frontiers | Research Topics

Organic chemistry education research into practice

Edited by
Jay Wackerly, Sarah Zingales, Michael Wentzel,
Gautam Bhattacharyya and Brett McCollum

Published in
Frontiers in Education

ISSN 1664-8714
ISBN 978-2-8325-6166-9
DOI 10.3389/978-2-8325-6166-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Organic chemistry education research into practice

Topic editors
Jay Wackerly — Central College, United States
Sarah Zingales — United States Coast Guard Academy, United States
Michael Wentzel — Augsburg University, United States
Gautam Bhattacharyya — Missouri State University, United States
Brett McCollum — Thompson Rivers University, Canada

Citation
Wackerly, J., Zingales, S., Wentzel, M., Bhattacharyya, G., McCollum, B., eds. (2025). *Organic chemistry education research into practice*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6166-9

Table of contents

TYPE Editorial
PUBLISHED 11 March 2025
DOI 10.3389/feduc.2025.1578608

OPEN ACCESS

EDITED AND REVIEWED BY
Lianghuo Fan,
University of Macau, Macao SAR, China

*CORRESPONDENCE
Jay Wm. Wackerly
✉ wackerlyj@central.edu

RECEIVED 18 February 2025
ACCEPTED 20 February 2025
PUBLISHED 11 March 2025

CITATION
Wackerly JW, Zingales SK, Wentzel MT, McCollum BM and Bhattacharyya G (2025) Editorial: Organic chemistry education research into practice. *Front. Educ.* 10:1578608. doi: 10.3389/feduc.2025.1578608

Editorial: Organic chemistry education research into practice

Jay Wm. Wackerly[1]*, Sarah K. Zingales[2], Michael T. Wentzel[3], Brett M. McCollum[4] and Gautam Bhattacharyya[5]

[1]Chemistry Program, Department of Natural Science, Central College, Pella, IA, United States, [2]Department of Chemical and Environmental Sciences, United States Coast Guard Academy, New London, CT, United States, [3]Department of Chemistry, Augsburg University, Minneapolis, MN, United States, [4]Centre for Excellence in Learning and Teaching, Thompson Rivers University, Kamloops, BC, Canada, [5]Chemistry and Biochemistry Department, Missouri State University, Springfield, MO, United States

KEYWORDS

organic chemistry, undergraduate curriculum, chemistry education research (CER), practice, learning, instruction

Editorial on the Research Topic
Organic chemistry education research into practice

Scholarship in chemical education has grown in remarkable ways over the past century. What started with a focus in primary and secondary education soon spread to first-year courses in higher education. By 2008 there was enough research on topics beyond the introductory undergraduate level, that the Royal Society of Chemistry's *Chemistry Education Research and Practice* dedicated a Research Topic to "advanced courses" (Bodner and Weaver, 2008). Along with the expansion into all levels of education, scholarship in chemical education has advanced to the point that journals devoted to chemical education are now dominated by theory-grounded research studies using quantitative, qualitative, and mixed methodologies (Cooper and Stowe, 2018)!

Ostensibly, chemical educators worldwide use the resulting bodies of research to inform, and reform, their instruction. However, these innovations are infrequently reported because of the absence of peer-reviewed journals in which the associated scholarship can be published (Sweder et al., 2023). We are excited to provide this forum for presenting evidence-based instructional practices in organic chemistry. To use an analogy from organic synthesis, CER articles are equivalent to methodology articles; and evidence-based practice articles—like the ones in this special issue—are like total syntheses. Just as we recognize the importance of total syntheses in showcasing and extending/refining respective methodologies, the articles in this issue, similarly, expand the knowledge base of CER and are clearly a valued form of scholarship in chemical education.

The contributions to this special issue of *Frontiers in Education* share several key attributes. First, each group of authors designed learning experiences that are grounded in research literature and/or theoretical frameworks from social sciences and philosophy. Second, the articles include detailed descriptions of the context in and methods by which the authors implemented their developed learning materials. Third, the authors demonstrate the efficacy of their evidence-based course innovations. Critically, all the presented data in this issue are consistent with one or more levels in St. John and McNeal's (2017) strength of evidence pyramid.

This Research Topic contains 12 articles, divided into three themes: (1) generally-applicable instructional strategies; (2) imaginative repurposing of instructional agents and virtual platforms; and (3) innovative approaches to assessment. In the brief descriptions of the contributions in the following paragraphs, we use one of the following abbreviations after the authors' names to designate the *Frontiers* manuscript category to which the article belongs: Curriculum, Instruction, and Pedagogy (CIP), Hypothesis and Theory (HT), Original Research (OR), Perspective (P), or Review (R).

Each of the contributions to the first theme, generally-applicable instructional strategies, presents concepts that are applicable to teaching across the spectrum of topics in organic chemistry. Popova (P) describes how research-practitioner partnerships can be used to create more effective course materials and, therefore, pedagogical implementation of research findings. Using representational competence as an example, the author explains how one such partnership was used to explicitly address an area of learner skill development that is often left implicit. MacNeil et al. (CIP) follow with a report on instruction in metacognition delivered concomitantly with course content. Using seminal works from cognitive and educational psychology, the authors developed a combination of learning task inventories, confidence self-assessments, and performance predictions and post-dictions. They found that learners improved their ability to engage cognitive processes involving planning, monitoring, and evaluating knowledge acquisition. Wackerly et al. (HT) then propose that abductive reasoning skills, essential in scientific problem-solving and medical diagnosis, are crucial for career interests of students that present in the 2^{nd}-year undergraduate course. The authors provide examples of how instructors can integrate abductive reasoning into their teaching and, thereby, enhance students' problem-solving abilities. Concluding this section, Graulich and Lieber (R) assert that effective chemistry learning requires engaging students in meaningful tasks that go beyond rote exercises. They explain that contrasting case comparisons are meaningful because they tend to induce students to use multiple cognitive operations simultaneously, which helps in their overall problem-solving ability.

The authors for the second theme, imaginative repurposing of instructional agents and virtual platforms, meticulously describe their adaptations and successful creation or adaptation of instructional methodologies for virtual and in-person learning. Schuessler et al. (OR) present their conversion of assessment tasks from pencil-and-paper formats into a digital ones. In their multi-institutional study, the authors demonstrate how these types of transitions need to be carefully and purposefully executed. Using cognitive load theory, the research team used several cycles of implementation and feedback to identify and minimize extraneous cognitive load resulting from the change in medium. Griffin et al. (CIP) describe their use of chemical education and peer-learning research literature to simultaneously design a new lab curriculum alongside a new Learning Assistant (LA) program in which undergraduate students worked with the graduate teaching assistants (GTAs). The authors discuss how interactions with LAs positively impacted several affective factors for students in non-majors courses. Additionally, the students found LAs to be especially helpful when their GTAs were working with other students. Ward et al. (CIP), explore how an augmented reality (AR) app, H NMR MoleculAR, helps students understand proton NMR in organic chemistry labs. The study highlights the challenges and benefits of using AR tools in different learning environments. In the final article of this section, Gallardo-Williams and Dunnagan (P) present their use of extended reality to address factors related to access to instructors during introductory-level organic chemistry labs. Initially developed for virtual instruction, the authors provide a research-based methodology for fostering constructive and thoughtful interactions between students and their lab instructors in research-focused institutions.

The articles in the final theme, innovative approaches to classroom assessment, offer compelling evidence demonstrating the potential of non-standard methods of assessment. Mio (CIP) reviews alternative grading methods, such as "ungrading" and standards-based assessments, and describes how these can reduce students' stress and anxiety while improving their metacognition. Gaines and Burrows (CIP) implemented oral examinations in two different classrooms during the disruption in educational settings caused by the pandemic. They found that oral exams allowed students and instructors to collaboratively identify strengths and weaknesses. Moster and Zingales (CIP) describe specifications-based grading in an online graduate organic chemistry course, wherein students earned grades by meeting specific learning objectives rather than accumulating points. The flexible system allowed students to choose assessments, work at their own pace, and use tokens for extensions or retakes, leading to more content-focused interactions and a slight increase in pass rates. This Research Topic concludes with Ferguson and Bonner (P), who share their perspective on "ungrading" across the curriculum and how they implement it in their organic chemistry courses. Like Mio, they propose "ungrading" as a promising strategy for increasing student metacognition.

Above all, we would like to thank the more than 30 authors who contributed to this Research Topic. The authors afford readers unique opportunities to learn about new and effective instructional strategies, some of which may have been previously unknown. Furthermore, several manuscripts demonstrate how creative adaptation of existing resources can lead to ground-breaking change.

As co-Editors, we recognize that a single Research Topic, cannot comprehensively alter the landscape of teaching and learning in organic chemistry. Rather than being definitive or prescriptive, our main hope is that this issue will stimulate healthy debates in the global chemical education community about ways to improve the student experience. Though we may have differences in approaches and proposed remedies, as instructors of organic chemistry we can certainly agree that there is room for improvement.

Finally, we strongly feel that the contributed articles demonstrate the immense value of practice-focused, evidence-based scholarship in chemical education, and the clear need for more venues to publish articles like the ones in this issue. In fact, the American Chemical Society Statement on Scholarship (American Chemical Society Committee on Education (SOCED), 2010) exhorted, *"the chemistry community* [to] *accept and act upon* *a broader definition rewarding faculty for the wide range of activities needed to bring about a modern and effective research and education infrastructure.*" To that end, journals need to establish clear and consistent guidelines for evidence of instructional efficacy that do not mandate research studies. We hope that the readers will join us in advocating for these future opportunities.

Author contributions

JW: Writing – original draft, Writing – review & editing. SZ: Writing – original draft, Writing – review & editing. MW: Writing – original draft, Writing – review & editing. BM: Writing – original draft, Writing – review & editing. GB: Conceptualization, Writing – original draft, Writing – review & editing.

Acknowledgments

We would like to express our heartfelt gratitude to the authors, for sharing their innovative work, the reviewers for their thoughtful evaluations of the manuscripts, and the *Journal's* editorial staff for making this Research Topic a reality.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

American Chemical Society Committee on Education (SOCED) (2010). *Statement on scholarship*. Available at: https://www.acs.org/content/dam/acsorg/about/governance/committees/education/statement-on-scholarship.pdf (accessed December 13, 2024).

Bodner, G. M., and Weaver, G. (2008). Introduction: Research and practice in chemical education in advanced courses. *Chem. Educ. Res. Pract.* 9, 81–83. doi: 10.1039/B806596A

Cooper, M. M., and Stowe, R. L. (2018). Chemistry education research – From personal empiricism to evidence, theory, and informed practice. *Chem. Rev.* 118, 6053–6087. doi: 10.1021/acs.chemrev.8b00020

St. John, K., and McNeal, K. S. (2017). Strength of evidence pyramid: One approach for characterizing the strength of evidence of geoscience education research (GER) community claims. *J. Geosci. Educ.* 65, 363–372. doi: 10.5408/17-264.1

Sweder, R. D., Herrington, D. G., and Crandell, O. M. (2023). Chemistry education research at a crossroads: Where do we need to go now? *J. Chem. Educ.* 100, 1710–1715. doi: 10.1021/acs.jchemed.3c00091

frontiers | Frontiers in Education

TYPE Curriculum, Instruction, and Pedagogy
PUBLISHED 26 March 2024
DOI 10.3389/feduc.2024.1367087

OPEN ACCESS

EDITED BY
Brett McCollum,
Thompson Rivers University, Canada

REVIEWED BY
Jess Allingham,
Thompson Rivers University, Canada
John Canal,
Simon Fraser University, Canada
Fun Man Fung,
National University of Singapore, Singapore

*CORRESPONDENCE
Renée D. Link
rlink@uci.edu

RECEIVED 08 January 2024
ACCEPTED 06 March 2024
PUBLISHED 26 March 2024

CITATION
Griffin JH, Thompson JC, López PA and Link RD (2024) Leveraging undergraduate learning assistants when implementing new laboratory curricula.
Front. Educ. 9:1367087.
doi: 10.3389/feduc.2024.1367087

Leveraging undergraduate learning assistants when implementing new laboratory curricula

James H. Griffin, Jordan C. Thompson, Pía A. López and Renée D. Link*

Department of Chemistry, University of California, Irvine, Irvine, CA, United States

At University of California, Irvine, a large-enrollment research university, undergraduate chemistry courses for non-chemistry majors were delivered remotely during the 2020–2021 academic year, with a return to in-person instruction planned for January 2022. Because this return to in-person instruction coincided with the transition of second-year students from general chemistry to organic chemistry laboratory courses, the instructional staff recognized a need for remedial laboratory curricula for students with no prior in-person laboratory experience. Simultaneously, we desired to implement undergraduate Learning Assistants (LAs) in non-chemistry major organic chemistry laboratories for the first time at our university. In this paper, we describe our approach for leveraging undergraduate LAs to (1) test new laboratory curricula and (2) address feelings of comfort and safety for students with no prior in-person laboratory experience. Benefits of our LA program perceived by students include increased laboratory efficiency and improved student learning from near-peer instructors; benefits perceived by LAs include the development of professional skills and teamwork with graduate student teaching assistants. We provide an outline of resources and strategies to enable instructors to simultaneously implement undergraduate LAs and new laboratory curricula.

KEYWORDS

learning assistants, organic chemistry laboratory, near-peer assisted learning, affective domain in science education, student-centered learning, collaborative learning

1 Introduction

Learning Assistants (LAs) — undergraduate students who serve as assistant instructors for courses they have previously taken — are increasingly being leveraged in STEM courses (Emenike et al., 2020; Barrasso and Spilios, 2021). Peer learning, in which experienced students guide current students' learning through a zone of proximal development, underpins LA programs (Thompson et al., 2020). In traditional learning communities, students learn solely from a senior instructor, who can be perceived as unapproachable or intimidating. Vertical learning communities seek to address this student-instructor gap by introducing a near-peer instructor (Bourne et al., 2021). LAs are approachable due to closeness in age and experience and can provide mentorship to students. Additionally, LAs improve students' comfort and confidence in the classroom (Ten Cate and Durning, 2007). Student comfort and confidence fall under the affective domain, which — along with the cognitive and psychomotor domains — is an essential part of meaningful learning in the laboratory (Bretz, 2001; Galloway et al., 2016). Therefore, peer instructors have the potential to foster an improved overall learning environment.

The COVID-19 pandemic and subsequent statewide shutdown in 2020 resulted in evacuation of our campus and remote learning that lasted until December 2021 (Lawhon, 2020). During the Spring 2021 term, the organic chemistry laboratory (OCL) course sequence for non-chemistry majors piloted an LA program to facilitate student guidance during a time of isolation and uncertainty.

Returning to in-person learning presented unique challenges for the OCL instructional team. Under usual circumstances, students would have completed two prior in-person general chemistry laboratory courses. Due to remote learning, both students entering the OCL series and pilot program LAs who completed OCL courses online would have little to no hands-on chemistry laboratory experience. Fall 2021 was an opportunity for the simultaneous testing of new experiments and training of LAs with little prior in-person experience before a complete return to in-person teaching in Winter 2022. Herein, we outline a strategy by which we simultaneously tested remedial laboratory curricula designed for this unique cohort of students and trained an initial group of LAs to assist with teaching these modified courses. This strategy was informed by the existing literature on LAs, vertical learning communities, and addressing the affective domain in laboratory instruction. Our overarching research question was: "Does the implementation of LAs in response to instructional discontinuity lead to beneficial outcomes for a cohort of students lacking prior in-person laboratory experience?" Specifically, these beneficial outcomes would include student-and LA-perceived improvements to the laboratory learning community and affective domain.

2 Pedagogical frameworks

2.1 Vertical learning communities with near-peer instructors

Traditional undergraduate learning communities consist of students instructed by a graduate student teaching assistant (GTA) and/or professor with several years of experience and training in the subject matter. In this hierarchical organization, students may feel disconnected or even intimidated, creating a relational gap between the student and instructor (Hall et al., 2014; Bourne et al., 2021). The effects of this structure can be exacerbated in challenging courses, such as OCL (Micari and Pazos, 2012). Implementation of LAs in courses results in a vertical community of scholars; LAs are inserted into the traditional relationship of students and instructors, reducing the gap between their experience and labels (Bourne et al., 2021; Frosch and Goldstein, 2021). An LA or near-peer instructor is an individual who is close in age and education level but is one or more years senior in their educational progress to a student and seeks to provide mentorship and guidance (Bulte et al., 2007; Akinla et al., 2018). Price et al. suggest that the social element of vertical learning communities encourages students to develop collaborative problem-solving skills (Price et al., 2019). Our LA program was inspired by previously-established peer-learning programs, such as the Learning Assistant Program at the University of Colorado Boulder and the Undergraduate Teacher-Scholar Program at UC Berkeley (Otero et al., 2010; Bourne et al., 2021).

2.2 Affective domain in laboratory instruction

The affective domain is defined as encompassing students' attitudes, motivations, values, expectations, and emotions in the context of the learning process. Galloway has emphasized the relevance of the affective domain in Novak's framework of meaningful learning (Bretz, 2001; Galloway et al., 2016), which requires complete integration of the cognitive and affective domains with the psychomotor domain. For meaningful learning to occur in an OCL setting, the instructor should aim for holistic treatment of motivational and attitudinal aspects in addition to conceptual and procedural aspects when designing laboratory curricula. Seery argues for preparing and supporting students by managing their expectations for challenges and difficulties in the complex learning environment of chemistry laboratory courses, which LAs could facilitate (Seery et al., 2019).

Despite a relative lack of research on the affective domain in the chemistry laboratory compared to the cognitive or psychomotor domains, considerable effort has been made in assessing the effects of various learning interventions on affect (Penn and Ramnarain, 2019; Flaherty, 2020). Implementation of a process-oriented guided inquiry learning introductory chemistry course resulted in improved self-efficacy and confidence in students with little prior chemistry knowledge and experience (Vishnumolakala et al., 2017). The introduction of LAs during the return to in-person classes had the potential to positively impact students' comfort and safety in the laboratory, with the additional benefit of improving learning assistants' confidence and attitudes towards chemistry (Smith, 2008; Kornreich-Leshem et al., 2022).

3 Learning environment

The courses described herein took place at University of California, Irvine, a large public research university in the western United States, designated as a Minority Serving Institution, with the federal designations of Asian American and Native American Pacific Islander-Serving Institution and Hispanic-Serving Institution. The overall student population included in the study is 66% female and 34% male, based on self-reported responses to a binary-choice question about biological sex at time of admission. Students self-reported as 55% Asian, 17% Hispanic/Latinx, 13% white, 3% Black or African American, 8% Native Hawaiian and/or Pacific Islander, less than 1% American Indian and/or Alaskan Native; 3% declined to state. Overall, 44% of students self-identified as first-generation college students and 33% were identified as low-income. First-generation status was defined as neither parent completing a 4-year degree. Students who did not self-report income were assumed to be nonlow-income.

These courses are part of an OCL series consisting of a three-course sequence for non-chemistry majors (Figure 1). Each term is 10 weeks. Laboratory sections meet once per week for 4 h. Additionally, students are expected to attend a one-hour laboratory lecture section

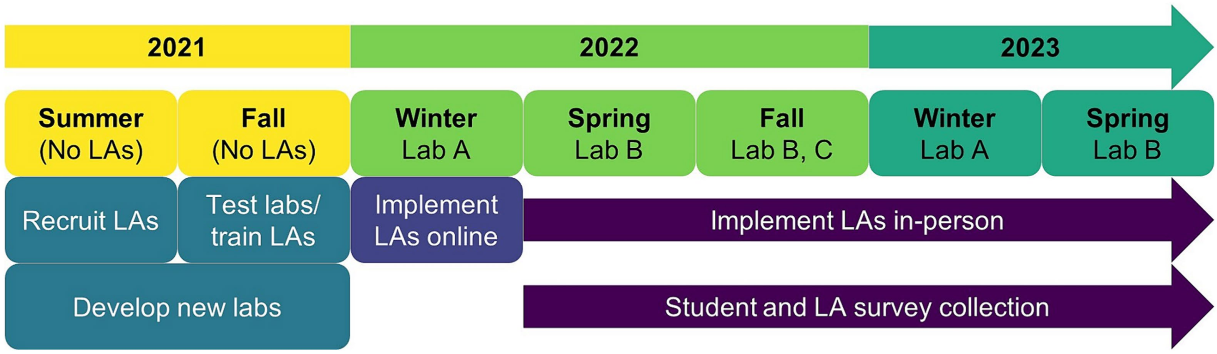


FIGURE 1
Timeline of curricula development, curricula testing, LA training, LA implementation, and survey collection.

once per week. Although the laboratory course, including laboratory sections and laboratory lecture sections, is separate from the organic chemistry lecture course, students enroll in both courses concurrently. Depending on the term and the specific course, an OCL course may have between 150–1,200 students enrolled. Up to 34 GTAs are required during a large-enrollment quarter, with each GTA assigned as the sole instructional staff member present for two 20-student laboratory sections per week.

The OCL course sequence described here is coordinated and taught by a single faculty member (RDL). Because of the scale of these courses, the instructional team includes multiple Head GTAs (JHG, JCT, PAL) who provide logistical, administrative, and pedagogical support to the instructor, including GTA scheduling, addressing grading discrepancies, writing exam questions, and hosting weekly office hours.

4 Results

4.1 Methods

4.1.1 LA implementation

This work represents the first implementation of LAs during the laboratory component of an OCL course at UC Irvine. Students serving as an LA for the first time in any course are required to enroll in the university-wide Certified Learning Assistants Program (CLAP), in which a certified instructor trains new LAs in pedagogical theory and strategies for facilitating classroom teaching. In Summer 2021, the remedial laboratory curricula were developed, and initial recruitment applications were sent to students who had performed well in the relevant course series (B+ or better) within the two prior academic years (Figure 1). In Fall 2021, the first cohort of LAs was accepted, and simultaneous laboratory safety/technique training and experiment testing took place. Implementation of LAs in lab sections occurred in Winter 2022, but continuing disruptions to instruction caused by the COVID-19 pandemic precluded survey data collection for this term. Spring 2022 represents the first term in which student and LA survey data were collected for a full in-person implementation of laboratory LAs.

Applications were sent to prospective LAs using Google Forms (Supplementary material). The application comprises three sections: (1) potential for effective peer instruction, (2) reflection on transferable professional skills, and (3) an example of answering student questions. Applications scoring highly on a rubric were accepted without a limit on the possible number of acceptances (Supplementary material). Application forms were distributed 4 weeks before the start of instruction; the application remained open for 2 weeks. Accepted LAs were notified 2 weeks before the start of instruction. One week before the start of instruction, LAs were assigned to laboratory time slots and were assigned based only on their individual availability such that each laboratory section was led by either one GTA *or* one GTA and one LA. In an average term of 60 individual laboratory sections, roughly 50% of sections had exactly one LA present, and the remainder did not have any LAs present.

During Fall 2021, LAs who had no prior experience handling chemicals and equipment in an instructional laboratory setting participated in training; these LAs earned the same course credit that they would have as lecture LAs. Training was carried out over five weekly two-hour periods. The first period was dedicated to safety training and familiarity with laboratory equipment. The latter four periods involved testing of both existing and new laboratory experiments. Students were provided with access to a draft version of an experiment handout where they could provide feedback. LA cohorts after Fall 2021 did not participate in training or experiment testing, as they had prior in-person experience and no new curricula were being tested.

4.1.2 LA responsibilities

All LAs attended a weekly 30-min meeting in which the Head GTA reviewed LA feedback from the previous week's experiment and summarized the upcoming experiment. Specific time was set aside for LAs to develop a plan to address anticipated student challenges or common misconceptions and mistakes. For example: At the beginning of our laboratory sections, students complete a collaborative set of questions concerning safety, equipment, and chemical principles; accordingly, LAs were provided with follow-up questions to guide student learning.

The primary responsibilities of LAs during laboratory time were to supplement GTA instruction by facilitating student completion of experimental work and achievement of related learning outcomes (GTA duties did not change). This was accomplished by addressing challenges and answering questions related to content, equipment, and procedure. During the laboratory section, LAs were free to develop an instructional plan with their GTAs based on information discussed during the weekly meeting. We were comfortable with LAs having the freedom to develop an independent instructional plan because of their CLAP training, but instructors at institutions without a CLAP analogue may want to be more prescribed in what in-laboratory activities are expected of their LAs.

The primary responsibility of LAs between weekly laboratory sections was to provide guided feedback on the completed experiment, addressing both the experiment itself and how students were or were not able to achieve learning goals efficiently. Our framework for collecting LA feedback was inspired by the implementation of "10-min journals" in peer-led team learning (Wilson and Varma-Nelson, 2021). Specifically, LAs provided answers to the following five questions about the laboratory experiment: (1) What went well? (2) What were "traps" or challenges for students? (3) Do you have suggestions for things that can be changed? (4) Do you have feedback on the writing of the experiment handout itself? (5) What information do students need clarification or additional instruction on *before* attending lab? This feedback was then aggregated by Head GTAs to be discussed in the following week's LA meeting. We used this feedback on a regular basis to make incremental improvements to the phrasing or organization of course materials.

4.1.3 Student and LA surveys

Our surveys were adapted from work by Bourne et al. on the implementation of a large-scale laboratory LA program at UC Berkeley (Bourne et al., 2021). Specifically, their study analyzed the types of questions students approach GTAs, LAs, and/or peers with during recitation/discussion and laboratory sections. We were interested in whether these findings were consistent for our student population, who were returning from pandemic-related educational disruptions. Additionally, we investigated student and LA perceptions of (1) which LA duties were appropriate and (2) student affect in the laboratory with or without an LA present. The student survey addressed four major themes — LA duties, student learning from LAs and GTAs, student affect, and laboratory time management — using a combination of Likert-scale and open-ended questions. The LA survey included open-ended questions designed to reference the LA application, specifically the professional and academic goals and skills sections (Supplementary material).

