHEM 241* (4 units); b. *PSYC 211* (4 units); and c. either *PSYC 212 (4 units) or PSYC 207C (3 units).* 2. Completion of a chemistry education research project anytime after the student advances to candidacy. The scope of this project is expected to be such that it comprises one chapter in the final PhD dissertation. This project may be completed under the guidance of a chemistry department professor of teaching (who would act as a co-advisor overseeing the chemistry education research), or could be completed independently under the approval of the Chemistry Education DE Program Committee and the student's research advisor. *CHEM 299* (4 units).	

Daniel Ozer, Ph.D.
Professor and Chair
DEPARTMENT OF PSYCHOLOGY
RIVERSIDE, CA 92521-0128
daniel.ozer@ucr.edu

Jack F. Eichler, PhD
Professor of Teaching in Chemistry
Department of Chemistry
University of California, Riverside
Riverside, CA 92521

Dear Professor Eichler,

I have had opportunity to review your request to list Psychology courses (211, 212, and 207C) as satisfying a requirement in the proposed Designated Emphasis in Chemistry Education. Contingent on instructor approval (as per the Catalogue description of each of these courses) and enrollment limits, the Psychology Department would welcome your students to out courses.

Sincerely,

Daniel Ozer
Professor & Chair