This study was approved by the Institutional Review Board as an exempt study (IRB #741). Surveys were available to students, LAs, and GTAs for 1 week during the final week of instruction through Qualtrics. Students were offered one credit toward a "token" for completing this and other research surveys in each quarter; tokens can be exchanged in the course grading system for options such as late passes or the opportunity to revise and resubmit unsatisfactory assignments (McKnelly et al., 2023).

Student survey responses ($n = 1{,}194$, 35%) were de-identified before analysis. Student survey response rates varied by term (S22 $n = 564$, 58%; F22 $n = 131$, 48%; W23 $n = 140$, 12%; S23 $n = 359$, 36%), but results remained consistent across terms regardless of response rate. Summary statistics of demographics for survey respondents matched those of the overall courses. LA survey responses (overall $n = 55$, 63%; S22 $n = 19$, 83%; F22 $n = 9$, 75%, W23 $n = 10$, 37%; S23 $n = 17$, 65%) were collected anonymously. GTA responses were not analyzed due to low response rates of two or fewer GTAs per term.

Analysis of Likert-type questions and multiple-select questions was conducted using the statistical programming language R (Garnier, 2018; R Core Team, 2019; Wickham et al., 2019). Responses from students to open-ended questions were analyzed using the Taguette free, open-source qualitative research tool (Taguette, n.d.). Students' own wording was used to identify themes based on identified relations, similarities, and differences that were grouped conceptually.

4.2 Survey results

4.2.1 LA duties

We adapted survey questions used by Bourne et al. to confirm whether LAs were performing expected duties during laboratory sections and gauge the perceived appropriateness of those duties by students and LAs. The actual duties of LAs within the program were to (1) provide information, (2) monitor laboratory safety, (3) supervise instrument use, (4) act as a role model, and (5) facilitate discussion (Figure 2). When students were asked how appropriate these five LA duties were, the majority of respondents (57–60%) indicated that duties 1–4 were "very appropriate," while the majority deemed duty 5 "usually appropriate" (41%) (Figure 2A). LAs were asked to self-assess these same duties and responded that 1–5 were "very appropriate" at higher rates than the students, particularly for duty 5, with 70% of LAs compared to 37% of students (Figure 2C). A similar trend emerged with student observations of their LAs during laboratory sections. The majority of students reported observing duties 1–4, but only 38% of respondents reported observing duty 5 (Figure 2B). Duty 1 was the most commonly observed by students at 91%, while duties 2 and 3 were also frequently observed at 79 and 75%, respectively. The most common duty performed by LAs was duty 2, in which 100% of LAs reported monitoring laboratory safety, and 95% of LAs reported performing duty 1 (Figure 2D). The largest differences between student and LA responses were that LAs reported performing duty 5 and duty 4 at higher rates than students reporting observing these activities, with discrepancies of 34 and 23%, respectively.

In addition to the actual LA duties, responsibilities that are instead assigned to other instructional staff were also included in this question. These non-LA duties included (6) planning laboratory activities, (7) creating course resources, and (8) grading student work. While students tended to underrate the appropriateness of duties 1–5 lower compared to LAs, the opposite trend was observed for duties 6–8. Student responses to the appropriateness of duty 6 were not as straightforward as for duties 1–5, with a broad range spanning "sometimes appropriate" (33%), "usually appropriate" (26%), and "very appropriate" (19%). LA responses skewed towards "sometimes appropriate" (50%), with 18% indicating "usually appropriate" and 8% indicating "very appropriate." Similarly, a small majority of students stated that duty 7 is "sometimes appropriate" at 32%; however, the majority of LA responses (48%) indicated it was "sometimes appropriate." Lastly, the majority of both students and LAs indicated that it would be "sometimes appropriate" for LAs to perform duty 8 (33 and 52%, respectively) with "not at all appropriate" being the second most common response (27% of students and 35% of LAs). Despite the fact that these duties were not assigned to LAs, a minority of both students and LAs (≤10%) observed or self-reported LAs performing duties 6–8.

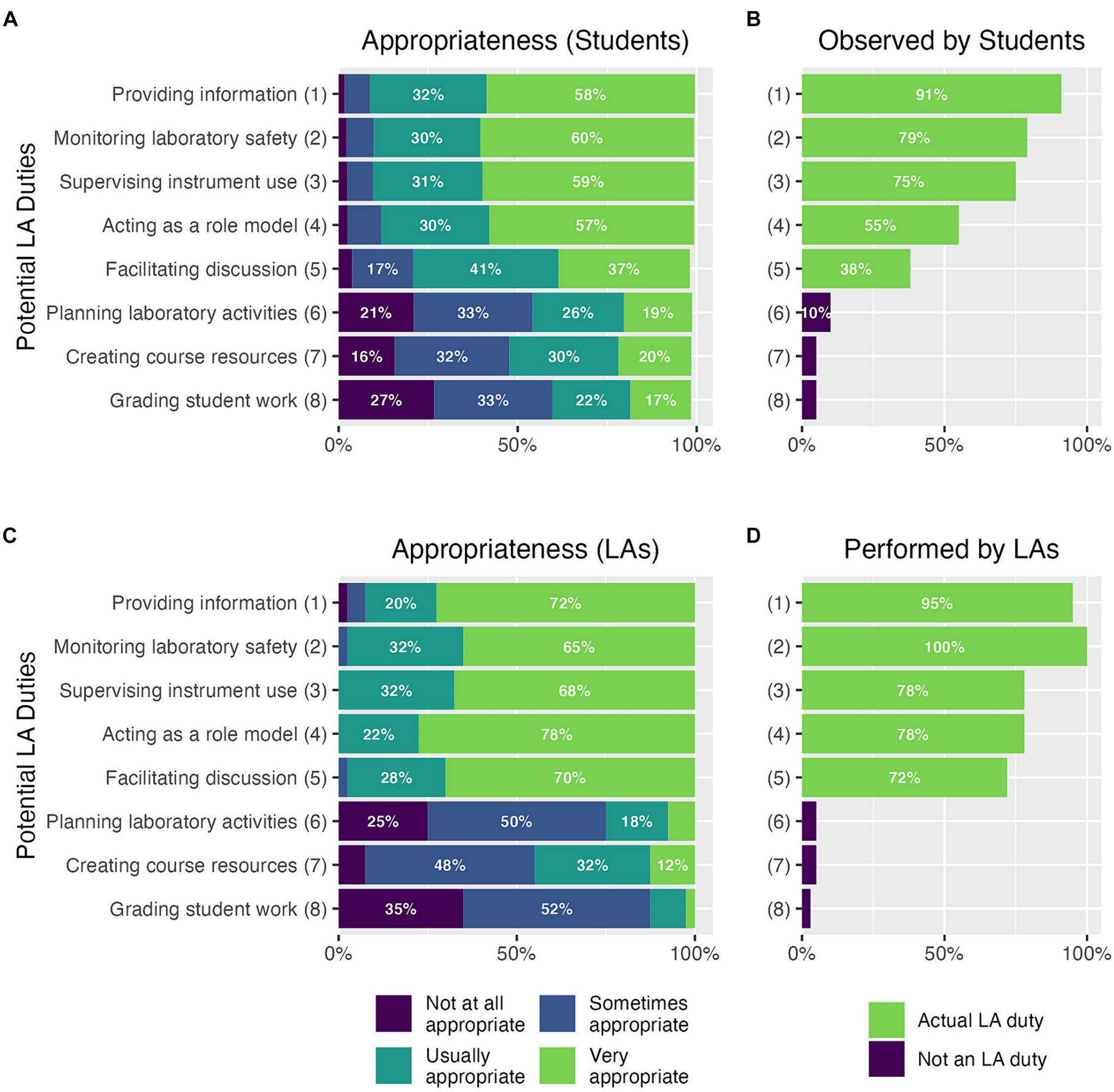


FIGURE 2
Student and LA responses to survey questions about the appropriateness and observation of LA duties. **(A)** Appropriateness of potential LA duties as determined by students. **(B)** Observation of LAs performing these duties as reported by students. **(C)** Appropriateness of potential LA duties as determined by LAs themselves. **(D)** Self-reporting of LAs performing these duties.

4.2.2 Student learning from LAs and GTAs

Following the work of Bourne et al., we investigated the order in which students preferred to ask GTAs, LAs, and peers certain types of questions (Figure 3). For content, equipment, and procedure questions, the majority of students indicated that they would approach a GTA first (51–64%). A smaller proportion of students indicated that they would approach a peer first (24–29%), and a small minority of students indicated that they would approach an LA first (6–20%). LAs were consistently the most popular second choice for asking these types of questions (52–56%), while students indicated that they would approach GTAs second 20–26% of the time and peers second 14–16% of the time. Grading did not follow the trend observed for content/equipment/procedure questions, as a larger majority of respondents (87%) indicated that they would approach GTAs first for questions about grading. Life questions were the only category in which students did not indicate GTAs as their first priority, as 48% reported that they would first approach a peer. Across all question categories, LAs were consistently considered students' second priority (grading: 57%; life: 48%).

4.2.3 Affective domain

A primary research question in this study was "Do students with no prior in-person laboratory experience self-report increased feelings of comfort, confidence, and safety when a laboratory LA is present?" In Figure 4A, we have separated results from the Spring 2022 term compared to Fall 2022–Spring 2023 (Figures 4B,C). Spring 2022 was the first term at our university since 2020 in which OCL courses were offered fully in person. For Spring 2022, a dramatic difference in students' self-assessment of their comfort, confidence, and safety was noted compared to later terms. Confidence appeared to be split fairly evenly between "agree" (no LA 48%, LA 54%) and "disagree" (no LA 52%, LA 46%) responses, while comfort and safety lean slightly

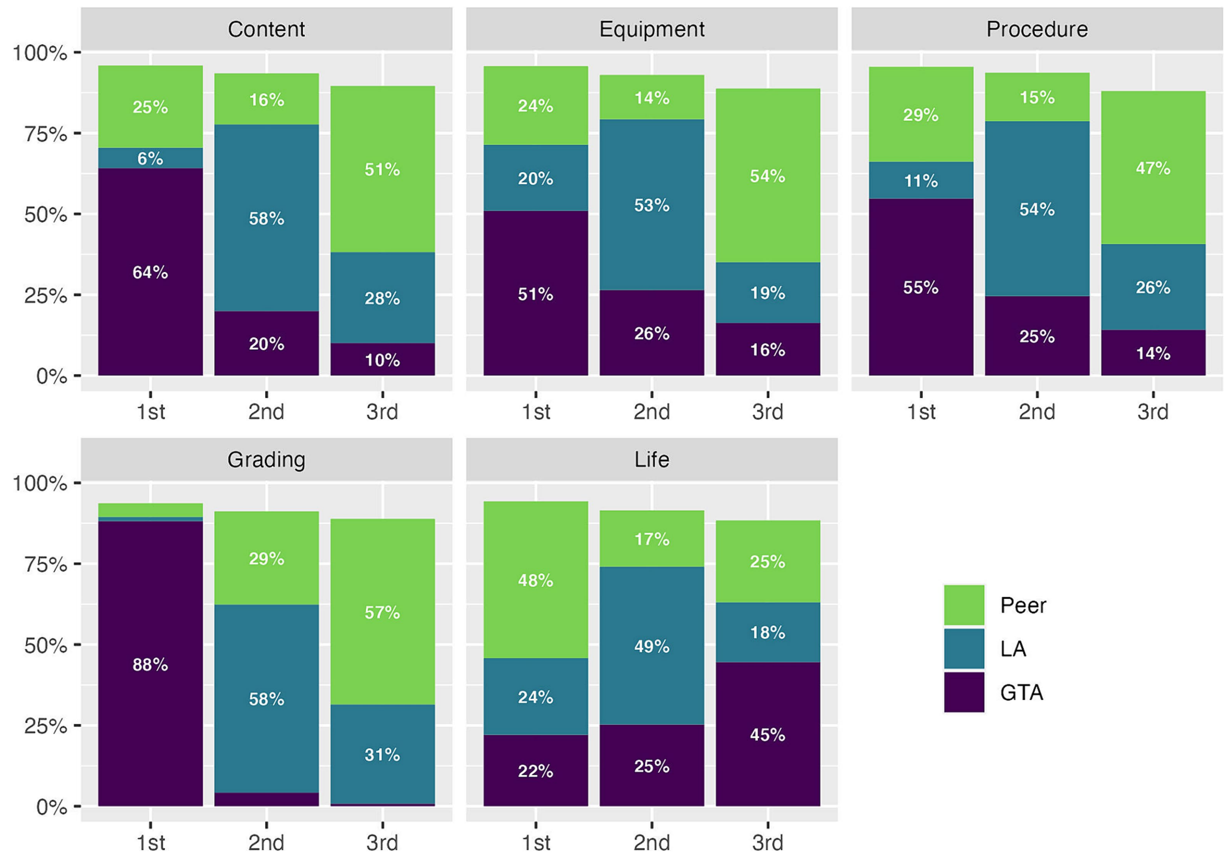


FIGURE 3
Student responses to the question "Please indicate the order in which you would approach the following people (GTA, LA, peer) to ask the following types of questions." Survey respondents were not required to give a 1st, 2nd, and 3rd priority for each category of question, so percentages may not add up to 100%.

towards "agree" responses (comfort: no LA 63%, LA 66%; safety: no LA 63%, LA 66%). For students with no LAs, comfort, confidence, and safety were strongly disagreed with 35, 47, and 34% of the time, respectively. With an LA, these results were 31, 40, and 31%, respectively.

Expectedly, students reported feeling more comfortable, confident, and safe when they had prior laboratory experience compared to when they did not: For Fall 2022–Spring 2023, responses overwhelmingly skew towards "agree" for comfort (no LA 96%, LA 96%), confidence (no LA 93%, LA 93%), and safety (no LA 97%, LA 98%). Anxiety, a negative affect trait compared to the positive traits of comfort, confidence, and safety, was consistently more varied in student responses: Furthermore, students appeared to self-report higher anxiety in terms other than Spring 2022, as "agree" responses increased (no LA: 32 to 49%; LA: 39 to 47%).

4.2.4 Open-ended survey question responses

Responses from students and LAs to open-ended questions were used to identify major conceptual themes describing the benefits of the laboratory LA program (Table 1). As a result of LAs answering student questions, students identified improvements to experimental efficiency and LAs identified improvements to communication skills. Students described LAs as being effective near-peer instructors, highlighting previous experience in the course, approachability, and their similar institutional knowledge. Students described LAs as beneficial to the learning experience because LAs were able to support both GTAs and students with their experienced perspective. Additionally, students felt that LAs promoted a safer laboratory environment.

5 Discussion

5.1 Comparison to previous results

Despite differences in our implementation of laboratory LAs (the COVID-19 pandemic, our LA training process, 10-week course length, etc.), our results are consistent with Bourne et al. Students approached TAs first for all question categories other than life, where peers are instead ranked first (Figure 3). LAs were ranked as second for each question category. For grading specifically, TAs were overwhelmingly ranked first. Furthermore, we observed similar student observation and LA self-assessment of the various LA duties; for each of the actual LA duties, a majority of both student and LA respondents reported these duties (with the exception of "facilitating discussion;" see Limitations) (Figure 2). For each of the duties LAs were *not* intended to perform, the majority of students and LAs did *not* observe or report these duties, respectively.

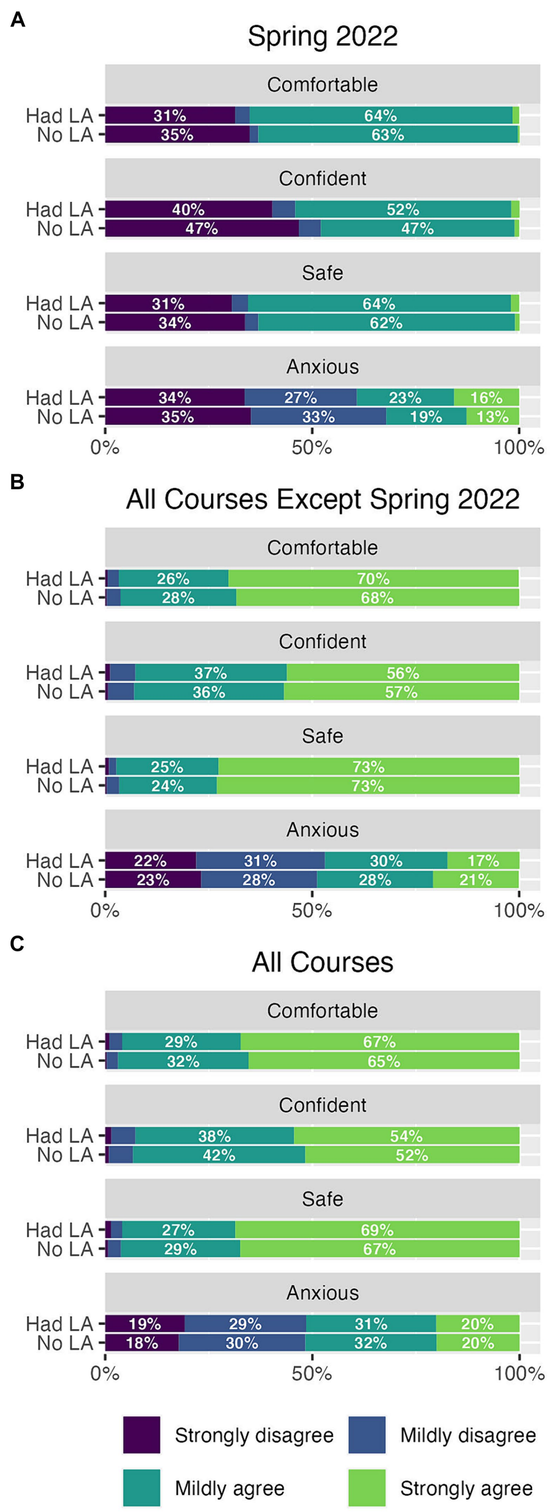


FIGURE 4
Student responses, separated by presence or absence of an LA in their laboratory section, to the question "Please rate the degree to which you agree with the following statement: 'I felt [adjective] performing in-person laboratory experiments this quarter.'" **(A)** Student responses for Spring 2022 only, the first instructional period since the beginning of the COVID-19 pandemic in which the full 10-week laboratory course was offered in person with no interruption. **(B)** Student responses for each term during the study period other than Spring 2022. **(C)** Student responses for all terms included in the study period.

5.2 Tandem LA training/experiment testing

We find that LAs are useful for testing new or revised curricula. Because of their previous experience in the course, LAs are motivated to improve the clarity and organization of course materials. Following a session of LA experiment testing, we propose organizing LA feedback into (1) identified problems and (2) proposed solutions. While students and LAs are effective at identifying problems, their proposed solutions are not always actionable; this necessitates the review of LA comments and suggestions by an experienced instructor before changes are made to course materials. We find that LAs feel prepared by their training (Supplementary Table S2). Their familiarity with the tested experiments further equipped them to address student questions.

5.3 Student and LA perspectives of laboratory LA benefits

5.3.1 LAs answer student questions

Students report that LAs improve laboratory efficiency. Many students recalled a long line to ask the GTA a question; the presence of an additional instructor in the form of an LA increased the rate at which questions were answered. This is consistent with Figure 3, in which most students would approach the LA second for content, equipment, and procedure questions. Many students reported directing "minor" questions to LAs, while TAs addressed more in-depth questions about chemical principles or specific experimental troubleshooting. However, students indicated that, during periods of an experiment where they had few questions, LAs passively waited until there were questions to answer rather than approaching students to initiate discussion.

LAs consistently indicated the importance of observing, learning, and practicing teaching and communication skills in both their motivations for participating in the program and their primary outcomes from participating. To assess achievement of personal goals, LAs were asked to recall and reflect on their motivations for joining the program and the transferable skills identified as part of their application to the program. Responses overwhelmingly reflected the desire to practice teaching and communication skills, both for career development and to help fellow undergraduate students. Some LAs expressed a desire for more direction and did not feel comfortable approaching students to initiate discussions when students were not asking questions, but overall, LAs felt more confident in their abilities to solve problems, clearly communicate information, and guide students in their learning process. These skills were specifically identified as being transferable to education, healthcare, and GTA positions in graduate programs.

5.3.2 Near-peer instructors facilitate student learning

Students report experiencing emotional stress during laboratory experiments — which may be caused by time management issues, the desire to obtain perfect results, or concerns about their grade — that could inhibit learning in the course. Although student responses to Likert-scale questions about their comfort in the laboratory did not differ based on whether or not an LA was present (Figure 4), open-ended responses indicated an increased sense of comfort from LAs

TABLE 1 Representative examples of common responses to open-ended survey questions.

	Student Perspective	LA Perspective
LAs answer student questions. *Answering questions, laboratory efficiency, communication*	"Yes I think having multiple upper division role models will help when students have a lot of questions. It also makes labs go by much faster when there are more sets of hands to aid in conducting the lab. My LA was also very helpful when my TA was busy helping someone else. I could ask her for anything about the experiment or conceptually questions and she mostly has the answers to help." "Labs are usually very hectic and many students are constantly asking questions, so having a second option to refer to for questions greatly helped keep everyone moving efficiently." "Having the sense of mentorship is a nice touch. Having an extra set of eyes and supervisor helps cut down the time spent on waiting for the TA answer your questions after something goes wrong."	"I wanted to improve on my active learning skills during my time as an LA. I feel as though the lab course I chose to assist helped immensely with this because it challenged me to come up with certain tactics to use in assisting students rather than just giving them the answer. This helps both me and the student because it strengthens problem solving skills while also allowing the students to use their own knowledge to get to the answer themselves." "As far as I can recall, the most important professional skill I wanted to to refine was my communication/teaching skills because in the field of medicine it is required of people to be able to elaborate certain medical knowledge that can be difficult to explain without a particular level of education. I believe that so far this position has been great at helping me rethink how I explain things and I have been able to compartmentalize what knowledge is important to knowing the concept vs. what knowledge is going overboard better than I have been able to do before." "It is a great experience if someone is trying to get exposed to more teaching positions. It also strengthens communication and problem-solving skills which is an important skill to have for the future and life in general."
Near-peer instructors facilitate student learning. *Near-peer instructors, approachability, community*	"I see LAs as the middle ground between TA and peer. They are super helpful because sometimes students do not want to go straight [to] the TA for what they may think is a stupid question. If a peer does not know, LA is the next best. I find this to be pretty common in lab settings. Maybe someone messed up the experiment, but they are too embarrassed to ask for help. A kind and understanding LA would be awesome!" "LAs are more relatable for undergraduates, and just having them present is reassuring because they were in our shoes somewhat recently and they passed the class." "It's nice seeing someone relatively our age be passionate about chemistry. It encourages learning [in] the lab environment."	"I think having another student in the lab really helps both the TA and students. It makes it easier for students to approach other fellow students and can facilitate a more comfortable environment where discussion and questions are encouraged. I think that the main thing [is] that I am only a year older than most of the students, it provides another person of 'authority' that the students can depend on while also being more comfortable with as there is a very tiny age gap…" "I think the lab in general can be pretty long and tiring which can exhaust students sometimes, but seeing someone who has taken the lab and come back to LA can make them feel like they are capable of getting through it. It also gives them the opportunity to ask questions about their current course content and future courses in a bit of a peer-to-peer way rather than [professionally]." "I feel like I succeeded since many students ended up enjoying ochem lab. It wasn't a stressful experience and it made people open their eyes to how great chemistry is."
GTAs and LAs form a cooperative teaching team. *GTA-LA teamwork, LAs supplement and support TAs, experienced student perspective*	"Yes Having the LA program is beneficial for both the student and the LA. In the case of the students, it allows a different perspective of the experiment and being an undergraduate student compared to the graduate student TAing." "If in-lab Learning Assistants are present in the lab, the TA will not be too busy tending to students' questions and will have time to go over crucial concepts more thoroughly with the lab section. The chances of safety and waste violations such as breaking equipment and items being placed into the wrong waste containers might be lowered. Overall, having more eyes and hands to monitor the multiple reactions happening in lab will make things more efficient and safe" "I had an LA during Winter 2022. They were really helpful with answering questions about experiment procedure and safety when the TA was busy with other students. I was also able to clarify concepts with the LA during down time if the TA was busy."	"Quick rundown of what the experiment run should look like, potential issues, common questions we'll get, demonstrations and theory we need to go over before the lab, what can I do as the LA to help her and the lab run smoother." "Before each lab, we would talk about how 'tricky' students may find the experiment, or if I got a lot of the same question I would let them know so they could make an announcement or address it in some way" "With both TAs I worked with, I was able to converse with them freely. They both made me feel like we ran lab as a unit, a team. Both completing the same duty of answering the students questions. There were often times where I did not know how to answer a students question, so I asked my TA and got back to them. When there was down time, my TA and I would sometimes stand towards the front and talk about random stuff."

Responses are categorized by common themes and highlight the unique perspectives of students and LAs.

that made the laboratory sections more enjoyable. Students perceive LAs to be more relatable than GTAs or professors because they are closer in age and experience to the students and have recently been in their position. Students identified mentorship and role-modeling as additional benefits of the LA's presence in the laboratory, which is consistent with the majority of students identifying "acting as a role model" as being an appropriate LA duty (Figure 2). LAs can "empathize" and "understand the struggle" students are encountering and provide an experienced perspective on how to succeed in the course. A number of students shared that LAs gave them advice about navigating their undergraduate degree. LAs additionally brought camaraderie to the laboratory sections, making them more "fun" and "interactive" while still ensuring that experiments were conducted safely.

Many students addressed the teaching hierarchy that vertical learning communities with near-peer instructors seek to mitigate. Students reported feeling intimidated to approach their GTA with certain questions if a mistake was made or for fear of being judged. LAs serve to assist both parties by answering student questions and reducing the burden on the GTA. Accordingly, students reported that LAs are generally more approachable than GTAs, citing that LAs had no power over grades and that LAs tended to explain concepts in a way more digestible to undergraduate students.

Many LAs were motivated by personal experience, joining the LA program out of a desire to reduce stress and increase confidence for students by being an approachable source of support and familiarity. LAs commented on their personal struggles when enrolled in the course and wanted to share their expertise. Throughout an instructional term, LAs described developing a rapport with students by discussing subjects outside of chemistry to help ease chemistry-related discussions. LAs emphasized the near-peer aspect of the program, in which students who were intimidated by their GTA could instead ask someone closer in age and experience. In "bridging the gap" between the GTA and students, LAs reported connecting professionally and personally with both the students and the GTA, fostering a sense of community in the laboratory.

5.3.3 GTAs and LAs form a cooperative teaching team

Students overwhelmingly recommended that the LA program be continued, with the most common reason being that LAs help to supplement GTAs in the laboratory. Students recognize that GTAs are often busy running the laboratory section and cannot help every student simultaneously. A common example was an experiment in which the GTA operated an instrument in an adjacent room while most students remained in the main laboratory space; LAs assisted GTAs by being where the GTA could not. Students recognized that the ability for LAs to supervise students while the GTA was busy improved overall laboratory safety (Figure 2). Finally, students recognized that LAs provide a useful and complementary perspective to the GTA, as LAs are current undergraduate students and have already performed well in the OCL series. This perspective reaches beyond course content, as 85% of students indicate that "My LA helped me improve my understanding of how to navigate UCI as an undergraduate" (Supplementary Figure S1).

LAs and GTAs were expected to work as a team, conferring at the beginning of a laboratory period to discuss how to optimize time management and students' general experience with the experiment at hand. Correspondingly, LAs were surveyed regarding interactions with their laboratory section's GTA. A small number of LAs mentioned consulting with the GTA before the laboratory period began to get a general sense of how the laboratory period should proceed. This type of response was less common than expected, which may indicate that our implementation of LAs would benefit from increased structure and clearer expectations of LAs. Based on survey responses, LAs understand their role as being supplemental to and supportive of GTAs; in other words, LAs recognize that their participation can benefit both the students and the GTA.

6 Limitations

The primary limitation of this study is the use of surveys that are not validated instruments. Specifically, we observe that certain words or phrases, such as "facilitate discussion," may be interpreted differently by students and LAs and that those interpretations may differ from our intent. In the case of "facilitate discussion," students may interpret this to refer to a recitation section of the course as opposed to the laboratory component (Figure 2). LAs may instead interpret "facilitate discussion" to mean "facilitated discussions/conversations about concepts with students," which is closer to our intent. Additionally, the results presented are in aggregate and may not represent the experiences of students who hold specific marginalized identities. Due to the scale of our OCL series and the complexity of undergraduate student scheduling, it is unlikely that we will ever be able to provide an LA for each laboratory section in a single term. This is a limitation of our implementation, as LAs are not assigned evenly throughout different section types (i.e., day of week and time of day).

7 Conclusion

We have described the process by which we implemented a laboratory LA program in non-chemistry major OCL for the first time at our institution. This was done in response to instructional discontinuity caused by the COVID-19 pandemic, which necessitated the development of remedial laboratory curricula. In order to (1) test these new curricula and (2) train LAs in hands-on laboratory techniques, LAs participated in the development of the new curricula. Following implementation of both LAs and the new experiments in the OCL series, survey results from students and LAs were compared to the previous study by Bourne et al., which took place prior to shutdowns caused by COVID-19. We find that students correctly identify LAs duties and prioritize LAs over peers when asking questions about experimental content, equipment, or procedure, which is consistent with the previous study. We identify three major categories of student and LA open-ended survey responses which describe the benefits that LAs bring to the teaching laboratory.

We plan to repeat this strategy of curricular development/LA training in the near future as we transition our OCL format to Argument-Driven Inquiry (Walker and Sampson, 2013; Howitz et al., 2023; Saluga et al., 2023). LA feedback indicated that the program could benefit from increased structure, such as additional prescribed leading and exit questions to engage student groups during experiments. Qualitative GTA feedback (excluded from this work) indicated that LA-GTA teamwork could be improved if GTAs were provided with a specific list of LA responsibilities. We have created and compiled resources with which other instructors in a broad range of learning environments can recruit, train, and implement LAs while developing new laboratory curricula. The LA application form, associated rubric, and surveys for both students and LAs are included in the Supplementary material. Tandem LA training/curricular design proved useful in responding to the instructional interruption caused by the COVID-19 pandemic, but we believe that this strategy is generalizable to any kind of curricular innovation/reform in a chemistry laboratory course series. Although the large-scale disruptions to in-person courses necessitated by the onset of the COVID-19 pandemic have passed, other events such as labor actions, natural disasters, or civil unrest could result in a cohort of students entering laboratory courses without in-person laboratory experience; these students may benefit from the presence of LAs in their laboratory courses.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by University of California, Irvine Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JG: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. JT: Formal analysis, Writing – original draft, Writing – review & editing, Investigation. PL: Formal analysis, Writing – original draft, Writing – review & editing. RL: Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to recognize MaryAnn Robak, Peter Marsden, and Anne Baranger for inspiring this OCL LA program. JG participated as an Undergraduate Teacher-Scholar at UC Berkeley during the time of Bourne et al.'s study, which was indispensable in developing our own program. Sarah Wang helped to implement a preliminary version of the current program in the Spring 2021 term. We would like to recognize Philip Lechner for assisting in the continuation of the LA program in the Fall 2023 term. Denise Bui and Celso Lopez provided integral support in OCL course implementation. We thank Kameryn Denaro and UCI PERI2 for consultation on statistical analyses and course demographic data. Finally, the authors would like to thank the participating students and LAs for their survey feedback, as well as Joshua Arimond and Alessandra Pantano for coordinating LA recruitment and the CLAP training program.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1367087/full#supplementary-material

References

Akinla, O., Hagan, P., and Atiomo, W. (2018). A systematic review of the literature describing the outcomes of near-peer mentoring programs for first year medical students. *BMC Med. Educ.* 18:98. doi: 10.1186/s12909-018-1195-1

Barrasso, A. P., and Spilios, K. E. (2021). A scoping review of literature assessing the impact of the learning assistant model. *Int. J. STEM Educ.* 8, 1–18. doi: 10.1186/s40594-020-00267-8

Bourne, S. M., Limfat, S., Dilip, S., Han, S., Robak, M. T., Marsden, P., et al. (2021). Comparing near-peer and non near-peer instructors in laboratory courses. *J. Coll. Sci. Teach.* 50, 10–17. doi: 10.1080/0047231X.2021.12290521

Bretz, S. L. (2001). Novak's theory of education: human constructivism and meaningful learning. *J. Chem. Educ.* 78:1107. doi: 10.1021/ed078p1107.6

Bulte, C., Betts, A., Garner, K., and Durning, S. (2007). Student teaching: views of student near-peer teachers and learners. *Med. Teach.* 29, 583–590. doi: 10.1080/01421590701583824

Emenike, M. E., Schick, C. P., Van Duzor, A. G., Sabella, M. S., Hendrickson, S. M., and Langdon, L. S. (2020). Leveraging undergraduate learning assistants to engage students during remote instruction: strategies and lessons learned from four institutions. *J. Chem. Educ.* 97, 2502–2511. doi: 10.1021/acs.jchemed.0c00779

Flaherty, A. A. (2020). A review of affective chemistry education research and its implications for future research. *Chem. Educ. Res. Pract.* 21, 698–713. doi: 10.1039/C9RP00200F

Frosch, E., and Goldstein, M. (2021). Using a learning community model for virtual medical student support during the COVID19 pandemic. *J. Int. Assoc. Med. Sci. Educ.* 12, 136–139. doi: 10.5116/ijme.60e2.c777

Galloway, K. R., Malakpa, Z., and Bretz, S. L. (2016). Investigating affective experiences in the undergraduate chemistry laboratory: students' perceptions of control and responsibility. *J. Chem. Educ.* 93, 227–238. doi: 10.1021/acs.jchemed.5b00737

Garnier, S. (2018). Viridis: default color maps from matplotlib. Available at: https://CRAN.R-project.org/package=viridis (Accessed December 30, 2023).

Hall, S., Stephens, J., Andrade, T., Davids, J., Powell, M., and Border, S. (2014). Perceptions of junior doctors and undergraduate medical students as anatomy teachers: investigating distance along the near-peer teaching spectrum. *Anat. Sci. Educ.* 7, 242–247. doi: 10.1002/ase.1419

Howitz, W. J., Frey, T., Saluga, S. J., Nguyen, M., Denaro, K., and Edwards, K. D. (2023). A specifications-graded, sports drink-themed general chemistry laboratory course using an argument-driven inquiry approach. *J. Chem. Educ.* 100, 672–680. doi: 10.1021/acs.jchemed.2c00860

Kornreich-Leshem, H., Benabentos, R., Hazari, Z., Potvin, G., and Kramer, L. (2022). The cognitive and affective roles of learning assistants in science, technology, engineering, and mathematics college classrooms: an exploration of classroom experiences and students' metacognitive awareness and disciplinary identity. *Sci. Educ.* 106, 545–572. doi: 10.1002/sce.21703

Lawhon, C. (2020). Answering the call: the story of how UCI responded and adapted to the historic COVID-19 pandemic. Available at: https://specialreports.news.uci.edu/uci-during-covid/index.php (Accessed December 30, 2023).

McKnelly, K. J., Howitz, W. J., Thane, T. A., and Link, R. D. (2023). Specifications grading at scale: improved letter grades and grading-related interactions in a course with over 1,000 students. *J. Chem. Educ.* 100, 3179–3193. doi: 10.1021/acs.jchemed.2c00740

Micari, M., and Pazos, P. (2012). Connecting to the professor: impact of the student–faculty relationship in a highly challenging course. *Coll. Teach.* 60, 41–47. doi: 10.1080/87567555.2011.627576

Otero, V., Pollock, S., and Finkelstein, N. (2010). A physics department's role in preparing physics teachers: the Colorado learning assistant model. *Am. J. Phys.* 78, 1218–1224. doi: 10.1119/1.3471291

Penn, M., and Ramnarain, U. (2019). South African university students' attitudes towards chemistry learning in a virtually simulated learning environment. *Chem. Educ. Res. Pract.* 20, 699–709. doi: 10.1039/C9RP00014C

Price, S., Wallace, K., Verezub, E., and Sinchenko, E. (2019). Student learning assistants: the journey from learning advice to creating community. *J. Furth. High. Educ.* 43, 914–928. doi: 10.1080/0309877X.2018.1425379

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Saluga, S. J., Burns, A. M., Li, Y., Nguyen, M. M., and Edwards, K. D. (2023). A specifications-graded, spice-themed, general chemistry laboratory course using an argument-driven inquiry approach. *J. Chem. Educ.* 100, 3903–3915. doi: 10.1021/acs.jchemed.3c00433

Seery, M. K., Agustian, H. Y., and Zhang, X. (2019). A framework for learning in the chemistry laboratory. *Isr. J. Chem.* 59, 546–553. doi: 10.1002/ijch.201800093

Smith, T. (2008). Integrating undergraduate peer mentors into Liberal arts courses: a pilot study. *Innov. High. Educ.* 33, 49–63. doi: 10.1007/s10755-007-9064-6

Taguette (n.d.). Available at: https://www.taguette.org/ (Accessed December 7, 2022).

Ten Cate, O., and Durning, S. (2007). Peer teaching in medical education: twelve reasons to move from theory to practice. *Med. Teach.* 29, 591–599. doi: 10.1080/01421590701606799

Thompson, A. N., Talbot, R. M., Doughty, L., Huvard, H., le, P., Hartley, L., et al. (2020). Development and application of the action taxonomy for learning assistants (ATLAs). *Int. J. STEM Educ.* 7:1, 1–14. doi: 10.1186/s40594-019-0200-5

Vishnumolakala, V. R., Southam, D. C., Treagust, D. F., Mocerino, M., and Qureshi, S. (2017). Students' attitudes, self-efficacy and experiences in a modified process-oriented guided inquiry learning undergraduate chemistry classroom. *Chem. Educ. Res. Pract.* 18, 340–352. doi: 10.1039/C6RP00233A

Walker, J. P., and Sampson, V. (2013). Argument-driven inquiry: using the laboratory to improve undergraduates' science writing skills through meaningful science writing, peer-review, and revision. *J. Chem. Educ.* 90, 1269–1274. doi: 10.1021/ed300656p

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *JOSS* 4:1686. doi: 10.21105/joss.01686

Wilson, S. B., and Varma-Nelson, P. (2021). Implementing peer-led team learning and cyber peer-led team learning in an organic chemistry course. *J. Coll. Sci. Teach.* 50, 44–50. doi: 10.1080/0047231X.2021.12290507

frontiers | Frontiers in Education

TYPE Curriculum, Instruction, and Pedagogy
PUBLISHED 04 April 2024
DOI 10.3389/feduc.2024.1379216

OPEN ACCESS

EDITED BY
Jodye I. Selco,
California State Polytechnic University, Pomona, United States

REVIEWED BY
Justin Houseknecht,
Wittenberg University, United States
Troy Terry,
University of North Georgia, United States

*CORRESPONDENCE
Sarah Kathryn Zingales
✉ szingales@usj.edu

RECEIVED 30 January 2024
ACCEPTED 08 March 2024
PUBLISHED 04 April 2024

CITATION
Moster CA and Zingales SK (2024) Use of specifications-based grading in an online, asynchronous graduate organic chemistry course.
Front. Educ. 9:1379216.
doi: 10.3389/feduc.2024.1379216

Use of specifications-based grading in an online, asynchronous graduate organic chemistry course

Caleb A. Moster and Sarah Kathryn Zingales*

Department of Chemistry, University of Saint Joseph, West Hartford, CT, United States

Specifications-based grading is an alternative grading scheme that emphasizes student proficiency of learning objectives. Course grades are determined by the number of objectives completed rather than the number of points accumulated. At the University of Saint Joseph, CHEM 510 Intermediate Organic Chemistry is the foundation course that all incoming graduate students take in their first semester of the fully online, asynchronous MS programs in Chemistry and Biochemistry. Students in CHEM 510 complete the entire course online and at their own pace within the structured due dates, which presents unique challenges compared with synchronous learning modalities. With these considerations in mind, CHEM 510 was revised to use a specifications-based grading scheme with an a la carte assessment menu and token system. Generally, students found the alternative grading scheme helpful, but they needed additional instructions and time to adjust to the new grading system. By the end of the semester, students expressed their appreciation for the ability to choose their assessment method, work at their own pace, and use the token system for extensions/retakes. The instructor found that implementation of specifications grading took greater time for the initial course setup, but did not require more time than points-based grading once the course began. One large positive outcome was that student-instructor interactions were more frequently about the content of the course rather than grades. Overall, there was a slight increase in the course's pass rate compared to the pass rate prior to the change in grading modality. We believe that the implementation of the a la carte assessment menu accommodates a more diverse population of learners without sacrificing the integrity of student learning. Additionally, we believe that the diverse assessment opportunities were critical for the successful implementation of specifications-based learning in the online classroom environment, though further extension of the menu in synchronous, in-person classroom settings may be challenging.

KEYWORDS

organic chemistry, graduate education, specifications-based grading, online instruction, alternative grading

1 Introduction: background and rationale for the educational activity innovation

Students often identify Organic Chemistry as one of the most challenging subjects in the college curriculum (Johnstone, 2010) and with the recent expansion of online education initiated by the COVID-19 pandemic (Sunasee, 2020), the academic needs of these students have increased (Crucho et al., 2020). One approach to mitigating negative student perceptions of college-level chemistry courses involves changing the way that grades are distributed by employing various alternative grading methods (Herman, 1992; Brookhart et al., 2016). Instead of using the traditional points-based grading system, alternative grading methods such as standards-based learning (O'Connor, 2002), contract grading (Danielewicz and Elbow, 2009), mastery or competency-based learning (Bloom, 1968), and "ungrading" (Blum, 2020) have become popular alternatives. Alternative grading methods have seen positive results in the classroom setting (Atifnigar et al., 2020; Cain et al., 2022), but none are without their drawbacks: often, the change from traditional grading to alternative grading is associated with increased workload for instructors and/or increased stress for students (Peters and Buckmiller, 2014).

To mitigate these drawbacks, specifications-based grading (Nilson and Stanny, 2023) has emerged as a method for assessing student learning that focuses on clearly communicating learning objectives and encouraging student engagement with material (Howitz et al., 2021). Within Chemistry, specifications-based grading has been used in small lecture settings (Ring, 2017; Donato and Marsh, 2023) and large ones (McKnelly et al., 2023), and even has been implemented in lab-based (Bunnell et al., 2023; Howitz et al., 2023) and writing-based (McKnelly et al., 2021) courses. Among these successful implementations of specifications-based grading, several were done in Organic Chemistry; however, few, if any, publications have explored the impact of specifications-based grading in an online, asynchronous course or graduate level course. In this article, we explore the impact of specifications-based grading in CHEM 510 – an online master's level course in Organic Chemistry – designed for the university's online asynchronous master's programs in chemistry and biochemistry. We comment on the results and give insight into adjustments made to the course over the span of several academic semesters.

2 Pedagogical framework(s), pedagogical principles, competencies/standards underlying the educational activity

Specifications-based grading (specs grading) is one of several different alternative forms of grading developed by educators to enhance student learning. The design of specs grading was informed by the successes and shortcomings of other, older alternative grading methods (Nilson and Stanny, 2023). Below, we highlight the main features of some of the most prominent alternative grading methods and their strengths and weaknesses identified by previous publications.

2.1 Standards-based grading

Standards-based grading (SBG) courses are designed around specific learning objectives and use formative and summative assessments to determine whether a student has grasped the material of each objective (Marzano, 2010; Boesdorfer et al., 2018). Early and frequent informal assessment is used to provide feedback to students on their progress toward the ultimate performance goal without receiving a grade on their work (Iamarino, 2014). Since students are not penalized for misunderstandings exhibited during these early assessments, students are more willing to revisit difficult material to prepare for the formal scored assessments later in the course. Student grades on the formal assessments are determined by rubrics with clear guidelines for determining the level of student achievement in each of the learning standards (Curley and Downey, 2023). Instead of awarding points, a rubric specifies the criteria for distinct levels of student proficiency in each standard (Boesdorfer et al., 2018). Since student work cannot be scored as falling between proficiency levels in the rubric, each rubric is written with clear language to eliminate grading bias. Often, students are given multiple attempts to prove proficiency in a standard through formal assessment. Most publications of standards-based grading indicate a positive response from students, who feel as though their perspective on learning was changed for the better (Iamarino, 2014); however, other students have confessed to difficulties with motivation and a lack of self-initiated study habits that may have hindered their learning experience (Guskey, 2001).

2.2 Mastery- and competency-based learning

Mastery-based learning courses, like SBG courses, are divided into concept groupings based on related course material. However, in mastery-based grading, students may not advance in the course without demonstrating mastery on the current topic (Block and Airasian, 1971). Where standards-based grading provides multiple, though limited, opportunities for students to demonstrate proficiency, mastery-based learning allows students practicably limitless attempts, giving constructive feedback on their answers after each attempt. This prevents students from advancing to a new module without vital knowledge from a previous concept and allows them to focus on their weaknesses to master the objective (Bloom, 1968).

Competency-based grading builds on mastery-based learning, adding additional course-related work students may complete to increase their grade (Diegelman-Parente, 2011). This often means that students achieving mastery in a module will earn a letter grade of a B, while an A can be earned by the completion of additional work.

Students in mastery-based learning courses report drastically reduced anxiety over assessments and increased confidence in material (Peters and Buckmiller, 2014). The downside to these grading styles is the strain it places on instructors, who must find ways to generate and grade multiple assessments for each module and provide timely feedback to students (Bangert-Drowns et al., 1991). Instructors also must coordinate between students who progress through the material at different rates, adding to the complexity of maintaining student grades.

2.3 Contract grading

In contract grading, students generate a contract that outlines in detail what coursework they must complete to get a desired letter grade in the course (Taylor, 1980). Students present their contract to the instructor at the beginning of the year, then modify it until both parties are satisfied (Lindemann and Harbke, 2011). Contracts indicate how many assignments from each category (quizzes, discussions, laboratory experiments, etc.) must be passed to earn a given letter grade for the course.

This approach emphasizes transparency of expectations, the main strength of contract grading and a consistent weakness of points-based grading (Danielewicz and Elbow, 2009). When students and instructors find a compromise during the contract negotiating period, student autonomy and buy-in is balanced with the expectations of the instructor, leading to positive impressions for both the student and instructor (Hiller and Hietapelto, 2001).

2.4 Ungrading

"Ungrading" is slowly becoming a popular method of deemphasizing grades, in favor of enriching student engagement without the pressure of grades and assessments. Instructors who use ungrading reduce the number of graded assessments given to students to encourage students to focus on the learning experience instead of on the grade they might receive for their work (Blum, 2020). Students are often given a chance to grade themselves or their peers, allowing them to reflect critically and engage with course material instead of receiving a verdict from the instructor (Stommel, 2023). Though examples are limited, those who are implementing ungrading in their classes report greater student engagement, a class culture focused on understanding over completion, and lower levels of student anxiety about coursework (Masland, 2023). Potential drawbacks include a lack of effort from students due to a lack of accountability, a lack of student preparedness for the rigor of challenging work environments, and tensions between the student and instructor caused by different standards for nongraded work.

2.5 Specifications-based grading

Specifications-based grading combines aspects of several of these grading styles, leading to a better learning experience for students and instructors. Nilson and Stanny (2023) seminal book provides many positive outcomes of specs grading, which can be summarized by three major goals:

- Design a course that focuses on student learning rather than completion and achievement
- Remove ambiguity in assigning grades by providing clear expectations to students
- Balance a manageable workload for instructors while providing ample learning opportunities for students

Specs-grading focuses on giving students the opportunity to truly grasp difficult content while still holding them accountable for their learning by balancing the concessions and expectations given to students. Various aspects of specs grading came from the inspiration of other alternative grading methods such as those listed below:

- Specs grading divides course content into several small learning objectives, assessing student proficiency of each objective (Townsley, 2014). Like SBG, assessments are graded using rubrics which provide clear criteria that graders use to organize student work into various levels of proficiency such as "high pass," "pass," or "no pass" (Howitz et al., 2023).
- Students have multiple attempts to show proficiency in each objective and, like mastery-based learning, students only cover one at a time before assessment and advancement to the next objective (Bunnell et al., 2023). In specs grading, students have multiple (though not infinite) attempts on each assessment to reach passing criteria. However, in specs grading students advance to the next objective once the assessment due date passes or a student has used the maximum number of attempts, even if they do not achieve proficiency for the current outcome.
- Instead of students writing a contract, the instructor provides the contract detailing the assignments that must be completed for a student to earn a particular letter grade (Houseknecht and Bates, 2020). This maintains transparency of instructor expectations without requiring an instructor to endure the time-consuming meeting process with each student as is customary in contract-grading courses. Often, the contract will bundle objectives instead of assignments, grouping the outcomes into "Essential" and "General," designating how many of each objective must be passed for a student to earn a letter grade in the course (Howitz et al., 2021).

Specs grading has several strengths that make it a favorable option for instructors and students. The first is the focus on student learning; specs-based courses encourage students to build more regular study habits by breaking course material into smaller, more manageable outcomes that do not exceed the cognitive load of most students (Kishbaugh and Cessna, 2018). The emphasis on pairing smaller, more frequent assessments with timely instructor feedback helps students to focus on their mistakes and resolve learning gaps more quickly in comparison to traditional points-based courses that rely on large summative assessments to evaluate student learning (Schneider and Hutt, 2014).

Specs grading has also decreased grading issues between students and instructors by clearly articulating expectations and removing the ambiguity of partial credit, which is a commonly observed fault in classes with large enrollments and multiple graders (McKnelly et al., 2023). Instead, rubrics used to assess student performance use clear criteria, making it easy for graders to correctly identify whether the work submitted by a student has reached the desired level of proficiency (Howitz et al., 2023). Some assessments may require students to reach a certain benchmark – 80% is a common passing threshold when multiple-choice quizzes are used for assessment.

Other alternative grading methods provide students with multiple attempts on a given assessment, but the increased grading expectations that comes with multiple attempts can become exhausting for instructors (Bangert-Drowns et al., 1991). To mitigate this, a "token economy" is frequently implemented in a specs-grading classroom; students are provided a set number of "tokens" that can be spent to redo a failed assessment or extend a deadline on an assignment (Howitz et al., 2021; McKnelly et al., 2021). This allows the instructor to limit the number of retakes while still providing the freedom to reattempt any assessment that is particularly challenging. Additionally, the token economy removes the burden from faculty charged with arbitrating what is a valid reason for need of a retake or extension.

Various approaches have been taken toward the administering of a cumulative final. Some instructors choose to administer a summative

TABLE 1 Course learning objectives for CHEM 510.

Course objective	Specs module
Explain the unifying structure–property and structure-reactivity relationships upon which organic chemistry is based.	Across all modules
Describe molecular shape and stereochemistry and apply its effect on reactions.	EO1 (shape), EO2 (stereochemistry)
Propose reasonable mechanisms for organic transformations using curved arrow notation and apply these common reactivity pathways to unfamiliar reactions.	EO3
Predict structural effects on the acidity/basicity of an organic molecule.	EO4
Define and predict products for the major reactions involving nucleophiles and electrophiles, including nucleophilic addition, nucleophilic substitution at a carbonyl group and at a saturated carbon atom.	Across all GOs by functional group
Apply the numerous reactions that result in functional group transformations to the synthesis of organic compounds.	Across all GOs by functional group
Apply retrosynthetic analysis to the synthesis of organic molecules.	EO5, EO6
Evaluate and discuss current research in organic chemistry.	RP2, RP3

TABLE 2 Grading summary of intermediate organic chemistry prior to spring semester 2021.

Category and percent	Description
Exams (38%)	Midterm and final, each composed of five 60-min sections
Discussion Boards (31%)	6 collaborative problem sets
Quizzes (12%)	2 timed multiple choice/multiple answer questions
Project (10%)	Named Reaction slides presentation
Modules (9%)	2 60-min timed sections in long answer format

final as a necessary part of achieving a particular grade in the class and will include it in the grading contract (Bunnell et al., 2023). Other instructors choose to administer a final in an alternative format or as a type of competency-based assignment, offering a grade incentive to those who score well on it (Ring, 2017; Howitz et al., 2023). There is no standard way to approach final exams in specs grading.

While specs grading continues to gain traction in face-to-face instruction, the impact in an online setting has not been extensively studied. There have been reports of courses using specifications-based grading in a hybrid setting or in response to COVID-19 (Houseknecht and Bates, 2020; Noell et al., 2023), but few organic chemistry courses were initially designed to be online and utilize specifications-based grading.

3 Learning environment (setting, students, faculty); learning objectives; pedagogical format

Since 2010, the University of Saint Joseph (USJ) has offered Master of Science (MS) degrees in Chemistry and Biochemistry through fully online, asynchronous instruction. This asynchronous online format helps working professionals pursue a degree while continuing a normal working schedule, and over two thirds of the students enrolled in the program work in a full-time job. All courses in these programs are capped at 20 students. The first semester for students in both the MS Chemistry and Biochemistry programs includes CHEM 510 – Intermediate Organic Chemistry, a class designed to ensure that students are prepared for the rigor of the program by establishing a unified foundation of organic chemistry knowledge among all enrolled students. During our study, 97 students enrolled over 7 semesters in CHEM 510, making the average enrollment 14 students per semester, though class sizes varied from 7 students to the maximum enrollment of 20 students.

CHEM 510 covers the fundamental concepts of organic chemistry from undergraduate level classes, then builds upon them at the graduate level, preparing students for advanced coursework. Central topics covered in the course include organic structures, stereochemistry, mechanisms, acid/base reactions, selectivity, retrosynthesis, reactions of various functional groups, literature searching, proper citations, and named reactions. Table 1 lists the departmentally developed course learning objective associated with these topics. Historically, this course has been graded on a traditional points-based grading system with various categories of assignments, with the bulk of the points coming from discussion board assignments and the midterm and final examinations (see Table 2).

In Spring 2021, a new specifications-based grading structure was implemented in CHEM 510. The course material was divided into weekly modules arranged by content, including 6 Essential Objectives (EOs), 7 General Objectives (GOs), and 4 Required Projects (RPs). As detailed in Table 1, the course learning objectives were divided into foundational topics such as structure, stereochemistry, mechanisms, and acid/base chemistry (EOs 1–4); reactions of specific functional groups (GOs), tools for organic synthesis (EOs 5–6), chemical literature (RP2), and named reactions (RP3). RP1 contained the course introduction module and RP4 the final exam for the course.

A grading contract was developed by the instructor (Figure 1) to clearly communicate the number of objectives that must be met to earn each letter grade in the course. While this grade contract changed slightly during our study, the final iteration is described here. A grade of a B- is the minimum grade necessary to pass the course, requiring students to complete 5 EOs with a minimum assessment score of 80%,

Base grade	Minus	Standard	Plus
A	6EO + 4RP + 6GO	7GO	
		OR: %increase in RP4/RP1 >1 SD	
B	5EO + 4RP + 2GO	4GO	4GO and %increase in RP4/RP1 >1 SD
		OR: %increase in RP4/RP1 >1 SD	OR: 6GO
		OR: 6EO + 4RP + 3GO	OR: 6EO + 4RP + 4GO
Grades below here are <u>not</u> passing			
C	4EO + 3RP + 2GO	4GO	4GO and %increase in RP4/RP1 >1 SD
		OR: %increase in RP4/RP1 >1 SD	OR: 6GO
		OR: 4RP	OR: 4GO and 4RP
D	3EO + 2RP + 2GO	4GO	4GO and %increase in RP4/RP1 >1 SD
		OR: %increase in RP4/RP1 >1 SD	OR: 6GO
		OR: 4RP	OR: 4RP and 4GO
F	<3EO		

FIGURE 1
Grading matrix for intermediate organic chemistry for the 2023 summer semester.

a minimum of 2 GOs with an 80% score, and completion of all 4 RPs with a passing grade. To achieve a higher score, students could complete the 6th EO, additional GOs, or increase their score from the initial pretest (RP1) to the final post-test (RP4) by greater than one standard deviation. For example, to earn an A in the course, a student would need to meet the criteria for an A- (6 EO, 4 RP, and 6 GO) plus complete an additional GO or increase their score by greater than one standard deviation for pre−/post-test increase.

Weekly deadlines for the EOs and RPs were given, though students could work ahead if they chose. Students had six weeks to complete two to seven GOs with a suggested weekly schedule for students to follow culminating in a firm deadline for the GO section.

4 Results to date/assessment (processes and tools; data planned or already gathered)

To meet the proficiency criteria for EOs/GOs, students were offered three different forms of assessment:

- Timed multiple-choice quiz (two attempts)
- Timed written response test
- Open-ended tutorial presentation teaching how to solve a multi-step problem

Students could attempt one or more assessments before the due date. The quizzes were automatically graded, so students could complete two quiz attempts; whereas the exams and tutorials were manually graded, so they only had to be turned in before the due date.

After an assessment, the instructor would give the student feedback to inform them of any mistakes or misunderstandings displayed in the assessment. Quizzes generated instant feedback to students, while tests and tutorials required instructor feedback, which was usually provided within 24 h of completion. If the student did not reach the 80% minimum passing score, they were encouraged to read the feedback provided and strengthen their understanding of the material. Then, students were free to attempt another assessment before the due date or use a token for additional attempts (*vide infra*).

RPs were graded on completion only and were distributed throughout the semester. The RP modules included an introductory module with a pre-test, a section on finding, citing, and reading scientific literature, a named reaction project, and the post-test final exam.

After the first semester, a token economy was adopted as an opportunity for students to improve their standing in the course and remove the need for faculty to approve reasons for extensions or retries. Three tokens were provided to each student at the beginning of the semester. The instructor planned to modify the number of tokens if needed, but this number was adequate for most students and still low enough to encourage participation to earn more tokens. To spend a token, students would submit a Microsoft Form to request a deadline extension, additional quiz attempts beyond the two standard attempts, or changing assessment type after the due date (e.g., switching to tutorial after not passing the exam). Additional tokens could be earned by identifying errors in the course materials and completing course surveys (e.g., midterm course survey, office hours poll). Students also submitted their token earned requests through a Microsoft form. Both the "tokens earned" and "tokens used" were tracked in the LMS for easy student reference.

5 Discussion on the practical implications, objectives, and lessons learned

5.1 Results

Four semesters of data from the points-based grading system (two instructors) and eight semesters of data from the specifications-based grading system (one instructor that had also taught in the points-based system) were analyzed to see whether any general trends in course grades could be seen. When comparing the number of passing students between the two modalities, we see that more students earned a passing grade with specifications-based grading, with an increase of 10% more passing scores (Figure 2). The increase in passing scores in this introductory course also coincided with a decrease in students being put on probation or dismissed from the program for poor grades and increased retention of students beyond the first semester.

When every student passed in that first semester, it raised questions about potential grade inflation caused by the grading scheme. To compare student knowledge to their final letter grade, two 50-question pre- and post-tests were added to the course for the following seven semesters. Test questions covered all content from the 13 objectives in the course, and completion was mandatory for students to pass the course. Although it is a challenging assessment, students are told in advance that their score cannot negatively impact their grade, but that a good score may increase their final letter grade in the course (Figure 1).

We analyzed the pre-test and post-test scores completed by recent students in the course (Figure 3). All students who passed the class had an increase in their score from the pre- to post-test, with the A students showing larger increases than the B students. None of the students who failed the course had an increase in pre/post, and most of them did not take the post-test. In addition, 54% of passing students had more than a 50% increase in their score from the pre-to the post-test, a significant growth in knowledge over the course of one semester. We can see that a majority of passing students experienced substantial growth in their organic chemistry knowledge, with only a few students experiencing less than 25% increase. This collective increase in performance on the summative post-test makes a compelling case for the efficacy of specifications-based grading.

We noticed that students who experienced less growth in their test score generally entered the course with significant prior knowledge of organic chemistry, as indicated by higher-than-average pre-scores. Since this is an entry-level course taken by all newly admitted students, a wide range of abilities and experience is represented by the data. Students with stronger organic chemistry backgrounds had less room to grow and accounted for most of the lower growth scores. Since this course's goal is to bring all students to a similar level of proficiency in organic chemistry, varying degrees of growth are expected between students with varying degrees of prior knowledge in organic chemistry.

We also noted the general trend that students who completed more objectives during the semester - earning a higher letter grade in the course - also scored higher on the post-test and experienced more growth. This is promising, as it shows the efficacy of specs grading for increased student absorption and retention of course material. We believe this data confirms that the increase in passing scores from the traditional points-based grading scheme to the specifications-based grading scheme is due to increased student knowledge of the course materials.

When looking at the raw grades on the post-test considering the growth data, we see a few additional trends that may be noteworthy

FIGURE 2
Grade comparison between traditional points-based grading system and specifications-based grading system in intermediate organic chemistry at the University of Saint Joseph. A and B are passing scores and C and D are not. Grade data included is from Fall 2019-Summer 2023. $N = 183$.

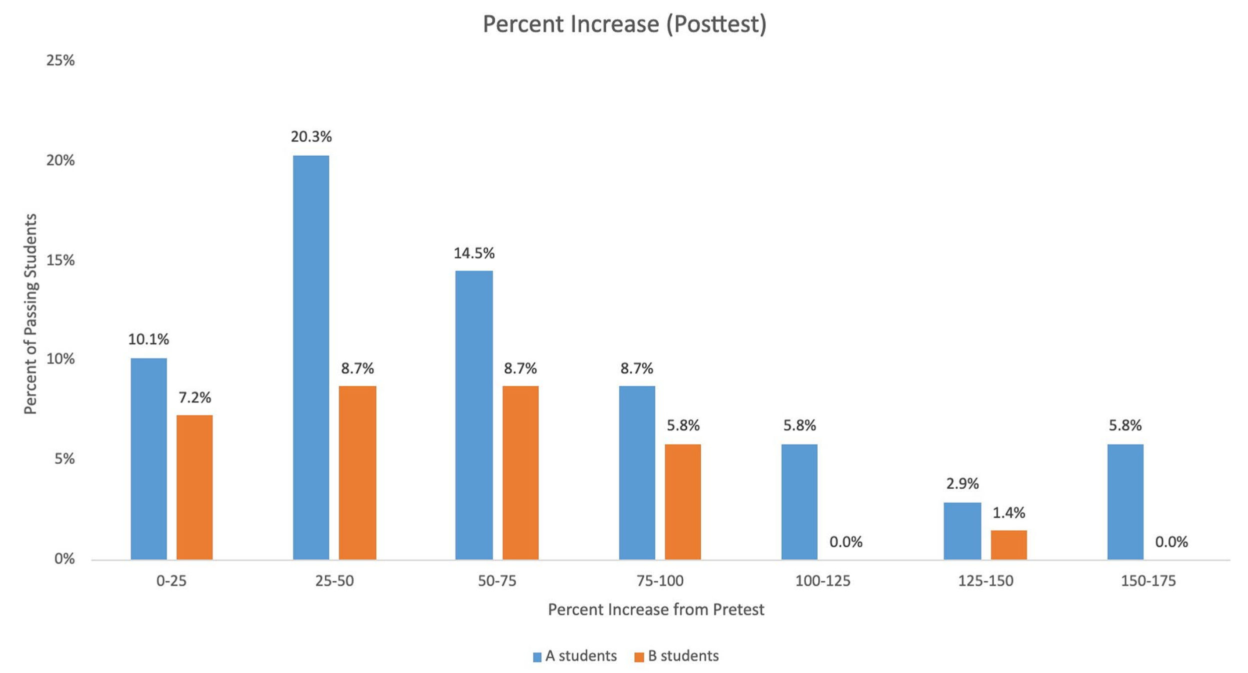


FIGURE 3
Increase in score from the pre-to the post-test for students with a passing grade in the course since the summer semester of 2021. *N* = 69.

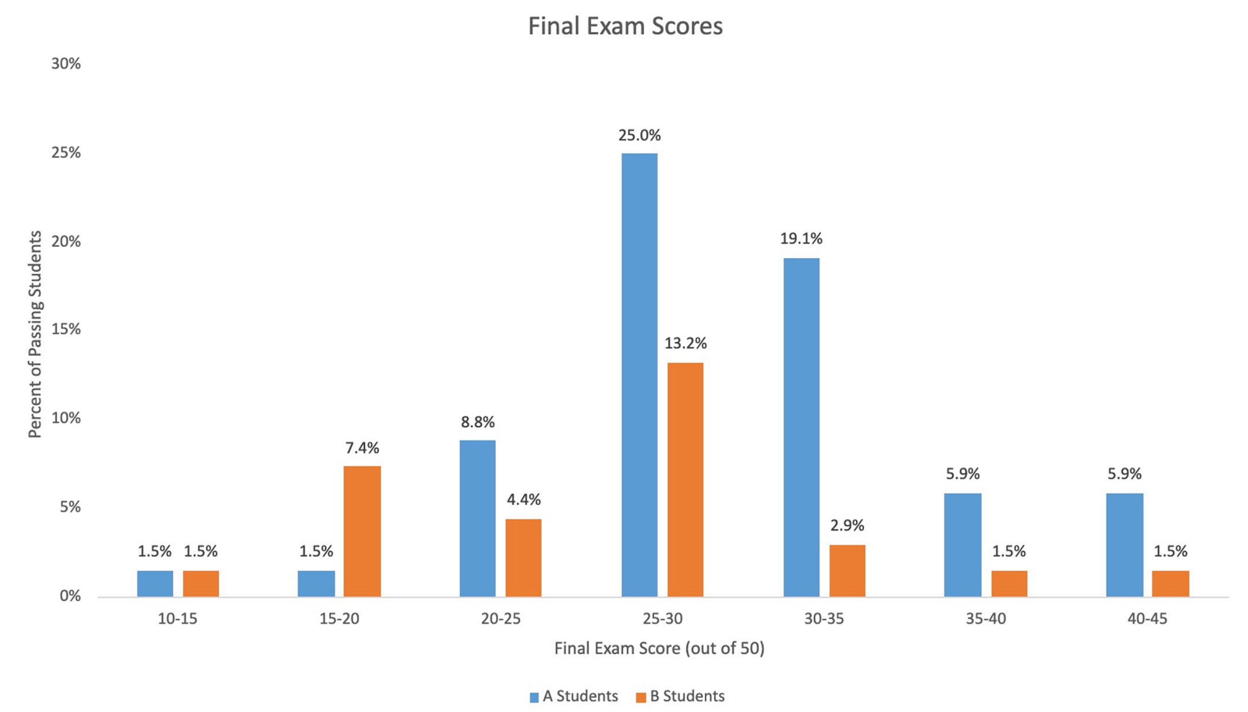


FIGURE 4
Post-test scores (out of 50) of students who passed the specifications-based style of the course. *N* = 69.

(Figure 4). Since students were required to complete a minimum of 5 EO modules and 2 of the 7 GO modules, passing students had to demonstrate proficiency in 54% of material covered in the course before taking the post-test. Since proficiency in an objective of this course is 80% accuracy, this means that passing students should have been able to answer a minimum of 43% of questions in the course correctly to earn the grade they received. As discussed above, 90% of passing students achieved this benchmark by answering a minimum of 22 questions correctly, with a vast majority scoring higher than that. The average score on the final was 29.125/50, 58.25%, which is the equivalent of a student achieving 80% proficiency on 9 course objectives. By this metric, student scores on the post-test are indicative of proficiency in organic chemistry.

It also should be noted that since many of these students did not take a long-form exam once during the course, the testing format of the post-test was likely unfamiliar to most participants. Additionally, since the post-test does not have much weight over the score a student earns in the class, it is possible that scores were lower due to a lack of motivation for students to adequately prepare for the test. Ultimately, the course was not designed with a focus on testing; it was focused on designing individual modules to maximize content exposure and retention. However, it is worth considering whether the delivery of the course or the thresholds for passing grades should be modified to improve the scores on the post-test.

5.2 Impressions

While data from the pre-and post-test provide a quantifiable indication of student learning, the numbers do not tell the whole story; student and instructor feedback reveal a more complete picture of the impact of specs grading on CHEM 510. To get feedback from students about their general impression of the course, an optional survey was offered during the week of midterm and final exams, which students could complete to earn an additional token for the course. Overall, student impressions were positive. The main strengths highlighted in these responses were centered around the multiple assessment options, reduced test anxiety, and the ability to improve upon previous mistakes in the learning process on the way to achieving proficiency.

While many students reported initial confusion about how final grades were distributed in the course, an overwhelming majority of students indicated that once the grading procedures were understood, they felt positively toward the specifications-based grading system. The most common comment made by students on the survey was the positive impact that multiple attempts and multiple assessment options had on the learning experience; students reported a greater sense of autonomy, reduced stress levels, and an enriched learning experience. When asked about the grading scheme, one student summarized a common sentiment: "It helps the student re-attempt the material no matter how low of a grade they achieved [on their first attempt]. It shows they need to brush up on a section they did not understand, [and builds confidence] down the road." (Response 37) Another student felt that they were "given the opportunity to actually learn the material without the pressure of trying once and failing" (Response 18) and appreciated the chance to review mistakes before attempting another assessment. These students asserted that the multiple attempts were essential to the reduced levels of stress they felt and added that they believed they learned more as a result.

In addition, several students had strong preferences for one method of assessment over another, with different students preferring each of the three assessment methods. One student wrote: "I feel much more comfortable with the tutorials because I can fully express in my own words what I think the goal of that module is." (Response 18) Another student preferred the instantaneous feedback offered by taking quizzes, writing "sometimes I would take the quiz feeling like I understood the concepts... but there was always feedback on the missed questions that [supported further learning]." (Response 24) Yet another student preferred exams because "the exams are structured in a way in which I am able to display a top-to-bottom understanding of the material." (Response 42) Each of these three students expressed preferences for different forms of assessment and felt positively that they had the freedom to choose the method that worked best for them. Though this is not a requirement of specifications-based grading, and is uncommon in chemistry courses, we believe that this is one of the greatest strengths of the specs grading scheme used in CHEM 510 and assert that it is an essential part to diversifying the learning experiences of the students.

When students' success and satisfaction are increased, a positive impression is left on the instructor as well, and many of the positive comments made by the instructor are identical to those made by the students. In addition, simplified grading processes in this specifications-based course benefited the instructor, who commented on the lack of ambiguity when grading quizzes, tutorials, and tests. Quizzes were automatically graded and provided instant feedback to students, reducing the amount of time the instructor spent grading and providing direct feedback to the student. Tests did not require additional time to grade, and the rubrics provided clear, unambiguous criteria for differentiating the student's level of proficiency. Even tutorials provide student work that – in our experience – usually provides a clear picture of a student's aptitude in the material, allowing for easy identification of proficiency. The instructor had previously felt frustrated that their feedback in traditional points-based courses was not looked at by the students but saw an increase in students accessing and applying the feedback in the specs-based system. Thus, the instructor could spend more time focused on giving feedback that would be used by students who would implement it.

The instructor appreciated the ways in which the course was designed to accommodate multiple learning styles from an instructional standpoint, as a wide variety of educational tools and resources were provided to the students. There were several passages from the textbook that students were expected to read, but the integration of lecture videos, instant chat features, and handwritten or electronic handouts allowed the instructor to generate a diverse learning experience for students. This diversity of instruction led to increased ways for students to engage with the material and contributed to the virtual classroom culture of inclusivity in assessment and instruction. However, providing a diversity of assessment and instructional methods is not without drawbacks: the initial design of the course was a considerable time commitment for the instructor, and grading multiple types of assignments submitted at inconsistent times throughout the week adds to the complexity of grading. In the semesters with more students (for example, 19 students in Fall 2022), the time spent on grading was increased and the instructor would not recommend exceeding the 20 student cap unless additional graders are used, or grading can be automated.

The instructor would also like to note that certain accommodations made in this class may not provide students the best chance to succeed in future classes taken later in the program, as no other course is graded in this way. The freedom offered to the students may come at a price: students who grow accustomed to this way of learning may not be prepared for the long discussion posts and high-stakes exams they will encounter in other courses later in their program. The token system added freedom for students to adjust their grade in a straightforward way but places the responsibility of processing token requests and adjusting due dates on the instructor and not the student. The autonomy experienced by students in this course is unique, as the flexibility and freedoms offered to students in future, traditional courses is significantly reduced.

6 Conclusion

Efforts to develop and implement sound grading practices continues to yield unique and promising alternatives to traditional points-based grading schemes. Specifications-based grading shows promise as a potential alternative that assigns student grades that reflect the student's understanding of content rather than effort given or points earned. We have found that in an online, asynchronous setting, specifications-based grading serves as a viable option for educators looking to improve student learning; our results show that specs grading courses can lead to improved assessment scores, targeted learning opportunities, and a more positive overall experience for students and instructors. We found that a critical piece in the success of specs grading is the opportunity for students to attempt assessment multiple times, and in our online asynchronous course, students found additional success by accessing multiple different assessment types, including short quizzes, long tests, and tutorial assessments. The positive impact that specs grading has had in CHEM 510 comes primarily from offering multiple types of assessment to students and providing multiple attempts to complete them. We would like to encourage other instructors considering specs grading for their online class structure to also consider including multiple assessment methods to reduce stress and increase learning opportunities for students, without adding excess time commitments to the instructor in the process.

7 Acknowledgment of any conceptual, methodological, environmental, or material constraints

Limitations to this report include the small sample size and the fact that the course design was being evaluated retroactively rather than prospectively. In addition, there was only one instructor implementing teaching this course and multiple minor changes were made each semester to improve the specs grading design.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the datasets could be provided upon reasonable request. Requests to access these datasets should be directed to SZ, szingales@usj.edu.

Ethics statement

The studies involving humans were approved by the University of Saint Joseph Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because this was a retroactive study looking at data collected by the instructor for formative feedback and no identifying information was used.

Author contributions

CM: Writing – original draft, Writing – review & editing. SZ: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

SZ thanks Renée Link for encouraging her to use specs-grading and providing continued support in the development and implementation of this course.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Atifnigar, H., Alokozay, W., and Takal, G. M. (2020). Students' perception of alternative assessment: a systematic literature review. *Int. J. Ling. Lit. Transl.* 3, 228–240.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., and Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Rev. Educ. Res.* 61, 213–238. doi: 10.3102/00346543061002213

Block, J.H., and Airasian, P.W. (1971) *Mastery learning: theory and practice*. New York: Holt, Rinehart and Winston.

Bloom, B.S. (1968). *Learning for mastery. Instruction and curriculum. Regional education laboratory for the Carolinas and Virginia, topical papers and reprints, number 1*. Regional education Laboratory for the Carolinas and Virginia, mutual plaza (Chapel Hill and Duke Sts. 1, 1–12. Available at: https://eric.ed.gov/?id=ED053419 (Accessed: 28 January 2024).

Blum, S. (2020). *Ungrading: Why rating students undermines learning (and what to do instead)*. Morgantown, WV, USA: West Virginia University Press.

Boesdorfer, S. B., Baldwin, E., and Lieberum, K. A. (2018). Emphasizing learning: using standards-based grading in a large nonmajors' general chemistry survey course. *J. Chem. Educ.* 95, 1291–1300. doi: 10.1021/acs.jchemed.8b00251

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., et al. (2016). A century of grading research: meaning and value in the most common educational measure. *Rev. Educ. Res.* 86, 803–848. doi: 10.3102/0034654316672069

Bunnell, B., LeBourgeois, L., Doble, J., Gute, B., and Wainman, J. W. (2023). Specifications-based grading facilitates student–instructor interactions in a flipped-format general chemistry ii course. *J. Chem. Educ.* 100, 4318–4326. doi: 10.1021/acs.jchemed.3c00473

Cain, J., Medina, M., Romanelli, F., and Persky, A. (2022). Deficiencies of traditional grading systems and recommendations for the future. *Am. J. Pharm. Educ.* 86:8850. doi: 10.5688/ajpe8850

Crucho, C. I. C., Avó, J., Diniz, A. M., and Gomes, M. J. S. (2020). Challenges in teaching organic chemistry remotely. *J. Chem. Educ.* 97, 3211–3216. doi: 10.1021/acs.jchemed.0c00693

Curley, B., and Downey, J. (2023). Implementation of alternative grading methods in a mathematical statistics course. *J. Stat. Data Sci. Educ.*, 1–11. doi: 10.1080/26939169.2023.2249956

Danielewicz, J., and Elbow, P. (2009). A unilateral grading contract to improve learning and teaching. *Coll. Compos. Commun.* 61, 244–268. doi: 10.58680/ccc20099471

Diegelman-Parente, A. (2011). The use of mastery learning with competency-based grading in an organic chemistry course. *J. Coll. Sci. Teach.* 40, 50–58. doi: 10.2505/3/jcst11_040_05&print=true

Donato, J. J., and Marsh, T. C. (2023). Specifications grading is an effective approach to teaching biochemistry. *J. Microbiol. Biol. Educ.*, e00236–e00222. doi: 10.1128/jmbe.00236-22

Guskey, T. R. (2001). 'Helping Standards Make the Grade', *ASCD*, 59. Available at: https://www.ascd.org/el/articles/helping-standards-make-the-grade.

Herman, J.L. (1992) *A practical guide to alternative assessment*. Association for Supervision and Curriculum Development, 1250 N: 1–98. Available at: https://eric.ed.gov/?id=ED352389 (Accessed January 28, 2024).

Hiller, T. B., and Hietapelto, A. B. (2001). Contract grading: encouraging commitment to the learning process through voice in the evaluation process. *J. Manag. Educ.* 25, 660–684. doi: 10.1177/105256290102500605

Houseknecht, J. B., and Bates, L. K. (2020). Transition to remote instruction using hybrid just-in-time teaching, collaborative learning, and specifications grading for organic chemistry 2. *J. Chem. Educ.* 97, 3230–3234. doi: 10.1021/acs.jchemed.0c00749

Howitz, W. J., Frey, T., Saluga, S. J., Nguyen, M., Denaro, K., and Edwards, K. D. (2023). A specifications-graded, sports drink-themed general chemistry laboratory course using an argument-driven inquiry approach. *J. Chem. Educ.* 100, 672–680. doi: 10.1021/acs.jchemed.2c00860

Howitz, W. J., McKnelly, K. J., and Link, R. D. (2021). Developing and implementing a specifications grading system in an organic chemistry laboratory course. *J. Chem. Educ.* 98, 385–394. doi: 10.1021/acs.jchemed.0c00450

Iamarino, D. L. (2014). The benefits of standards-based grading: a critical evaluation of modern grading practices. *Curr. Iss. Educ.* 17, 1–12. Available at: http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1234

Johnstone, A. H. (2010). You can't get there from here. *J. Chem. Educ.* 87, 22–29. doi: 10.1021/ed800026d

Kishbaugh, T.L.S., and Cessna, S.G. (eds) (2018) *Increasing retention of under-represented students in stem through affective and cognitive interventions*. Washington, DC: American Chemical Society (ACS Symposium Series).

Lindemann, D. F., and Harbke, C. R. (2011). Use of contract grading to improve grades among college freshmen in introductory psychology. *SAGE Open* 1:215824401143410. doi: 10.1177/2158244011434103

Marzano, R.J. (2010) *Formative assessment & standards-based grading*. Bloomington, IN: Marzano Research Laboratory.

Masland, L. C. (2023). Ungrading: the joys of doing everything wrong. *Zeal: A Journal for the Liberal Arts.* 1, 88–93.

McKnelly, K. J., Howitz, W. J., Thane, T. A., and Link, R. D. (2023). Specifications grading at scale: improved letter grades and grading-related interactions in a course with over 1,000 students. *J. Chem. Educ.* 100, 3179–3193. doi: 10.1021/acs.jchemed.2c00740

McKnelly, K. J., Morris, M. A., and Mang, S. A. (2021). Redesigning a "writing for chemists" course using specifications grading. *J. Chem. Educ.* 98, 1201–1207. doi: 10.1021/acs.jchemed.0c00859

Nilson, L.B., and Stanny, C.J. (2023) *Specifications grading: Restoring rigor, motivating students, and saving faculty time*. (*1st Edn.*). New York: Routledge.

Noell, S. L., Rios Buza, M., Roth, E. B., Young, J. L., and Drummond, M. J. (2023). A bridge to specifications grading in second semester general chemistry. *J. Chem. Educ.* 100, 2159–2165. doi: 10.1021/acs.jchemed.2c00731

O'Connor, K. (2002). *How to grade for learning: linking grades to standards*. (*2nd Edn.*). SkyLight professional development, 2626 South Clearbrook drive, Arlington Heights, IL.

Peters, R., and Buckmiller, T. (2014). Our grades were broken: overcoming barriers and challenges to implementing standards-based grading. *J Educ.Lead. Action* 2, 1–26. doi: 10.62608/2164-1102.1070

Ring, J. (2017). Confchem conference on select 2016 BCCE presentations: specifications grading in the flipped organic classroom. *J. Chem. Educ.* 94, 2005–2006. doi: 10.1021/acs.jchemed.6b01000

Schneider, J., and Hutt, E. (2014). Making the grade: a history of the A–F marking scheme. *J. Curric. Stud.* 46, 201–224. doi: 10.1080/00220272.2013.790480

Stommel, J. (2023) *Undoing the grade: Why we grade, and how to stop*. Denver, CO: Hybrid Pedagogy.

Sunasee, R. (2020). Challenges of teaching organic chemistry during covid-19 pandemic at a primarily undergraduate institution. *J. Chem. Educ.* 97, 3176–3181. doi: 10.1021/acs.jchemed.0c00542

Taylor, H. (1980). *Contract grading*. ERIC clearinghouse on tests, measurement, and evaluation, Educational Testing Service, Princeton, NJ. 1–28.

Townsley, M. (2014). 'What is the Difference between Standards-Based Grading (or Reporting) and Competency-Based Education?', *Competency Works Blog*, 11 November. Available at: https://aurora-institute.org/cw_post/what-is-the-difference-between-standards-based-grading/ (Accessed January 28, 2024).

frontiers | Frontiers in Education

TYPE Perspective
PUBLISHED 15 April 2024
DOI 10.3389/feduc.2024.1394042

Check for updates

OPEN ACCESS

EDITED BY
Jay Wackerly,
Central College, United States

REVIEWED BY
Matthew Mio,
University of Detroit Mercy, United States
Nicole Graulich,
University of Giessen, Germany

*CORRESPONDENCE
Jalisa H. Ferguson
✉ fergusj@eckerd.edu

RECEIVED 29 February 2024
ACCEPTED 29 March 2024
PUBLISHED 15 April 2024

CITATION
Ferguson JH and Bonner LA (2024) Ungrading in organic chemistry: students assessing themselves and reflecting on their learning.
Front. Educ. 9:1394042.
doi: 10.3389/feduc.2024.1394042

Ungrading in organic chemistry: students assessing themselves and reflecting on their learning

Jalisa H. Ferguson* and Lisa A. Bonner

Department of Chemistry, Eckerd College, St. Petersburg, FL, United States

The focus on grades has diminished the focus on learning. One strategy that aims to return students' attention to what they are actually learning (and not just earning) is ungrading. Ungrading is thought of as any strategy in which instructors do not assign a number or letter grade to students' assignments and assessments. Instead, faculty may (1) provide thorough feedback and engage in dialogue with students about their work, and perhaps, (2) allow students to assign their own grade. Whichever style of ungrading they choose, the scholars that have been forging the path for ungrading come from a variety of fields and perspectives, including STEM instructors in more recent years. The focus on incorporating ungrading practices into the organic chemistry curriculum provided here is adapted from a variety of practitioners, and especially the foundational work of chemistry professor Clarissa Sorensen-Unruh. In addition to discussing the current ungrading practices in various fields, we will use this perspective article to share our own experience with and lessons learned from beginning to incorporate ungrading in the undergraduate organic chemistry curriculum, both as it relates to the implementation of the practice and our own perceptions of the student experience and learning outcomes. Ultimately, the goal is to allow students to see the significance of the process of learning and to engage in some metacognitive work that they can apply to different assignments, whether in our class or not. If we want students to focus on learning, perhaps they should do the grading themselves.

KEYWORDS

grading, ungrading, organic chemistry, chemistry education research, metacognition

1 Introduction

Attention toward assessment has been increasing in recent years for multiple reasons (Digital Science, 2018 and **Supplementary Figure 1**). As demands on faculty time continue to grow, there are attempts to decrease instructor time and energy spent grading. It is also important to help students learn that their grade in a particular class should not be considered a reflection of how much time they spent reading, studying, or working practice problems. This common misconception, that time spent is automatically equivalent to knowledge obtained, and thus grade earned, is one that many instructors report as an issue for their students (Kemmerer, 2014; Carver et al., 2017). Students spend time outside of class on their coursework, but it is the quality and focus of that study time, like the use of self-testing, that is typically positively associated with grade outcomes (Hartwig and Dunlosky, 2012; Fergus, 2022). Students may understand that higher quality studying leads to better grade outcomes, but they don't always know how to implement it (Tomes et al., 2011; Fergus et al., 2021; Carpenter et al., 2024). To help *students* better assess their own learning, we must take a more meaningful approach to how *instructors* assess their learning.

Several reviews offer historical highlights of traditional assignments of alphanumeric grades. Cureton's article from 1971 indicates that "percentage grading," established as early as the mid-1800s, occurs when instructors determine what proportion of the maximum grade (usually out of 100) a student should earn on a given essay or examination (Cureton, 1971). By the beginning of the 20th century, due in part to increased student populations and a need to communicate student abilities between institutions, grading on a letter scale gained popularity as a universal standard (Schinske and Tanner, 2014). Furthermore, using a normal distribution curve for assigning letter grades was accepted as a meaningful measure of student abilities, while minimizing subjectivity of individual instructor's marks (Schinske and Tanner, 2014).

Though he wasn't the only skeptic of the time, in 1968, Bloom questioned the validity and "sacredness" of the normal distribution curve for grades and encouraged instructors to use strategies that improved mastery of learning (Bloom, 1968). Referencing Carroll's *A Model of School Learning* (Carroll, 1963), Bloom says "if the students are normally distributed with respect to aptitude, but the kind and quality of instruction and the amount of time available for learning are made appropriate to the characteristics and needs of *each* student, the majority of students may be expected to achieve mastery of the subject" (Bloom, 1968). As instructors have attempted to modify their teaching methods to improve student learning, they have also adjusted their expectations about the normal distribution of the number of As and Fs granted. This change in philosophy has led many to question what grades really mean - or at least, what they're supposed to mean - and alter their approaches to assigning grades.

1.1 Support for ungrading

Even well-known proponents of the strategy tend to be fairly vague when defining ungrading, perhaps trying not to exclude practitioners willing to disrupt conventional grading techniques. Jesse Stommel says it means "raising an eyebrow at grades as a systemic practice, distinct from simply "not grading"" (Stommel, 2023). Since the 2020 release of her edited book *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead)*, Susan D. Blum has taken the perspective that the term should be an umbrella term, describing a variety of approaches (Blum and Kohn, 2020). Katharine E. Johanesen takes it further, saying ungrading describes "a variety of practices that decenter grading in a class. [It] often involves eliminating or reducing numeric scoring in favor of descriptive feedback and/or reflection" (Johanesen et al., 2023). Ungrading can involve allowing students to assign their own grade at the end of a course, giving students tools to grade their own individual assignments, or anything else in between and beyond.

Kohn compiles several issues with assigning traditional grades, ranging from adverse effects on student learning and motivation, to the negative effect of grades on the relationship between student and teacher (Kohn, 1999). Compared with traditional grading, ungrading provides some hope for instructors that are looking to improve students' intrinsic motivation. Researchers like Butler, Green, and Grant have observed that using grades as a source of external motivation might push students to work for a better grade, not necessarily to learn the material, or engage in classroom assignments and discussions (Butler, 1988; Grant and Green, 2013). Ungrading can be used to help students focus on what they need to learn - or what they *have* learned - as opposed to assigning an alphanumeric grade.

Importantly, ungrading has the potential to minimize or eliminate inequity in college classrooms, though the true benefit seems yet undetermined. In *Undoing the Grade: Why We Grade and How to Stop*, Jesse Stommel argues that the current approach to grading is already harmful, and that any new approaches, such as ungrading, should be designed with flexibility, care, and structure (Stommel, 2023). Supiano's 2022 *Chronicle* article discusses math professor Robert Talbert's question of whether ungrading makes the equity gap worse (Supiano, 2022). Notably, removing grade "guideposts" may make it harder for marginalized students to know where they stand. But students who are already at an advantage, by having college-educated parents for example, are having to learn a new system just like everyone else, potentially leveling the playing field. In their recent book *Grading for Growth: A Guide to Alternative Grading Practices That Promote Authentic Learning and Student Engagement in Higher Education*, Clark and Talbert discuss various benefits of alternative grading, including ungrading, like providing clear standards, allowing for reassessment without penalty, and "focusing on eventual understanding", which is expected to minimize biased grading (Clark, 2023). With appropriate intention and care, instructors may find ways to enhance equity in their classes with the help of ungrading.

1.2 Examples of ungrading strategies in STEM

We were inspired by STEM educators who have recently begun exploring the ungrading landscape in their courses. Riesbeck and Cangialosi teach computer science and biology, respectively, yet their approaches to ungrading are quite similar - and they've been in practice for over 20 years. Essentially, they both provide feedback, without assigning grades, to student assignments and allow the student to resubmit their work with revisions and improvements (Riesbeck, 2017; Cangialosi, 2020). Cangialosi goes a step further by asking students to complete self-assessments to describe their activities and how successful they think they've been. She reports that this is a rather illuminating experience, as students are not typically equipped to accurately translate their efforts into an appropriate grade. For instance, she mentions that a student who had not yet started on a project gave themselves a "C" grade, instead of a failing grade (Cangialosi, 2020). This disconnect allows her to have conversations with her students about what grades actually mean, while still giving them the agency to improve based on feedback she provides.

More recent STEM ungraders have leaned into the element of student self-reflection and assessment. Katharine Johanesen's excellent 2023 report describes a number of ungrading iterations in her geosciences courses that she has made over the years, in conversation with other examples in the field (Johanesen et al., 2023). Because of the clear emphasis on self-reflection and structures provided for students to assess themselves, she reports that students felt less anxious and more supported by the end of the semester.

In chemistry, two prominent ungrading figures are Courtney Sobers and Clarissa Sorensen-Unruh (Jarvis, 2020). Sobers describes her experience with ungrading in a general chemistry II laboratory class: she gives feedback, allows for student revision and self-assessment, and confirms that their revisions are correct (Rutgers School of Arts & Science-Newark, 2021). Sorensen-Unruh has incorporated ungrading to varying degrees in several of her general and organic chemistry classes, including requiring students to grade their own mid-term exams (Jarvis, 2020).

Sorensen-Unruh's motivation for ungrading resonated with us as well, particularly her remarks in her online blog and book chapter in Blum's *Ungrading*: "I came to believe grading undermines learning daily by focusing student interest on achievement and not on learning" (Sorensen-Unruh, 2020). We encourage interested parties to read the full book chapter, but to briefly summarize her approach, students took their mid-term exams as normal, Sorensen-Unruh would write feedback on each exam and keep track of scores assigned for specific problems in her own private spreadsheet. She then returned the exams to students, and they would grade themselves. Sorensen-Unruh accepted the student's grade as long as it was within a standard deviation of her own assessment. Further, the more accurate the student was in their self-assessment, the more likely she was to assign extra credit, incentivizing accurate corrections and grade assignment. Sorensen-Unruh's experience reflects a cooperative and amicable relationship between faculty and students as it relates to their grade assignment. She highlights the importance of trust in the relationship, but also accountability and structure. Even her students recognize the value of ungrading, with one saying, "I feel like ungrading focuses on the higher level concerns and therefore encourages a deeper understanding of the material, despite some potential issue with lower level ideas" (Sorensen-Unruh, 2020). Getting students to understand the significance of higher order learning, as opposed to nitpicky point determination, is exactly the sort of experience we were hoping to replicate in our own work.

2 Our experience

2.1 Background

Like many others, we were interested in making some pedagogical changes in aftermath of the Covid-19 pandemic, which prompted our foray into the relatively unexplored ungrading territory. For the Fall 2020 semester, our institution offered one or two condensed courses at a time to help mitigate the spread and impact of Covid-19 on our campus. To facilitate a productive learning environment given much longer lecture periods of 3 h, we flipped all organic chemistry courses in both the first and second semester. Students could learn the material through recorded lecture videos before class, and we spent lecture time together solving relevant practice problems. After this first academic year of teaching the courses this way, we wanted to include more avenues for students to take control of their own learning, rather than relying on our assigned grades. We also sought to encourage students to reflect on (or at least *acknowledge*) the feedback we provided them, which often went unread. In this new era of intensive active learning in the course, it was important that students fully realized their power in both learning and assessing their own learning - ungrading provided a helpful avenue for this metacognitive work (Carpenter et al., 2024).

2.2 First iteration

Like many others report regarding their first experience with ungrading, we were apprehensive to get started. Inspired by Sorensen-Unruh's incorporation of the strategy in her own chemistry classes, particularly her desire to "divorce grades from feedback" (Jarvis, 2020), we decided to start small with the four in-class quizzes we provide over the course of a semester. The model that we follow for quizzes is adapted from Sorensen-Unruh's approach with exams (Sorensen-Unruh, 2020). In our initial iteration of ungrading, students took a 15-min quiz in class, which was immediately collected by their instructor. After class, we scanned the original copies of the completed quizzes and returned them to students the same day, as indicated in the instructions (**Figure 1A**). Using an available rubric on our learning management system, students graded themselves over the course of several days, and turned in their completed ungraded quiz at the next class meeting. We assessed their ungraded quizzes, including confirming a numeric grade, and returned them the following class meeting. Importantly, the provided rubric did not have correct answers. Instead, it included how many (if any) partial points to assign based on the correct answer(s) given by students. The intention was that students would review their notes, the textbook, and/or visit office hours to confirm their answers or gain insight into how to solve the problems. This approach was fairly manageable for us with class sizes of approximately 20–35 students.

2.3 Revisions

Reflecting after the first semester of offering these ungraded quizzes, we identified a few modifications that would provide a bit more structure to students as they did this reflective work, many of them for the very first time (**Figure 1B**). First, we made the rubrics more detailed (**Supplementary Figure 2**). In addition to providing the points students should assign themselves for various parts of the answers, we directed students to the appropriate sections of the required textbook and/or included helpful context for solving the problems. We had found that students needed more instruction to determine whether or not they even answered the questions correctly.

A

CH221 FA21 Organic Chemistry I
Dr. Bonner, Dr. Ferguson, and Dr. Grove
Quiz 1
September 16, 2021

Name ____________________
First initial and last name only

Pledged ____________________

<u>INSTRUCTIONS</u>: This quiz will be scanned and saved after class so that we have a copy of the original. Please pick it up from the ledge outside of CMLS 139 this afternoon. After picking it up, grade yourself according to the Quiz 1 Rubric on Moodle and turn it in to your professor at the beginning of your next class (Tuesday September 21, 2021).

B

CH221 FA23 Organic Chemistry I
Dr. Bonner, Dr. Ferguson, and Dr. Grove
Quiz 1
September 14, 2023

Name ____________________
First initial and last name only

Pledged ____________________

<u>INSTRUCTIONS</u>: You will have a choice for grading. Read the descriptions below and circle one option before turning in your completed quiz.

Please circle one: Option 1 or Option 2

Option 1: *Graded by your professor.* When you turn this quiz in, that's all that's required of you. You will receive your graded quiz during the next class period. This option does not have an opportunity for additional bonus points.

Option 2: *Ungraded by you, with reflections and corrections for EACH question, even the ones you answered correctly.* Please pick it up from the ledge outside of CMLS 139 this afternoon. After picking it up, grade yourself according to the Quiz 1 Rubric on Moodle. *Additionally, you need to provide corrections and a reflection (using words), explaining the process for solving the problem and making improvements for any missed questions. This reflection may be typed or handwritten.*
Turn the ungraded quiz and your reflection in to your professor at the beginning of your next class (**Tuesday September 19, 2023**).

A well-graded quiz with thorough reflection will receive up to two additional bonus points. A reflection should include your problem-solving process, even for questions you answered correctly.

FIGURE 1
Organic Chemistry I ungraded quiz instructions for **(A)** Fall 2021 and **(B)** Fall 2023.

Second, students were allowed to receive up to two additional bonus points for providing appropriate corrections and reflections - even if they answered all questions correctly. Recognizing mistakes and correcting them is an obviously helpful tool that we have encouraged for years, but now those corrections are incentivized in a way that they weren't before. The reflections (recommended by colleague Sarah K. Zingales) reinforce the notion that having students see their mistakes, make corrections, and meaningfully reflect on their problem-solving process, *and how that process might need to change*, is so valuable for the metacognitive work we need students to engage in.

Lastly, we realized that we needed to give students more choice. Most, if not all, students enrolled in either semester of Organic Chemistry are also taking at least one or two other science courses with associated 3-h labs. Many of our best students are also teaching assistants, tutors, resident advisors, and/or working in the on-campus, student-run emergency response team. Although we recognize the impact of making corrections and self-reflection on their own learning, sometimes time constraints limit a student's ability to meaningfully engage in this work. Students are allowed to opt-out of ungrading for each quiz if they so choose, and are thus ineligible for the two bonus points. Most students still choose to ungrade themselves, anywhere from 80 to 100% of the time.

These three changes were implemented in the very next semester, Spring 2022, during the Organic Chemistry 2 lecture courses, and we still use them in Spring 2024.

3 Discussion and future directions

3.1 Impact on faculty time

We feel the need to be very clear about one thing - this version of ungrading does not reduce the grading burden on faculty. In fact, in our experience, it takes *longer* to evaluate students' ungraded efforts and reflections than it would to simply grade the quizzes. In order to ensure the students actually graded themselves correctly, we essentially grade them twice at once - both the original answers and the students' assessment of their answers. Reviewing student reflections is also time-intensive, but we find it to be a worthy component to help achieve our metacognitive goals.

Ultimately, if your goal is to reduce the amount of time you spend grading, use a different ungrading approach. If your goal is to help students meaningfully reflect on their own learning, this could be a valuable strategy for you to try.

3.2 Impact on student outcomes

From our viewpoint, most students tolerate the ungrading strategy and are willing to participate in the process of grading themselves. The most well-prepared students do well when making quiz corrections and reflecting on their thought processes, and those insights seem to last as they prepare for exams. We suspect that a correlation exists between student accuracy in ungrading quizzes and exam performance, but we do not yet have the data to verify this suspicion. As such, we have concerns that the students who were already going to perform well in the class are the ones that typically perform better on their ungraded quizzes, as well as their reflections. Incentivizing accurate corrections and reflections with bonus points does seem to encourage the average and lower performing students, but it is clear that the primary beneficiaries are the top students.

A challenge we have faced is that some students will over-assess their performance on assignments or the course, giving themselves a higher grade than they actually earned, despite our inclusion of the detailed rubric with instructions on how to find correct answers. In our experience, the discrepancy is less likely because of an attempt to "game the system", and more likely a result of not attributing the necessary time and effort toward accurately assessing themselves. It is unclear whether they do not recognize their mistakes, do not refer back to the required resources, or simply underestimate the amount of time and effort it takes to critically reflect on their work.

We believe that this process of ungrading and self-reflection helps students learn how to learn, a valuable skill that can be applied throughout their college and professional careers.

3.3 Future directions

As ungrading becomes increasingly prominent in the academic and popular literature, we will continue to look to our STEM peers for insights to refine the process in a way that works for us and our students. As we seek to develop and identify improvements to our ungrading process, it is critical that we consider the important balance between giving students ownership of their grade, and holding them accountable for learning the content. Organic chemistry is a gateway course into upper-level chemistry courses, as well as many health-related professional graduate programs. Making sure that students use ungraded assignments as opportunities to reflect on their learning, and to think about their thinking, is critical to their success in the future.

In the future, we would like to investigate the longitudinal outcomes of applying this assessment structure. As mentioned, we expect a correlation between ungrading accuracy and exam performance. Perhaps other correlations may exist, such as grade outcomes in the second semester of organic chemistry or other upper-level chemistry courses. It will be important for chemistry education researchers to compare course outcomes for students who reflect on their answers, both correct and incorrect ones, with those who just make corrections on original answers. If differences are present, we could evaluate whether the differences can be attributed to student effort, faculty expectations, prior chemistry preparation, or something else. Regardless of the findings, it is important that we learn more about ungrading, and that we don't give up the practice.

Data availability statement

The original contributions presented in this study are included in this article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

JF: Writing – review and editing, Writing – original draft. LB: Writing – review and editing, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by Eckerd College for the publication of this article.

Acknowledgments

This work is dedicated to David Grove.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1394042/full#supplementary-material

References

Bloom, B. S. (1968). *Learning for mastery. Instruction and curriculum. Regional education laboratory for the carolinas and virginia, topical papers and reprints, number 1. Evaluation comment, 1*. Washington, DC: Office of Education.

Blum, S. D., and Kohn, A. (2020). *Ungrading: Why rating students undermines learning (and what to do instead)*, 1st Edn. Morgantown: West Virginia University Press.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluations of interest and performance. *Br. J. Ed. Psychol.* 58, 1–14. doi: 10.1111/j.2044-8279.1988.tb00874.x

Cangialosi, K. (2020). *A special kind of hell, 2020* Available online at: https://karencang.net/teaching/a-special-kind-of-hell/ (accessed February 29, 2024).

Carpenter, T., Fritz, J., and Penniston, T. (2024). "Banking on adaptive questions to nudge student responsibility for learning in general chemistry," in *Data analytics and adaptive learning*, eds P. D. Moskal, C. D. Dziuban, and A. G. Picciano (New York, NY: Routledge), 189–210. doi: 10.4324/9781003244271-14

Carroll, J. B. (1963). A model of school learning. *Teach. Coll. Rec.* 64, 723–733. doi: 10.1177/016146816306400801

Carver, L. B., Mukherjee, K., and Lucio, R. (2017). Relationship between grades earned and time in online courses. *Online Learn. (Newburyport, Mass.)* 21:303. doi: 10.24059/olj.v21i4.1013

Clark, D. (2023). *Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education*. 1st Edn. Routledge: Taylor & Francis. doi: 10.4324/9781003445043-5

Cureton, L. W. (1971). The history of grading practices. *Meas. Educ.* 2, 1–8.

Digital Science (2018). *Dimensions analytical views*. Available online at: https://app.dimensions.ai (accessed March 21, 2024).

Fergus, S. (2022). Are undergraduate students studying smart? Insights into study strategies and habits across a programme of study. *J. Univ. Teach. Learn. Pract.* 19, 110–127. doi: 10.53761/1.19.2.8

Fergus, S., Heelan, A., Ibrahim, S., Oyman, H., Diaz-de-Mera, Y., and Notario, A. (2021). Insights into study strategies and habits: A study with undergraduate students in Spain and the U.K. *J. Chem. Ed.* 98, 3084–3089. doi: 10.1021/acs.jchemed.1c00126

Grant, D., and Green, W. B. (2013). Grades as incentives. *Empir. Econ.* 44, 1563–1592. doi: 10.1007/s00181-012-0578-0

Hartwig, M. K., and Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychon. Bull. Rev.* 19, 126–134. doi: 10.3758/s13423-011-0181-y

Jarvis, C. L. (2020). *Chemistry educators try 'ungrading' techniques to help students learn*. Washington, DC: Chemistry & Engineering.

Johanesen, K. E., Claiborne, L. L., Falk, E. S., Parsons Hubbard, K., Kohfield, K. E., Nadin, E. S., et al. (2023). Common-sense teaching for the 2020s: Ungrading in response to covid-19 and beyond. *J. Geosci. Educ.* 71, 1–16. doi: 10.1080/10899995.2023.2259784

Kemmerer, K. P. (2014). Rigor, grades, support, and the amount of time students spend outside of class: A comparison of full- and part-time faculty in an entertainment and music business program. *MEIEA* 14, 123–156. doi: 10.25101/14.5

Kohn, A. (1999). From degrading to de-grading. *High Sch. Magazine* 6, 38–43.

Riesbeck, C. (2017). *20 years gradeless: Having my cake and eating it too, teachers going gradeless*. Available online at: https://growbeyondgrades.org/blog/2017/07/10/20-years-gradeless/ (accessed February 29, 2024).

Rutgers School of Arts & Science-Newark (2021). *Chart a course: Ungrading with courtney sobers [Video]*.

Schinske, J., and Tanner, K. (2014). Teaching more by grading less (or differently). *CBE Life Sci. Educ.* 13, 159–166. doi: 10.1187/cbe.cbe-14-03-0054

Sorensen-Unruh, C. (2020). "A STEM ungrading case study: A reflection on first-time implementation in organic chemistry II," in *Ungrading: Why rating students undermines learning (and what to do instead)*, ed. S. D. Blum (West Virginia: West Virginia University Press), 140–160.

Stommel, J. (2023). *Undoing the grade: Why we grade, and how to stop*. Madison WI: Hybrid Pedagogy Inc.

Supiano, B. (2022). The unintended consequences of 'ungrading': Does getting rid of grades make things worse for disadvantaged students? *Chron. Higher Educ.* 68:46.

Tomes, J. L., Wasylkiw, L., and Mockler, B. (2011). Studying for success: Diaries of students' study behaviours. *Educ. Res. Eval.* 17, 1–12. doi: 10.1080/13803611.2011.563087

Frontiers in Education

TYPE Curriculum, Instruction, and Pedagogy
PUBLISHED 17 April 2024
DOI 10.3389/feduc.2024.1384129

OPEN ACCESS

EDITED BY
Sarah Zingales,
University of Saint Joseph, United States

REVIEWED BY
Fun Man Fung,
National University of Singapore, Singapore
Jalisa Ferguson,
Eckerd College, United States

*CORRESPONDENCE
Lyniesha Wright Ward
wardlyn@iu.edu

RECEIVED 08 February 2024
ACCEPTED 12 March 2024
PUBLISHED 17 April 2024

CITATION
Ward LW, Spencer D, Chauhan D and Oliver-Hoyo M (2024) Lessons learned: the use of an augmented reality application in organic chemistry laboratories.
Front. Educ. 9:1384129.
doi: 10.3389/feduc.2024.1384129

Lessons learned: the use of an augmented reality application in organic chemistry laboratories

Lyniesha Wright Ward[1]*, Dan Spencer[2], Daivik Chauhan[3] and Maria Oliver-Hoyo[4]

[1]Department of Chemistry and Chemical Biology, Indiana University Indianapolis, Indianapolis, IN, United States, [2]Department of Distance Education and Learning Technology Applications (DELTA), North Carolina State University, Raleigh, NC, United States, [3]Department of Informatics, Indiana University Indianapolis, Indianapolis, IN, United States, [4]Department of Chemistry, North Carolina State University, Raleigh, NC, United States

Immersive technologies such as augmented reality (AR) have the potential to enable students to remediate invalid assumptions about molecular structure through visualizing site-specific, non-observable chemical processes. In this study, we explore how this technology-embedded instruction impacted student perceptions and experiences in a collaborative face-to-face and independent remote organic chemistry laboratory, the latter of which occurred during the COVID-19 pandemic. While we acknowledge the emotional toll of the pandemic, it afforded a unique opportunity to compare the differences in implementation when covering the same material. We used a novel AR mobile application, H NMR MoleculAR, and a complementary worksheet to support students' understanding of proton nuclear magnetic resonance (^{1}H NMR) spectroscopy. We gathered data using a mixed-methods pre-post survey about students' perceptions and experiences in the remote and in-person environments. There were differences in student user experience and perceptions of NMR knowledge, with face-to-face students showing more positive rankings. Although lower than those in face-to-face environments, perceptions of the remote environment remained neutral or positive for all measures. There were no differences in the reported number of challenges faced, but there were unique challenges in the remote learning environment. Our findings illuminate the complexity of factors that must be considered when implementing novel technologies into instruction in face-to-face and remote environments. We conclude by describing concrete lessons learned and considerations for researchers and instructors leveraging augmented reality.

KEYWORDS

organic chemistry, augmented reality, laboratory learning, remote learning, proton nuclear magnetic resonance, chemistry education research

1 Introduction: background and rationale

In recent years, there has been an increase in the use of augmented reality (AR) to support student understanding of representations and visualizing molecules and chemical phenomena across a variety of topics (Behmke et al., 2018; Tee et al., 2018; Sung et al., 2020; Abdinejad et al., 2021; Mazzuco et al., 2022), though it still is not widely adopted (De Lima et al., 2022). AR affords the ability to overlay 3D virtual images in a real physical space, allowing for more comprehensive interactions with molecular structures, incorporating depth and stereoscopic perception into the understanding of chemical phenomena (Behmke et al., 2018; Goddard et al., 2018), and enabling the visualization of site-specific, non-observable chemical processes (Huwer et al., 2018).

AR-embedded instruction has been known to advance the affective domain by increasing motivation, interest, and confidence in learning; the cognitive domain by supporting learning, improving performance, and increasing knowledge retention; and the psychomotor domain by developing spatial skills and allowing the visualization of abstract concepts (Cheng and Tsai, 2013; Mazzuco et al., 2022). Furthermore, the ability to use AR on a mobile device means that students can learn in multiple locations. Combined, these AR affordances can support students and develop their understanding of chemistry.

Leveraging the affordances of AR would be especially advantageous with chemistry content, such as ^{1}H NMR spectroscopy, that demands the visualization of molecular spectra but is known to be challenging for students (Anderson et al., 2020). Studies have shown that textbooks (Anderson et al., 2020) and instructors (Connor and Shultz, 2018) struggle to scaffold ^{1}H NMR instruction adequately, suggesting additional resources are needed to supplement these barriers. Resultingly, we leveraged an AR application to provide technology-embedded instruction about ^{1}H NMR. Herein, we discuss how we implemented an activity that consists of H NMR MoleculAR, a novel AR application, and an accompanying worksheet to support students in understanding the concepts and problem-solving processes underlying ^{1}H NMR spectroscopy. We describe the lessons from incorporating this activity in face-to-face and remote learning environments. Specifically, we answer the following questions:

1 How do students experience the activity in each environment (RQ1)?
2 How do students perceive the activity in each environment (RQ2)?
3 To what extent does the activity adjust students' perceptions of the importance of visualizations for learning chemistry and their knowledge of ^{1}H NMR (RQ3)?

2 Pedagogical frameworks underlying the educational activity innovation

The H NMR MoleculAR application (the App) uses static and dynamic visualizations to render virtual 3D representations of molecular structure, molecular orbitals, and electrostatic potential maps over unique targets that are 2D images of molecules or spectra. The content in the App and its associated worksheet is structured using the Compare-Predict-Observe-Explain (CPOE) cycle and the contrasting cases framework (Alfieri et al., 2013; Graulich and Schween, 2018). Students are given structures that differ in one important feature (a contrasting case). They are expected to compare these structures, make a prediction about how the difference in features will be reflected on the spectrum, observe the spectra, and then explain the underlying concepts or principles that connect the structural features to the spectral features (see an example in the Supplementary Figure S1). These frameworks increase the interactivity associated with the App while requiring students to orient to key concepts relevant to each task to problem solve. For more details on the App's design and development, see Wright and Oliver-Hoyo (2021).

3 The learning environments

Sharples and colleagues have a two-part definition of mobile learning. Mobile learning can be education (a) supported by mobile devices or (b) that occurs in unique spaces outside the formal classroom (Sharples et al., 2007). A holistic characterization of mobile learning must consider the design of technological tools and the context in which the learning occurs (Sharples et al., 2007; Imtinan et al., 2013). Therefore, we adopt the Task Model of Mobile Learning to characterize the remote and face-to-face learning environments in which students engaged with the activity (Taylor et al., 2006; Sharples et al., 2007).

The Task Model of Mobile Learning is grounded in activity theory and comprises six factors, each with a technological and semiotic layer. The factors are the learning goals or outcomes (*Objective*), the learner and their prior knowledge (*Subject*), the function of the medium or artifact used to facilitate learning processes (*Tool*), the social and pedagogical parameters that moderate learning (*Control*), the device's portability and the environment's relevance to the learning goals (*Context*), and the potential for interactions (*Communication*). Each factor coalesces with the others so that the individual completes the activity as a *Changed Object* with revised knowledge and skills (Taylor et al., 2006; Frohberg et al., 2009). Frohberg and colleagues used the semiotic layer of the Task Model of Mobile Learning, with technology as the enabler, to create a five-point rating scale for each factor (Table 1) and describe mobile learning environments (Frohberg et al., 2009). By comparing multiple instantiations using this framework, one can characterize activities and explore how differences influence the Subject's revised knowledge or skills (Sharples et al., 2007). We use these factors in Table 1 to characterize how students completed the same activity in different learning environments.

In both environments, students were expected to analyze and interpret spectra and molecular structures (*Objective* – analyze). On average, the students reported good prior knowledge (*Subject* – good previous knowledge). The App was designed for interactivity using the CPOE framework in that it is guided with a playful approach to increase motivation (*Tool* – interaction for motivation and control) through engagement with structured content that has space for students to make decisions during the learning process (*Control* - mainly teacher control). However, there were a few differences between the learning environments.

3.1 Face-to-face environment

During the Spring 2020 semester, the activity was administered on instructor–provided tablets within a single three–hour laboratory (*Context* – formalized) and with a teaching assistant in person. The students completed the activity with a partner (*Communication* – tightly coupled pairs). The partners shared an iPad and had to advance through the AR experience together; however, they were each responsible for submitting individual work. During the spring semester, students would document their predictions and provide explanations on the worksheet while comparing and observing the AR components in the App.

3.2 Remote environment

The Fall 2020 semester was the first full semester of remote learning due to the COVID-19 pandemic. The students completed the activity in their own space (*Context* – independent), using personal devices. Students had an entire week to complete the activity. While students could reach out to their teaching assistant via Zoom throughout the week, students did not work with partners (*Communication* - isolated learners). During the fall semester, students only wrote their explanations on the worksheet. Each AR experience

TABLE 1 An overview of the scales for each factor [adapted from Frohberg et al., 2009].

Factors	Scales 1	2	3	4	5
Objective	Know	Comprehend	Apply	R F Analyze	Synthesize and evaluate
Subject	Novice	Little previous knowledge	R F Good previous knowledge	Much previous knowledge	Expert
Tool	Content delivery	R F Interaction for motivation & control	Reflective interaction	Reflective data collection	Content construction
Control	Full teacher control	R F Mainly teacher control	Scaffold	Mainly learner control	Full learner control
Context	R Independent context	F Formalized context	–	Physical context	Socializing context
Communication	R Isolated learners	Loose couples	F Tight couples	Communication within group	Cooperation

The green R describes the remote learning environment in the fall, and the purple F describes the face-to-face learning environment in the spring.

was modified to require students to compare, make a prediction, and make an observation within the App.

4 Participants and data collection and analysis measures

4.1 Participants

The face-to-face data were collected from 114 students across six laboratory sections in the spring. The remote data were collected from 154 students across eight laboratory sections in the fall. All participants were concurrently self-enrolled in the Organic Chemistry II laboratory course lecture. Students participated in the study by taking a pre-and post-survey to gather information about their perception and user experience. Participants were not given incentives to complete the activity or survey as it was part of their coursework; however, they had to consent for their feedback to be used for research per our Institutional Review Board guidelines. In the face-to-face and remote environments, most students were sophomores or juniors (80.7, 67.5%) and self-identified as female (57.9, 54.6%). They reported learning about ^{1}H NMR in their lecture course before the laboratory (64.9, 61.7%). Less than one-third of students reported previously using AR (32.9, 29.2%).

4.2 Data collection and analysis

Using Qualtrics software, participants completed a pre-survey before the activity and a post-survey immediately after. We analyzed the data related to each research question within each environment and compared environments.

To investigate student user experience (research question one, RQ1), we used the user experience questionnaire (UEQ) in the post-survey (Laugwitz et al., 2008; Schrepp et al., 2014). The questionnaire contains 26 items across six scales:

1 Attractiveness - individual's overall impression of the product.
2 Perspicuity - how easy the app is to understand/learn.
3 Efficiency - how easy the app is to use to solve tasks.
4 Dependability - how dependable the app is to use.
5 Stimulation - how exciting the app is to use.
6 Novelty - how creative or innovative the app is perceived.

The UEQ is a semantic differential survey in which students are given two opposing words and asked to indicate their preference using a seven-point scale (−3 to +3). Responses can be positive (greater than +0.8), neutral (between +0.8 and-0.8), or negative (less than −0.8) (Santoso et al., 2016). All six scales showed relatively acceptable levels of reliability across both semesters (Supplementary Table S1), indicating that every item in the scale was measuring something similar to other items within the scale (Taber, 2018). A MANCOVA was utilized to analyze data from the UEQ across semesters. The outcome variables for the analysis were the six UEQ scales, with the independent variable being the learning environment. Research has shown that prior knowledge influences experience and performance (Kohl and Finkelstein, 2006; Rittle-Johnson et al., 2009; Braithwaite and Goldstone, 2015). The analysis also controlled for students' prior experience with AR technology and NMR content. All assumptions of the statistical test were met (Field, 2009).

The post-survey contained more items to understand students' overall perception of the activity (RQ2). Students rated each item on a scale from 1 (Strongly disagree) to 5 (Strongly agree). The prompts were:

1 Given the chance, I would use the AR app again.
2 I found the lab structure more engaging than that of other labs I have taken in the past.
3 The App was effective in increasing my engagement with the content.
4 The worksheet was effective in increasing my engagement with the content.

Students were asked to indicate if they experienced challenges with the worksheet and the App via a yes/no item. In both cases, students could provide an open-ended response to describe their experiences with the activity:

1 Please describe those challenges or difficulties with using the worksheet (App).
2 Please provide 1–2 things you found most helpful about the worksheet (App).
3 Please provide 1–2 things you would improve about the worksheet (App).

Data from items relating to student perceptions of the App and worksheet violated the assumption of normality, and therefore, non-parametric tests were utilized. Mann–Whitney U tests were run for each item, with perception as the outcome variable and learning environment as the independent variable. A separate chi-square analysis was used to explore the frequency of students who stated they experienced challenges using the worksheet and application in each environment. The open-ended responses underwent a content analysis to identify common patterns across the participants within each environment (Patton, 2002). One coder analyzed 20 % of the data and developed a codebook. The codebook was shared with a second coder who coded the same 20 % of the data. The coders met and refined the codebook using constant comparative analysis (Glaser, 1965). Both coders separately analyzed the remainder of the data. After independent coding, the two researchers discussed their codes until a 100% negotiated agreement was reached (Campbell et al., 2013; Saldaña, 2013). Peer debriefing and negative case analysis were used to ensure the credibility of the findings (Lincoln and Guba, 1985).

Lastly, we used pre-post survey questions to investigate if the activity adjusted students' perception of the importance of chemistry visualizations or student knowledge of ^{1}H NMR (RQ3). The first four items were adapted from a survey to evaluate BiochemAR, another augmented reality educational tool (Sung et al., 2020). The last two items were author-generated and specific to ^{1}H NMR. Students rated the following items on a 1 (Strongly disagree) to 5 (Strongly agree) scale:

1 Seeing a visual helps me connect my knowledge and new information about molecules.
2 Manipulating something physically helps me connect what I know and new information about molecules.
3 Analyzing 2D images or molecules is helpful for learning organic chemistry.
4 Analyzing 3D images of molecules is helpful for learning organic chemistry.
5 I understand the concepts related to ^{1}H NMR.
6 I know how to solve ^{1}H NMR problems.

Data from these items violated the assumption of normality, and reverse score and log transformations were used to correct the data. A mixed (2×2) MANCOVA was run to understand differences between environments over time. The outcome variables for the analysis were item responses at the pre-and post-survey, with the independent variable being the learning environment. Interaction effects (differences between environments over time) and main effects (differences between environments or differences over time) were investigated.

5 Results

5.1 RQ1 – How do students experience the activity in each environment?

5.1.1 Face-to-face

Overall, a majority (> 50%) of students in face-to-face settings indicated that they had positive overall impressions of the App (attractiveness = 56.68% positive). When using the App, a majority indicated it was easy to learn (perspicuity = 61.70%), use (efficiency = 59.57%), and dependable (dependability = 54.26%). A slight majority also found the App exciting (stimulation = 52.13%) and perceived it as creative/innovative (novelty = 55.32%).

5.1.2 Remote learning

Student responses during remote learning were less positive, with less than half indicating they had positive overall impressions of the App (attractiveness = 43.51% positive). Less than half reported finding the App easy to use (efficiency = 46.10%) and viewing it as exciting (stimulation = 37.66%). However, most students did indicate the App was easy to learn (perspicuity = 57.14%), dependable (dependability = 51.30%), and perceived it as creative/innovative (novelty = 59.09%).

5.1.3 Environment comparison

Students in face-to-face had significantly higher ratings than students in remote learning for the efficiency (F (1,211) = 6.50, p = 0.012, eta = 0.03), dependability (F (1,211) = 4.40, p = 0.037, eta = 0.02), stimulation (F (1,211) = 12.32, p < 0.001, eta = 0.06), and attractiveness (F (1,211) = 10.38, p = 0.001, eta = 0.05) factors. There was no difference between the perspicuity (F (1,211) = 1.79, p = 0.18) and novelty (F (1,211) = 0.91, p = 0.34) factors (see Table 2 for mean and standard deviations).

5.2 RQ2 – How do students perceive the activity in each setting?

5.2.1 Face-to-face

Student responses were largely positive (Table 3), with ~70% of students agreeing that they found the lab structure more engaging than previous labs (69.15%) or would use the App again (73.40%). Students also agreed that the App increased their engagement (80.85%) and helped them understand the content (78.72%). Further, students said the worksheet increased their engagement (76.60%) and helped them understand the content (73.40%). Students reported encountering challenges or difficulties with the worksheet (54%) and App (44%).

TABLE 2 Distributions of student responses for UEQ scales.

	Negative (%)	Neutral (%)	Positive (%)	M (SD)
Efficiency				
Face-to-face*	7.45	32.98	59.57	1.05 (1.17)
Remote learning	12.99	40.91	46.10	0.66 (1.27)
Perspicuity				
Face-to-face	5.32	32.98	61.70	1.10 (1.20)
Remote learning	9.74	33.12	57.14	0.90 (1.34)
Dependability				
Face-to-face*	2.13	43.62	54.26	1.04 (0.90)
Remote learning	5.84	42.86	51.30	0.80 (0.99)
Stimulation				
Face-to-face*	4.26	43.62	52.13	1.01 (1.02)
Remote learning	13.64	48.70	37.66	0.50 (1.20)
Novelty				
Face-to-face	5.32	39.36	55.32	1.19 (1.16)
Remote learning	3.90	37.01	59.09	1.08 (1.06)
Attractiveness				
Face-to-face*	6.38	37.23	56.38	0.99 (1.23)
Remote learning	18.18	38.31	43.51	0.45 (1.33)

*Indicates significantly higher rating on that factor within that environment. Mean scores can be positive (> + 0.8), neutral (between + 0.8 and-0.8), or negative (< −0.8).

TABLE 3 Distributions of student responses for the UEQ scales.

	Disagree (%)	Neutral (%)	Agree (%)	Mean rank	Mann–Whitney U
I found the lab structure to be more engaging than other labs i have taken in the past					
Face-to-face*	22.34	8.51	69.15	141.10	$Z = -2.95$, $p = 0.003$, $\eta^2 = 0.04$
Remote learning	39.61	9.74	50.65	114.37	
Given the chance, I would use the AR app again:					
Face-to-face*	14.89	11.70	73.40	139.57	$Z = -2.68$, $p = 0.007$, $\eta^2 = 0.03$
Remote learning	29.87	16.23	53.90	115.30	
The app was effective in helping me understand course content:					
Face-to-face*	9.57	11.70	78.72	138.80	$Z = -2.56$, $p = 0.01$, $\eta^2 = 0.03$
Remote Learning	16.88	24.03	59.09	115.77	
The app was effective in engaging me in course content:					
Face-to-face*	6.38	12.77	80.85	143.72	$Z = -3.45$, $p < 0.001$, $\eta^2 = 0.05$
Remote Learning	18.83	22.73	58.44	112.77	
The worksheet was effective in engaging me in the course content:					
Face-to-face*	10.64	11.70	76.60	150.33	$Z = -4.67$, $p < 0.001$, $\eta^2 = 0.09$
Remote Learning	34.42	12.99	52.60	108.10	
The worksheet was effective in helping me understand course content:					
Face-to-face*	18.09	8.51	73.40	141.29	$Z = -2.99$, $p = 0.003$, $\eta^2 = 0.04$
Remote Learning	24.68	18.18	57.14	114.25	

*Indicates significantly higher rating within that environment.

5.2.2 Remote learning

Student responses during remote learning were less positive (Table 3), with a slight majority agreeing that they found the lab structure more engaging than previous labs (50.65%) or would use the App again (53.90%). Students also agreed that the App increased their engagement (58.44%) and helped them understand the content (59.09%). A slight majority said the worksheet increased their engagement (52.60%) and helped them understand the content (57.14%). Students reported encountering challenges or difficulties with the worksheet (51%) and App (36%).

TABLE 4 Activity challenges supported by exemplary responses from students.

Primarily Application		Primarily Worksheet	
Challenges encountered within both environments			
Loading the AR image: *"Nothing too crazy, it was just frustrating sometimes when the images wouldn't scan or pick up."* – F042	Manipulating molecules: *"Rotating the molecule was difficult… It was also too small of a screen for me to be able to see some of the words and images."* – R031	Length: *"…learning a new topic such as this with just a worksheet is very hard to do. Also, it was very long…"* –R111	Unclear explanations: *"Some of the directions were unclear and the explanations about the topic were not as 'easy' as I would have liked if that was the first time, I looked at the topic."* – F091
Improvements suggested across both semesters			
Loading the AR image: *"The only problem I had was having the 3D molecules actually pop up so if there is a way to fix that, then I would be able to learn better."* – F064	Manipulating molecules: *"Provide a lock picture option for some examples to be able to move the molecule easier but not have to hold the phone up the whole time."* – R103	Shorten the worksheet: *"I really liked the structure of this worksheet, and I don't think many changes need to be made. I do think that there may be a few too many examples (it took a very long time to complete the lab), but I think everything was well written and helpful for learning."* – R047	Clearer explanations: *"Explaining what each peak is in the NMR earlier would make more sense for those who have never seen NMR before."* – F109
Helpful components across both semesters			
3D visualizations: *"The visuals were useful in explaining the electronegativity and H shielding. I did not fully remember how the shifting worked and feel much more confident after this worksheet."* – R005	Manipulating molecules: *"It provides easy-to-understand visuals that were also capable of being manipulated to provide different perspectives."* – F029	Detailed explanations: *"The worksheet was very informative and explained the difficult topics well. It was also clear and concise on what it was asking or going over."* – R153	Content progression: *"The problems built on one another and got more in-depth as we progressed."* – F023
Challenges unique to the remote environment			
Screen size: *"This is a bit difficult on a small screen - especially when trying to drag elements around."* – R135	Battery drained: *"The App drained my phone battery, and it was difficult having to hold the phone up while moving the molecules around at the same time"* – R103	Documenting responses: *"Drawing is hard to do with the computer"* – R130	Multiple screens: *"It was challenging constantly switching between my phone and the worksheet on my laptop. It was straining on the eyes and tiring."* – R127

The F and R student pseudonyms correspond to the face-to-face and remote environments.

5.2.3 Environment comparison

Students in the face-to-face were significantly more positive than those in remote learning about using the App and worksheet, as well as their impact on their engagement/learning (Table 3). There were neither significant differences in students who reported challenges or difficulties with the worksheet (χ^2 (1) =3.72, $p = 0.054$) nor the App within either environment (test was significant, χ^2 (1) =8.20, $p = 0.005$, however, standardized residuals did not meet significance). The open-ended responses about the worksheet were split among students as what some considered the most challenging others considered the most helpful (see example quotes in Table 4). The main worksheet challenges and desires for improvement concerned the length and the lack of clear content explanations. Coincidentally, students also reported that the detailed content explanations and step-by-step content progression within the worksheet were most useful for learning. The main App challenges were scanning the target for the AR image to appear and moving/rotating the augmented molecules. Those challenges were described as the suggested improvements for the App. Students stated the most helpful things about the App were the 3D visualizations that explained the content and the ability to manipulate the molecules.

A few responses were unique to the remote environment (Table 4). Remote students had the option to print their worksheet or complete it online. Many students chose to complete it online; however, students said it was difficult to draw on the document via Word or Google Docs and that switching between multiple screens (i.e., their mobile device and the computer) was overwhelming. Students in remote learning were also challenged by the size and power limitations of their mobile devices. None of these challenges were evidenced in the Spring.

5.3 RQ3 – to what extent does the activity adjust students' perceptions of the importance of visualizations for learning chemistry and their knowledge of ^{1}H NMR?

5.3.1 Face-to-face

Students overwhelmingly agreed at pre-and post-survey with the perceptions about the importance of seeing a visual (97.3, 93.6%), physically manipulating objects (92.9, 92.5%), and analyzing 2D (77.5, 84%) and 3D (93.7 and 93.6%) images. A slight majority of students felt they understood the underlying concepts (59.5%) and were able

TABLE 5 Distributions of student rankings for each perception.

	Frequencies at pre-survey (%)			Frequencies at post-survey (%)		
	Disagree	Neutral	Agree	Disagree	Neutral	Agree
Seeing a visual helps me make connections between what I know and new information about molecules.						
Face-to-face	1.80	0.90	97.30	0.00	6.40	93.60
Remote	4.80	4.00	91.30	1.90	7.10	91.00
Manipulating something physically helps me make connections between what I know and new information about molecules.						
Face-to-face	2.7	5.4	92.9	1.10	6.40	92.50
Remote	4.80	8.70	86.50	2.60	16.90	81.50
Analyzing 2D images or molecules is helpful for learning organic chemistry.						
Face-to-face	6.30	16.20	77.50	3.20	12.80	84.00
Remote	7.10	14.30	78.60	3.90	10.40	85.70
Analyzing 3D images or molecules is helpful for learning organic chemistry.						
Face-to-face	0.90	5.40	93.70	0.00	6.40	93.60
Remote	7.90	4.80	87.30	1.90	5.80	92.20
I understand concepts related to ^{1}H NMR						
Face-to-face	26.10	14.40	59.50	6.40	5.30	88.30
Remote	23.80	18.30	57.90	5.80	14.90	79.30
I know how to solve ^{1}H NMR problems.						
Face-to-face	32.40	14.40	53.20	9.60	3.20	87.20
Remote	31.70	15.10	53.20	7.80	14.90	77.30

Student disagreement was classified as a negative perception, and student agreement as a positive perception.

to solve ^{1}H NMR problems (53.2%) during the pre-survey, but this increased during the post-survey (88.3 and 87.2%, respectively).

5.3.2 Remote learning

Similarly, students overwhelmingly agreed at pre-and post-survey with the perceptions about the importance of seeing a visual (91.3, 91%), physically manipulating objects (86.5, 81.5%), and analyzing 2D (78.6, 85.7%) and 3D (87.3 and 92.2%) images. A slight majority of students felt they understood the underlying concepts (57.9%) and were able to solve ^{1}H NMR problems (53.2%) during the pre-survey, but this increased during the post-survey (79.3 and 77.3%, respectively).

5.3.3 Environment comparison

All significant effects are described below; for an overview of non-significant effects, see the Supplementary Table S2. There were significant interaction effects for items about student ability to understand (F (1, 215) = 8.14, p = 0.005, eta = 0.04) and solve ^{1}H NMR problems (F (1, 215) = 6.57, p = 0.011, eta = 0.03). Students in face-to-face and remote environments showed significant increases in items from pre-to post-survey (p < 0.001). No differences were found between environments at the pre-survey; however, at post-survey, those in the remote environment showed significantly lower ratings (p < 0.05) (Table 5).

A main effect of environment was found for the importance of manipulating physical objects (F(1, 215) = 4.24, p = 0.04, eta = 0.02). Collapsed across time points, students in the remote environment indicated lower perceptions compared to those in face-to-face (p = 0.041). A main effect of time was also found for the importance of analyzing 2D images (F(1, 215) = 10.08, p = 0.002, eta = 0.05), with students in both environments showing a significant increase over time.

6 Discussion and implications

This paper describes using an AR application (H NMR MoleculAR) and accompanying worksheet to support undergraduate organic chemistry students in understanding ^{1}H NMR spectroscopy in two different environments. Students in the spring semester completed the activity in a face-to-face laboratory (*Context* - formalized), working in pairs (*Communication* - tightly coupled). Students in the fall semester completed the activity independently (*Communication* - isolated learners) and remotely (*Context* - independent). Our findings indicated that students had a neutral to positive user experience and relatively positive perceptions of the chemistry activity. However, students who completed the activity remotely had significantly lower perceptions and a less positive user experience than students who completed the activity face-to-face. Delivering this activity in two different settings led to multiple lessons learned.

6.1 Lesson 1: augmented reality experiences support students in connecting content to visualizations

Across both settings, students overwhelmingly valued the 3D visualizations and how they helped them to understand the content (see R005, Table 4) and "make connections between the theory and the practice" – R084. Students discussed how the color coding helped them link the hydrogen atoms to specific chemical shifts, and the animations helped them visualize how electrons interact with magnetic fields. However, some students struggled with the fact it was augmented. Like R103 in Table 4, participant R078 stated, "I did not like how I could not "freeze" the screen and look at it while I was looking at the questions. It made it hard to hold my phone up and also read on my laptop." Freezing the screen would make the image appear more like a 3D object on a screen than an augmented reality image. These sentiments lead to questions about the content and contexts in which AR affords more than non-augmented virtual representations on a desktop or website application when learning chemistry.

6.2 Lesson 2: augmented reality experiences may be more useful when reinforcing material than introducing new material

While the statistical analyses controlled for student exposure to AR and NMR, the open-ended survey responses revealed that students who had not covered ^{1}H NMR in their lecture struggled with the activity. This sentiment was present regardless of environment and can be evidenced by participants R111, F091, and F109 in Table 4. As participant R082 stated, completing this activity without prior content exposure requires learning "new material and devices I was not familiar with." This could potentially increase student cognitive load, a significant challenge in using AR in teaching chemistry (Cheng and Tsai, 2013). Students may perceive AR activities more positively if they have some experience with the content. Other studies have been designed in which the AR component was incorporated towards the end of the instructional unit (Behmke et al., 2018) or after foundational knowledge of the topic had been covered in the course (Cai et al., 2014; Sung et al., 2020). Educators could benefit from research that provides insight into how user prior knowledge, the amount of material, and task complexity impact chemistry learning with AR tools.

6.3 Lesson 3: collaborative environments may positively impact student perceptions of using AR for learning

Students' perceptions scored higher in the face-to-face environment where students could communicate with a partner. However, most perceptions were still rated positively in the remote and independent environment. Students in the face-to-face environment gave significantly higher ratings for their willingness to reuse the App, the effectiveness of the App in helping them study the content, the worksheet's effectiveness in assisting students with course content, and overall lab engagement. In both environments, students reported an increased understanding of ^{1}H NMR concepts and problem-solving ability, but this increase was significantly lower in the remote learning environment. While we cannot disentangle the remote from the independent work or the in-person from the collaborative, studies show that interpersonal interactions influence students' attitudes toward science (Wei et al., 2019), and when students collaborate with mobile learning activities, they perceive greater improvements in learning than without collaboration (Burke et al., 2021). Students who work collaboratively are more likely to discuss their reflections and ask one another questions than students who work independently (Chi and Wylie, 2014). Students in the remote environment echoed this point, as one stated they were challenged "without the feedback and engagement of a lab partner/table to do it with" (R060). Although mobile applications have the affordance of portability, there are other nuances to consider that can impact student learning and experience (Frohberg et al., 2009; Sung et al., 2020). Research around AR use in collaborative vs. independent chemistry learning environments, especially if some students are inexperienced with AR, would be beneficial to support educators in adapting these technologies in the classroom.

6.4 Lesson 4: overusing technology can reveal fewer positive perceptions and a lower user experience

Although the results show significantly lower ratings in user experience in the remote environment, student ratings still fell within the neutral zone of the scale. When face-to-face, students had paper copies of the worksheet and used the tablet and App to examine the worksheet. When remote, most students did not choose to print the worksheet. They reported using a mobile device to examine the worksheet projected on a laptop or desktop. This overuse of screens is not an efficient way to use AR and likely decreased student user experience and perceptions of the activity (R127, Table 4). Furthermore, though mobile devices are widely adopted, providing students with a device may be most beneficial, especially if using an app for the first time. Our data demonstrate that using personal devices made some students frustrated by the screen size or the battery's strength (see R135 and R103 Table 4).

7 Limitations

Several methodological constraints in evaluating student perceptions and experiences must be acknowledged. First, all data collected were self-reported. Thus, we are not discussing whether the activity helped students learn ^{1}H NMR but the extent to which they perceive the App impacted their learning. This distinction must be considered when interpreting the results and the lessons learned. Additionally, while we have qualitative evidence that the collaborative environment may positively impact students perceptions of AR, we cannot fully separate the collaborative from the in-person or the independent from the remote to quantify which had a more meaningful impact on students perceptions. Lastly, since the Fall semester occurred during the COVID-19 global pandemic, the context was not the standard online learning scenario where students opt into a remote learning environment. Students' overall dissatisfaction with remote learning may have impacted their perceptions. Even within these confines, this manuscript provides insight into lessons learned and suggestions for research that will help educators implement novel technologies into the chemistry classroom in face-to-face and online settings.

Data availability statement

The datasets presented in this article are not readily available per our current Institutional Review Board guidelines. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by North Carolina State University Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LW: Conceptualization, Data curation, Formal analysis, Writing – original draft. DS: Data curation, Formal analysis, Writing – review & editing. DC: Formal analysis, Writing – review & editing. MO-H: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was made possible with support from North Carolina State University and Indiana University Indianapolis.

Acknowledgments

We'd like to acknowledge all participants for providing their honest feedback, the teaching assistants who helped facilitate the laboratory, and Maria Gallardo-Williams, the laboratory coordinator.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1384129/full#supplementary-material

References

Abdinejad, M., Ferrag, C., Qorbani, H. S., and Dalili, S. (2021). Developing a simple and cost-effective Markerless augmented reality tool for chemistry education. *J. Chem. Educ.* 98, 1783–1788. doi: 10.1021/acs.jchemed.1c00173

Alfieri, L., Nokes-Malach, T. J., and Schunn, C. D. (2013). Learning through case comparisons: a meta-analytic review. *Educ. Psychol.* 48, 87–113. doi: 10.1080/00461520.2013.775712

Anderson, S. Y., Ong, W. S. Y., and Momsen, J. L. (2020). Support for instructional scaffolding with 1 H NMR spectroscopy spectral features in organic chemistry textbook problems. *Chem. Educ. Res. Prac.* 21, 749–764. doi: 10.1039/c9rp00252a

Behmke, D., Kerven, D., Lutz, R., Paredes, J., Pennington, R., Brannock, E., et al. (2018) 'Augmented reality chemistry: transforming 2-D molecular representations into interactive 3-D structures', *Proceedings of the interdisciplinary STEM teaching and learning conference*, 2, pp. 5–11.

Braithwaite, D. W., and Goldstone, R. L. (2015). Effects of variation and prior knowledge on abstract concept learning. *Cogn. Instr.* 33, 226–256. doi: 10.1080/07370008.2015.1067215

Burke, P. F., Kearney, M., Schuck, S., and Aubusson, P. (2021). Improving mobile learning in secondary mathematics and science: listening to students. *J. Comput. Assist. Learn.* 38, 1–15. doi: 10.1111/jcal.12596

Cai, S., Wang, X., and Chiang, F. K. (2014). A case study of augmented reality simulation system application in a chemistry course. *Comput. Hum. Behav.* 37, 31–40. doi: 10.1016/j.chb.2014.04.018

Campbell, J. L., Quincy, C., Osserman, J., and Pedersen, O. K. (2013). Coding in-depth Semistructured interviews: problems of unitization and Intercoder reliability and agreement. *Sociol. Methods Res.* 42, 294–320. doi: 10.1177/0049124113500475

Cheng, K. H., and Tsai, C. C. (2013). Affordances of augmented reality in science learning: suggestions for future research. *J. Sci. Educ. Technol.* 22, 449–462. doi: 10.1007/s10956-012-9405-9

Chi, M. T. H., and Wylie, R. (2014). The ICAP framework: linking cognitive engagement to active learning outcomes. *Educ. Psychol.* 49, 219–243. doi: 10.1080/00461520.2014.965823

Connor, M. C., and Shultz, G. V. (2018). Teaching assistants' topic-specific pedagogical content knowledge in 1H NMR spectroscopy. *Chem. Educ. Res. Pract.* 19, 653–669. doi: 10.1039/C7RP00204A

De Lima, C. B., Walton, S., and Owen, T. (2022). A critical outlook at augmented reality and its adoption in education. *Comput. Educ. Open* 3:100103. doi: 10.1016/j.caeo.2022.100103

Field, A. (2009) *Discovering statistics*. *3rd*. London: Sage Publications.

Frohberg, D., Göth, C., and Schwabe, G. (2009). Mobile learning projects - a critical analysis of the state of the art. *J. Comput. Assist. Learn.* 25, 307–331. doi: 10.1111/j.1365-2729.2009.00315.x

Glaser, B. G. (1965). *The Constant Comparative Method of Qualitative Analysis*. 12, 436–445. doi: 10.2307/798843

Goddard, T. D., Brilliant, A. A., Skillman, T. L., Vergenz, S., Tyrwhitt-Drake, J., Meng, E. C., et al. (2018). Molecular visualization on the holodeck. *J. Mol. Biol.* 430, 3982–3996. doi: 10.1016/j.jmb.2018.06.040

Graulich, N., and Schween, M. (2018). Concept-oriented task design: making purposeful case comparisons in organic chemistry. *J. Chem. Educ.* 95, 376–383. doi: 10.1021/acs.jchemed.7b00672

Huwer, J., Lauer, L., Seibert, J., Thyssen, C., Dörrenbächer-Ulrich, L., and Perels, F. (2018). Re-experiencing chemistry with augmented reality: new possibilities for individual support. *World J. Chem. Educ.* 6, 212–217. doi: 10.12691/wjce-6-5-2

Imtinan, U., Chang, V., and Issa, T. (2013). "Common Mobile learning characteristics - an analysis of three Mobile learning models and frameworks" in *Proceedings of the international conference Mobile learning* (Lisbon), 3–11.

Kohl, P. B., and Finkelstein, N. D. (2006). Effects of representation on students solving physics problems: a fine-grained characterization. *Phys. Educ. Res.* 2, 1–12. doi: 10.1103/PhysRevSTPER.2.010106

Laugwitz, B., Held, T., and Schrepp, M. (2008). "Construction and evaluation of a user experience questionnaire" in *HCI and usability for education and work. USAB 2008 lecture notes in computer science*. ed. A. Holzinger (Berlin, Heidelberg: Springer), 63–76.

Lincoln, Y. S., and Guba, E. G., (1985) *Naturalistic inquiry*. Thousand Oaks, CA: Sage Publications, *9*, 438–439.

Mazzuco, A., Krassmann, A. L., Reategui, E., and Gomes, R. S. (2022). A systematic review of augmented reality in chemistry education. *Review of Education* 10:e3325. doi: 10.1002/rev3.3325

Patton, M. Q. (2002) *Qualitative research and evaluation methods*. *3rd*. Thousand Oaks, CA: Sage Publications.

Rittle-Johnson, B., Star, J. R., and Durkin, K. (2009). The importance of prior knowledge when comparing examples: influences on conceptual and procedural knowledge of equation solving. *J. Educ. Psychol.* 101, 836–852. doi: 10.1037/a0016026

Saldaña, J. (2013). "The coding manual for qualitative researchers" in *Qualitative research in organizations and management: an international journal*, vol. *12*. *3rd* ed (Thousand Oaks, CA: Sage Publications), 169–170.

Santoso, H. B., Schrepp, M., Kartono Isal, R. Y., Yudha Utom, A., and Priyogi, B. (2016). Measuring the user experience. *J. Educ. Online* 13, 1–79. doi: 10.9743/JEO.2016.1.5

Schrepp, M., Hinderks, A., and Thomaschewski, J. (2014) 'Applying the user experience questionnaire (UEQ) in different evaluation scenarios', in A. Marcus (ed.) Design, user experience, and usability. Theories, methods, and tools for designing the user experience: Third international conference, DUXU 2014, Held as part of HCI international 2014. Heraklion, Crete, Greece. Proceedings, part I. Series: Lecture notes. Springer, pp. 383–392.

Sharples, M., Taylor, J., and Vavoula, G. (2007). "A theory of learning for the Mobile age" in *The sage handbook of Elearning research*. eds. R. Andrews and C. Haythornthwaite (London: Sage), 221–247.

Sung, R.-J., Wilson, A. T., Lo, S. M., Crowl, L. M., Nardi, J., St. Clair, K., et al. (2020). BiochemAR: an augmented reality educational tool for teaching macromolecular structure and function. *J. Chem. Educ.* 97, 147–153. doi: 10.1021/acs.jchemed.8b00691

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.* 48, 1273–1296. doi: 10.1007/s11165-016-9602-2

Taylor, J., Sharples, M., O'Malley, C., Vavoula, G., and Waycott, J. (2006). Towards a task model for mobile learning: a dialectical approach. *Int. J. Learn. Technol.* 2, 138–158. doi: 10.1504/ijlt.2006.010616

Tee, N. Y. K., Gan, H. S., Li, J., Cheong, B. H. P., Tan, H. Y., Liew, O. W., et al. (2018). Developing and demonstrating an augmented reality colorimetric titration tool. *J. Chem. Educ.* 95, 393–399. doi: 10.1021/acs.jchemed.7b00618

Wei, J., Treagust, D. F., Mocerino, M., Lucey, A. D., Zadnik, M. G., and Lindsay, E. D. (2019). Understanding interactions in face-to-face and remote undergraduate science laboratories: a literature review. *Discip. Interdiscip. Sci. Educ. Res.* 1, 1–16. doi: 10.1186/s43031-019-0015-8

Wright, L., and Oliver-Hoyo, M. (2021). Development and evaluation of the H NMR MoleculAR application. *J. Chem. Educ.* 98, 478–488. doi: 10.1021/acs.jchemed.0c01068

frontiers | Frontiers in Education

TYPE Review
PUBLISHED 26 April 2024
DOI 10.3389/feduc.2024.1374793

Check for updates

OPEN ACCESS

EDITED BY
Michael Wentzel,
Augsburg University, United States

REVIEWED BY
Dušica D. Rodic,
University of Novi Sad, Serbia
Patrick Willoughby,
Ripon College, United States

*CORRESPONDENCE
Nicole Graulich
Nicole.Graulich@dc.jlug.de

RECEIVED 22 January 2024
ACCEPTED 11 April 2024
PUBLISHED 26 April 2024

CITATION
Graulich N and Lieber L (2024) Why comparing matters – on case comparisons in organic chemistry.
Front. Educ. 9:1374793.
doi: 10.3389/feduc.2024.1374793

Why comparing matters – on case comparisons in organic chemistry

Nicole Graulich* and Leonie Lieber

Institute of Chemistry Education, Justus-Liebig-University Giessen, Giessen, Germany

When working with domain-specific representations such as structural molecular representations and reaction mechanisms, learners need to be engaged in multiple cognitive operations, from attending to relevant areas of representations, linking implicit information to structural features, and making meaningful connections between information and reaction processes. For these processes, appropriate instruction, such as a clever task design, becomes a crucial factor for successful learning. Chemistry learning, and especially organic chemistry, merely addressed meaningful task design in classes, often using more reproduction-oriented predict-the-product tasks. In recent years, rethinking task design has become a major focus for instructional design in chemistry education research. Thus, this perspective aims to illustrate the theoretical underpinning of comparing cases from different perspectives, such as the structure-mapping theory, the cognitive load theory, and the variation theory, and outlines, based on the cognitive theory of multimedia learning, how instructors can support their students. Variations of this task design in the chemistry classroom and recommendations for teaching with case comparisons based on current state-of-the-art evidence from research studies in chemistry education research are provided.

KEYWORDS

case comparisons, chemistry education, support, guidance, instruction

1 Introduction

As educators in chemistry, we would unanimously agree that understanding the relationship between the Lewis structure representations of organic molecules and their chemical properties, the molecular architecture, as named by Laszlo (2002), is essential for explaining or predicting chemical behavior. When learning chemistry, students, thus, encounter various ways of representing structures and processes (i.e., electron-pushing formalism) and must connect this to chemical and physical characteristics and energetic considerations (Goodwin, 2010). As a chemical entity has both a visible structural representation and an underlying conceptual aspect, difficulties in linking these two aspects can lead to a superficial understanding. Studies consistently show that students often focus on surface features or patterns when estimating the reactivity of molecules, overlooking functional or more abstract relational similarities (*cf.* Cooper et al., 2013; Anzovino and Bretz, 2016; Talanquer, 2017). They tend to equate visual similarity with chemical similarity, potentially missing out on understanding how different structural environments can lead to property changes, i.e., changes in chemical reactivity (Bhattacharyya, 2014; Graulich et al., 2019).

One may now ask, why comparing and contrasting should be an important part of learning in chemistry. The act of comparing is inherent to the discipline because it allows us to understand the properties of substances by comparing their behavior in different conditions (Goodwin, 2008). Chemists often compare different substances to identify similarities and differences of chemical and physical properties. In chemical synthesis, making small changes in functional groups at a target catalyst, for example, allows us to determine which ones are most effective at promoting specific chemical reactions (Afagh and Yudin, 2010). By comparing the behavior of chemical systems, chemists can gain a deeper understanding of the underlying principles of chemical processes to monitor and control chemical reactions or refine computational models. Comparing either experimental, machine learning or computational data allows us to estimate the magnitude of effects (Keith et al., 2021). Comparing, for instance, kinetic data of reactions helps determine the magnitude of reaction speed, for instance, influenced by changes of electronic substituent effects (Trabert and Schween, 2018). In some cases, we have this data at hand in terms of empirical properties, such as electronegativity or pK_a values, but in other cases, in which we do not have access to these data, chemists often express qualitatively the properties of a functional group or molecule, e.g., this leaving group or nucleophile is good, or this structure is stable (Popova and Bretz, 2018). However, to estimate what "good" means requires answering the question "Good, compared to what" and essentially answering the question "why is it better?." This is an inherently comparative process that requires knowledge about implicit properties, electron distribution, strength of effects, and energetic considerations. Purposeful case comparisons may engage learners in meaningful sense-making about organic reactions. This assumption is further supported by studies in psychology that have highlighted the educational value of using case comparisons to assist students in grasping new concepts (Schwartz and Bransford, 1998; Gentner et al., 2003). In particular, Gentner et al. (2003) found that comparing two cases simultaneously was more effective for learning than studying five single cases in sequence. By comparing and contrasting different cases, students learn to discern both common and distinctive characteristics that help differentiate and understand key concepts or phenomena. As the instruction continues, such comparisons offer a chance for learners to develop inferences and justifications for the specific features. A meta-analysis by Alfieri et al. (2013) has shown that this method significantly enhances learning. This perspective outlines the theoretical underpinning of case comparisons and highlights how instruction in chemistry can profit from well-designed and orchestrated cases.

2 Why should we learn with case comparisons? Theoretical underpinning

2.1 What does structure mapping theory tell us about comparing?

Learning by comparing cases can be rationalized from a cognitive psychology perspective because it taps into several important cognitive processes, essential for learning and problem-solving. When comparing cases, a learner is engaged in a process called analogical reasoning, which involves finding similarities and differences between cases and using those similarities and differences to make inferences and draw conclusions. This analogical reasoning is a fundamental cognitive process that allows transfer knowledge and skills from one domain to another, or from one context to another (Gick and Holyoak, 1983). The structure mapping theory by Gentner (1989) and Gentner and Markman (1997) explains how this analogical reasoning works. When we compare two situations, objects, or reactions, we look for shared relationships. These relationships could either be similarities in surface features or relational features, such as causal or functional ones. Surface features are always visible features and details of a situation or object and, thus, are easy to discern. While relational structures refer to the abstract relationships between features and implicit information conveyed, they can, but do not necessarily share surface similarities. Comparing a set of correspondences between the surface or relational features of two cases leads to a structural alignment, i.e., discerning the information that two cases share. According to the structure mapping theory, the more shared relational features there are between two situations, the stronger the analogy, the easier to transfer our knowledge about one situation to reason about the other. For example, knowing that an electronegativity difference is needed to make a carbon-heteroatom bond polar, we can use that knowledge to infer that other carbon-heteroatom bonds might be polar as well, when there is a difference in electronegativity, even if the functional group looks different. However, attending to the relational similarity between cases is modulated by expertise. With increasing expertise, we can make use of abstract schemas and use them to categorize tasks based on implicit, conceptual aspects, whereas novice chemistry learners tend to focus on more explicit concrete features (Graulich et al., 2019; Lapierre and Flynn, 2020).

2.2 Cognitive load – the gatekeeper for accessibility

The Cognitive Load Theory (CLT) (Sweller and Chandler, 1994; Kalyuga et al., 1998) offers substantial insights into the use of case comparisons in learning chemistry, emphasizing how instructional design can manage cognitive resources to enhance learning (Paas et al., 2003). The CLT acknowledges the structure or extraneous load of a task (extraneous cognitive load), as well as the cognitive affordances that come with the content (intrinsic cognitive load) and the cognitive effort that a learner needs to activate for learning (germane cognitive load). When we compare cases, we activate our working memory system. However, the use of working memory and the associated capacity is limited, which is why sufficient available capacity must be accessible for effective learning or application of knowledge (Baddeley, 2010). CLT describes that learning is associated with cognitive load and that learning can be simplified or be more challenging depending on the circumstances. Intrinsic cognitive load is related to the difficulty or complexity of the learning material. Sweller (2003) focuses here on element interactivity. In concrete terms, this means that different elements must be processed simultaneously in the working memory during learning. This can happen sequentially, which causes a lower intrinsic cognitive load, or simultaneously, which results in an increased intrinsic cognitive load. If the elements are processed one after the other, e.g., in learning with single cases, this usually leads to memorization; if they are processed simultaneously, e.g., by comparing cases, links are created, which generates understanding but is also more demanding for the working memory (Sweller, 2010). The more prior knowledge learners have, the more links already exist and the lower the intrinsic cognitive load, even when processing elements simultaneously (Paas and Sweller, 2014). Two assumptions support the use of case comparison in light of the intrinsic cognitive load. On the one hand, as our working memory is limited in capacity, comparing cases instead of single cases helps us to be able to attend easily to differences and similarities and neglect other possibly irrelevant features of a situation or object (Schwartz and Bransford, 1998). Simultaneous processing of multiple and maybe irrelevant aspects can be challenging for learners; thus, the extraneous and intrinsic load can be reduced if cases help learners to focus on a reduced number of relevant aspects, as the one variable that needs to be compared can be focused on. This allows us to save capacity in our working memory. Furthermore, studying multiple cases allows learners to see how the same underlying principles apply to different contexts. This can help learners develop a deeper understanding of those principles and how they relate, which makes it easier to build conceptual chunks instead of memorizing single features (Schwartz and Bransford, 1998; Alfieri et al., 2013; Roelle and Berthold, 2015). Studying a single case in isolation may not give learners enough context or variation to understand the underlying principles involved fully (Alfieri et al., 2013). However, using case comparisons does not, *per se*, remediate mediocre ways of teaching. If the cases are not fully understood and the learner struggles to determine the relevant aspects, comparing cases might increase the intrinsic cognitive load compared to a single case, especially when multiple variables are involved (Schwartz and Bransford, 1998).

In contrast to the intrinsic cognitive load, the extraneous cognitive load is about *how* learning materials are designed (Sweller, 2010). The more superfluous or irrelevant information learners are presented with, the greater the possibility that they will not be able to distinguish between relevant and irrelevant information and will be distracted, which increases extraneous cognitive load. To minimize extraneous cognitive load for learners, it is therefore advisable to use design principles such as Mayer's, which are evidence-based and conducive to learning (Mayer, 2021). In relation to case comparisons, this means, for example, that in addition to reducing irrelevant information, the relevant information can be emphasized, e.g., by highlighting techniques (Rodemer et al., 2022).

The germane cognitive load describes the load that relates directly to learning as an activity and is considered productive (Paas and Sweller, 2014). The more a learner can focus on the learning itself, the more effectively links can be created. The germane cognitive load thus relates to the intrinsic cognitive load. Currently, there is an assumption "that germane cognitive load has a redistributive function from extraneous to intrinsic aspects of the task rather than imposing a load in its own right" (Sweller et al., 2019, p. 264). The lower the extraneous cognitive load is kept, the more space is given to the intrinsic cognitive load, which in turn results in an increased germane cognitive load (which is positive). However, this only becomes important with complex learning material, as the intrinsic cognitive load only becomes noticeable here. The simpler a task is, the lower the intrinsic cognitive load and the lower the germane cognitive load (Paas and Sweller, 2014). In relation to case comparisons, this means that the way in which the learning material is designed should be well considered so that there is more space for the germane cognitive load. Complex tasks can be chosen, whereby the complexity must match the prior knowledge and the capacity of the working memory to be able to generate effective learning and links (Sweller, 1994).

Overall, comparing cases as a task design can offload the working memory and engage multiple cognitive processes that are essential for learning and problem-solving when they match the capability of the learners (Roelle and Berthold, 2015).

2.3 Variation theory – instructional design principles

While Cognitive Load Theory (CLT) focuses on the capacity of working memory and how instructional design can be optimized to avoid cognitive overload, Variation theory is a learning theory that emphasizes the importance of variation in the design of instructional materials and activities and places emphasis on the importance of experiencing variations in the learning material to understand and discern the critical aspects of the content. While CLT is more about managing the quantity and complexity of information, Variation Theory is about the quality and structure of learning experiences. According to this theory, learners need to experience variations in the material they are studying in order to fully understand the underlying concepts, i.e., to abstract the relational connections beside surface similarities. Variation theory is based on the work of Swedish researcher Ference Marton and his colleagues, who developed the theory in the 1970s and 1980s (Marton, 1981). Marton (1981) was interested in understanding how students develop their understanding of complex concepts, and he observed that learners often struggle to transfer knowledge from one context to another.

Lo and Marton (2011) proposed that the key to understanding complex concepts is to focus on the variations in the material. They argued that learners need to experience different examples of a concept in order to fully understand it and develop a flexible understanding that can be applied to new contexts, advocating for a deep understanding of the subject matter instead of surface-level memorization.

Variation Theory of Learning helps further to support the use of case comparisons in chemistry education, as it emphasizes the importance of discerning critical features of a concept being taught. Using case comparisons (like different chemical reactions) helps students notice and understand the essential characteristics of each case; for example, contrasting an acid–base reaction with a redox reaction can help students understand the unique features of each type of reaction. Second, Variation Theory suggests that exposure to a range of examples, prototypical and non-prototypical examples, can help students see beyond single examples and support the ability to discriminate between different entities and recognize the significance of these differences. Certain elements become more salient to the viewer through variation, while other elements are kept invariant (Lo and Marton, 2011; Bussey et al., 2013), which allows learners to notice critical features more quickly (Bussey et al., 2013). Using case comparisons helps in achieving this by requiring students to apply principles to different scenarios, thereby promoting a deeper understanding of the underlying concepts (Roelle and Berthold, 2015; Bego et al., 2023). By focusing on these variations, variation theory aims to help learners develop a more nuanced and flexible understanding of the concept they are studying, which can be applied to new situations and contexts. The theory highlights the importance of experiencing variations in the material being studied in order to develop a flexible understanding that can be applied to new situations.

3 How good are students in comparing chemical reactions?

Multiple studies in chemistry education in the last decades documented that students when either not taught or not prompted appropriately to compare meaningfully, show a more surface-level-oriented comparison behavior when categorizing molecules or reactions. Moreover, by comparing two or more structures just because of their similar surface features, learners may overlook their properties (Talanquer, 2008; DeFever et al., 2015). Considering implicit properties and underlying processes of a reaction mechanism is crucial for higher modes of reasoning (Weinrich and Sevian, 2017) and leads to greater success when solving novel mechanistic problems (Grove et al., 2012). Stains and Talanquer (2007, 2008) compared the behaviors of undergraduate and graduate students while engaged in classifying different chemical representations and analyzed how often surface and deep-level attributes were used in the classification tasks. They determined that graduate students used more implicit information from the representations given than explicit ones for their classification. The most common approach used by undergraduates was a single attribute decision-making process. In the domain of organic chemistry, Domin et al. (2008) investigated the behavior of undergraduate students and experts while engaged in categorizing different cyclic or acyclic a-chloro derivatives of aldehydes and ketones. Consistent with Stains and Talanquer's findings, they found that students primarily categorized these compounds dichotomously by choosing a single surface-level attribute, such as aldehyde/ketone, cyclic/acyclic, or halogenated/non-halogenated. In Stains and Talanquer's study, experts tended to build similar categories as novices, also focusing on functional groups, but made the decision based on more implicit considerations, such as reactivity of the functional group toward the addition of nucleophiles. This increased focus on functional similarity, i.e., focusing on nucleophilicity/electrophilicity as well as reactivity of reactants, has been as well observed in various studies using card sorting activities (Graulich and Bhattacharyya, 2017; Galloway et al., 2018). It seems as if experts or advanced students in organic chemistry are able to generate more abstract schemas and store implicit information about molecules and reactions in bigger chunks, mirroring chemical reactivity patterns. Regarding investigating the development of expertise, a study revealed that successfully categorizing organic chemistry reaction cards is, with a large effect, correlated with the students' academic performance ($r = 0.62$). Moreover, the findings that academic performance is correlated with the successful online categorization were confirmed over the years (Lapierre et al., 2022). In a study from Graulich et al. (2019), learners were prompted to identify, for example, which two out of three nucleophiles would react similarly in a given substitution reaction. Thereby, the explicit properties of the given reactants matched or did match with the correct solutions. The findings revealed that students experienced greater challenges with items in which the structural representations of the correct answer did not share explicit similarity. Therefore, it might be helpful from time to time to use molecules or reactions with similar explicit surface features that are not undergoing similar reaction pathways or reactions that seem to be similar on the surface but undergo different pathways (Graulich and Schween, 2018). This could ideally induce cognitive dissonance in learners and challenge their strong focus on surface similarity. As a result, learners are required to use implicit properties to get to a proper solution and might be open to new explanatory concepts. Moreover, studies revealed that learners experience difficulties in activating the same concept knowledge in different contexts; thus, using a variety of molecules to introduce nucleophilicity might help students not to look only for negative charges and may help learners broaden their concept knowledge (Anzovino and Bretz, 2015; Popova and Bretz, 2018).

4 Designing and orchestrating cases

Case comparisons have been widely used as a task design across natural sciences and mathematics to foster students' ability to derive implicit features and weigh multiple arguments when reasoning. In their meta-analysis, Alfieri et al. (2013) found that case comparisons led to a higher number of identified variables than single cases ($d = 0.60$, 95% CI[0.47, 0.72]). Appropriately designed case comparisons offer the possibility to support learners to see how the same underlying principles apply to different chemical systems or to what extent reactions might occur differently (Graulich and Schween, 2018). This offers a chance to foster a deeper understanding of those principles and help students abstract from the explicit and sometimes misleading features of structural representations. Case comparisons seem to be more effective at the beginning rather than the end of an instructional topic, as it can prepare students to be sensitive to important features that need to be properly considered or to key features that must be transferred to new cases (Schwartz and Bransford, 1998; Schwartz et al., 2011).

When learners compare different chemical reactions that involve similar reactants and products but occur under different conditions, learners can experience how changes in conditions can affect the reaction rate and yield and relate this observation to the principles of thermodynamics and kinetics (Pölloth et al., 2022). Moreover, by comparing different cases, learners are forced to consider multiple influential factors and have to evaluate the similarities and differences. This can help them develop their ability to recognize patterns, make connections, and draw conclusions, which are essential skills in scientific inquiry and research (Alfieri et al., 2013). Figure 1 illustrates the differences between tasks based on single cases, contrasting cases with one variable and contrasting cases with two (or more) variables. When comparing a simple single case (Figure 1, upper part), the prompt is often only answered superficially, for example in stating as to whether reactions take place from a thermodynamic point of view. But when another case is added, such as changing the leaving group, this could be considered the simplest format of a case comparison, as only one variable of two displayed reactions is changed (Figure 1, middle part). This requires univariate reasoning and a strong focus on how the leaving group, in this case, the bromide or the chloride ion, is influencing the kinetic outcome of the reaction. Case comparisons can be adapted to more complex ones by changing a second variable, for example, several substituents or positions. The lower part of Figure 1 illustrates a case comparison that requires multivariate reasoning, as not only the leaving group (bromide or chloride-ion) but also the nature of the substrate (e.g., carbonyl vs. double bond) influences the reaction kinetic. Thus, learners have to weigh multiple arguments and justify their decisions based on the strength of implicit properties, in this case, mesomeric and inductive effects (Lieber and Graulich, 2022; Watts et al., 2023).

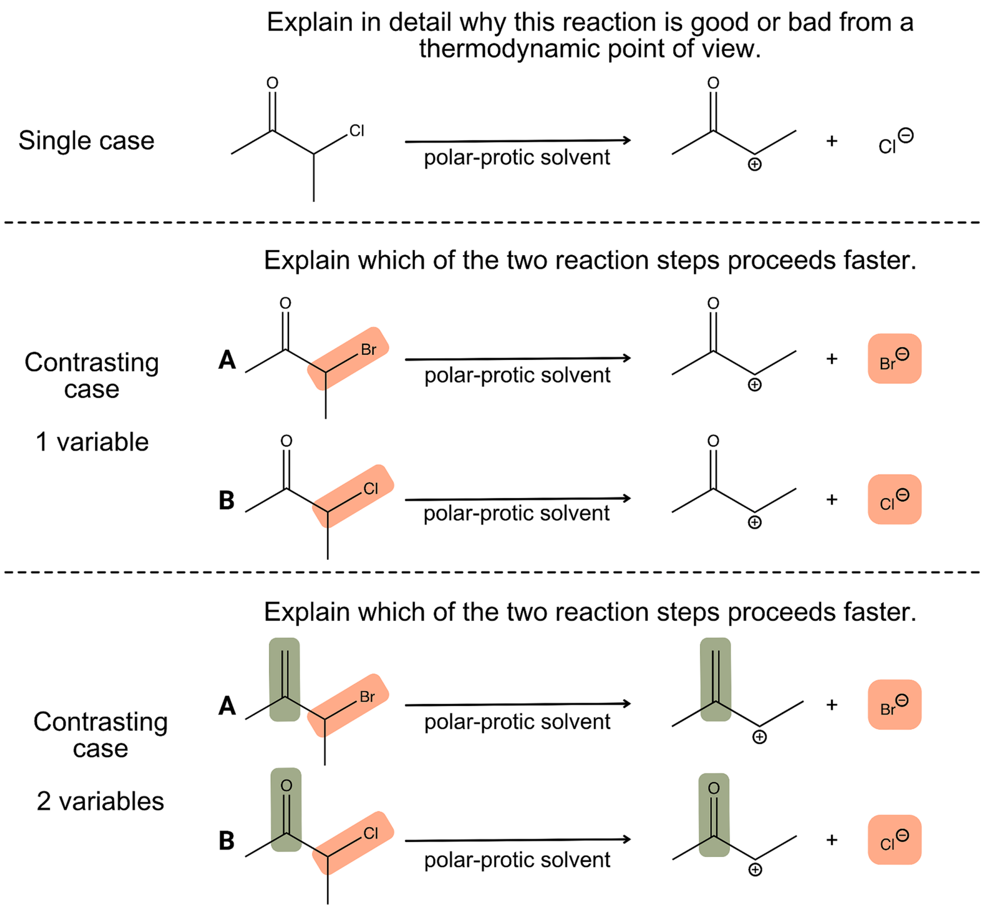


FIGURE 1
Example for a single case and case comparisons.

Case comparisons have been widely used in chemistry education studies, but the way in which these case comparisons were used differed (e.g., Bodé et al., 2019; Lieber and Graulich, 2022; Kranz et al., 2023). Figure 2 illustrates three different possibilities for using contrasting cases in argumentation processes. In the simplest case, an argument is divided into three parts: a claim, evidence and reasoning (evidence and reasoning can be combined as justification) (McNeill and Krajcik, 2012). One possibility for a task design involving case comparisons is that students compare two reactions at the beginning of the task to reason deeply about which reaction will proceed more likely. Thereby, the justification process can take place first and is guided by scaffolding which leads to a claim (Kranz et al., 2023) (see Figure 2, first example). Moreover, after comparing two reaction mechanisms at the beginning, it is also possible that learners first make a claim and justify their claim afterwards (Bodé et al., 2019; Deng and Flynn, 2021) (see Figure 2, second example). Besides comparing reactions at the beginning, it is also possible to build arguments on single reaction products of a reaction but contrast the reaction products at the end of the task. Thereby, students first claim if the respective reaction product is plausible or implausible, which is each justified with evidence and reasoning and compare the plausibilities of the reaction products in the end (see Figure 2, third example). This can lead to a revision of students' claims of most plausible reaction products toward a correct claim by weighing key concepts when contrasting them (Lieber et al., 2022; Lieber and Graulich, 2022). These studies indicate that the use of case comparison, at the beginning or at the end, has a beneficial effect for building arguments.

4.1 CPOE cycle – embedding case comparisons in inquiry processes

One way to combine the use of case comparisons with lab work is to embed these case comparisons in the CPOE cycle (Graulich and Schween, 2018), an adapted form of the Predict-Observe-Explain cycle (White and Gunstone, 2014) with an added "Compare" step. The cycle is based on learners first receiving a case comparison where they need to compare two given reactions (C), to predict (P) by generating a hypothesis which of the two reactions, for example, is faster than the other. This hypothesis can then be tested experimentally. By experimentally testing the hypotheses that have arisen from the case comparison, the outcome of the reactions is observed (O). Once the data has been analyzed, the final step takes place, in which conclusions are drawn about the previously formulated hypothesis based on the experimental results (E). Figure 3 illustrates the theoretical CPOE cycle by giving concrete examples how each step can look like, which is described in more detail in the following section.

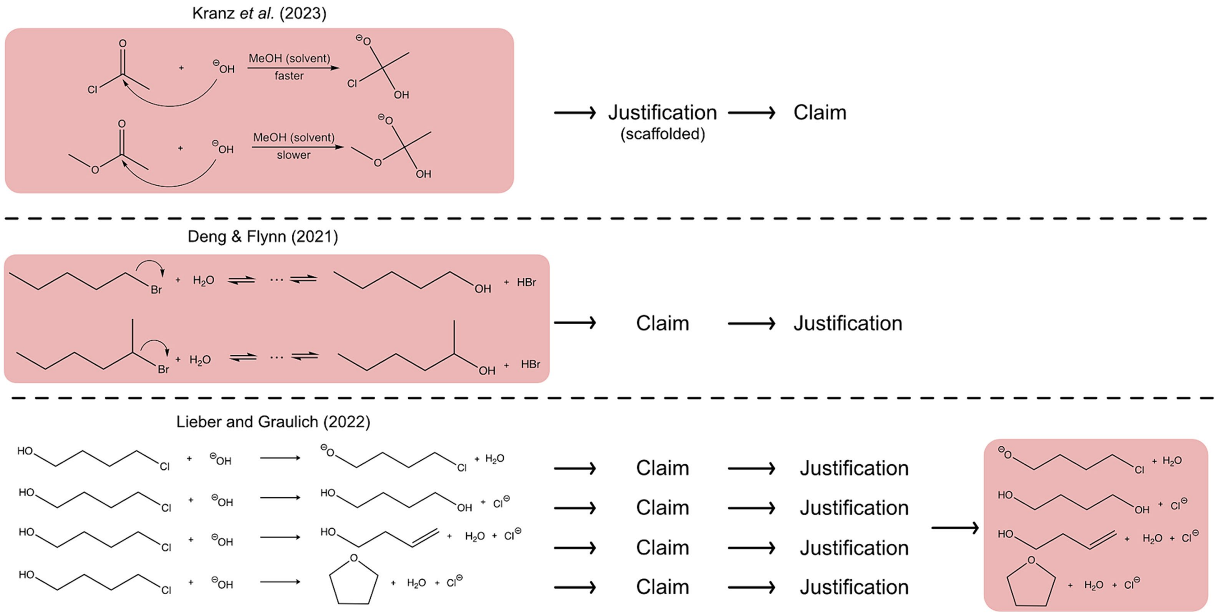


FIGURE 2
Illustration of different possibilities for the use of case comparisons in argumentation and reasoning processes. The red background highlights when the case comparison is used during the process.

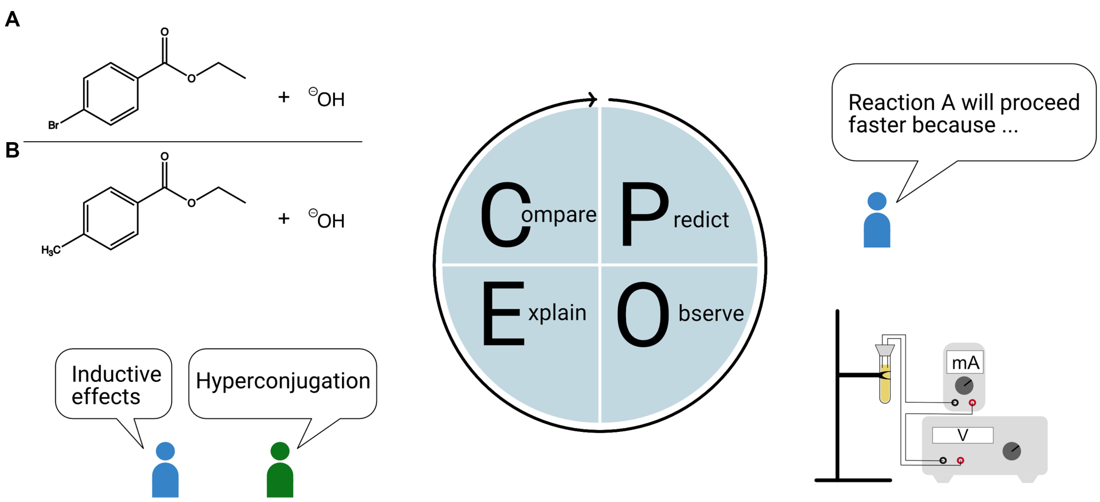


FIGURE 3
Embedding case comparisons in the CPOE cycle as illustrated with an example from Trabert and Schween (2020).

As the name suggests, however, this may not be a linear process with a defined end but a cycle that can be repeated based on new case comparisons. In this way, learners not only become familiar with scientific principles through independent experience, but the targeted choice of contrasting cases and experiments also enables a specific promotion of chemical concepts.

Schween's group has developed numerous experiments that make intermediate stages "visible," for example, based on conductivity measurements (*cf.* Trabert et al., 2023 for an overview). In each case, two or more reactions are compared with each other and learners are prompted to estimate the reaction with the higher reaction rate. Their work resulted in experimental case comparisons on electrophilic substitution on aromatic compounds, in which the sigma complexes were determined by conductivity measurements (Vorwerk et al., 2015), on the stability of carbenium ions, which makes intermediates directly and indirectly visible through color gradients as well as conductivity measurements (Schmitt et al., 2013), on the competition of primary and secondary haloalkanes in S_N2 reactions (Schmitt et al., 2018), as well as on electronic substituent effects in alkaline ester hydrolysis (Trabert and Schween, 2018). All these experiments can be used in a CPOE cycle. Figure 3 illustrates the linkage of Trabert and Schween's (2020) case comparisons of an alkaline ester hydrolysis, which is focused on inductive effects and their experimental design to the

CPOE cycle. Thereby, students first receive contrasting cases of ester hydrolysis, which differ in their substituents on the phenyl group (Figure 3, compare) Based on these two reactions, students have to predict which of the reactions proceed faster including a justification (Figure 3, predict). Students test their hypothesis afterwards in the laboratory with conductivity measurements (Figure 3, observe). Based on their observations, students are encouraged to explain the phenomenon and refer to their hypothesis (Figure 3, explain). When the shown cycle is used in teaching and learning, learners can transfer their knowledge of inductive effects into a second cycle. Therefore, learners can apply their knowledge of inductive effect on new reactions, which focus on the position of substituents. Thereby, learners complete the CPOE cycle a second time by comparing the position of substituents on aromatic compounds, predicting the reaction rate, observing the hypothesis by conducting experiments, and explaining the position dependency of inductive effects. The key aim of these experimental case comparisons is to engage learners in reflection about reaction rate, slowly increasing the sophistication of chemical concepts such as electronic effects that is not only supported by the experimental investigations but can also be advanced to other reactions and contexts. Those cases used in the lab and discussed in lecture might serve as a bridge between these two traditional course formats in organic chemistry.

5 Supporting students to learn meaningfully with case comparisons

When engaged in comparing, meaningful problem-solving requires attending to the relevant features of a representation, as well as linking the necessary implicit information to it (Mason et al., 2019). This may not be an intuitive process for students, as the connection between the feature of a carbonyl group (e.g., C=O) and its electron distribution has to be learned. The first visual selection process when looking at a structure is guided by learners' perception of saliency, their individual framing of what a given task entails, as well as their prior knowledge and the cognitive resources that a learner is able to activate (Bodé et al., 2019). Just comparing is not a one-size-fits-all solution, especially when implicit or functional information is more important than superficial features and might not result in the intended deeper reasoning about critical features (Bhattacharyya, 2023). For beginners, it might thus be necessary to be supported in attending to the relevant aspects, in order to decrease the extraneous and intrinsic load. The Cognitive Theory of Multimedia Learning (CTML) by Mayer (2021) allows informed instructional design to support students in these aspects. The key assumption of the CTML is that human cognition proceeds by two channels, a visual and a verbal channel, that need to be optimally synchronized in learning. It is thus beneficial to present information both visually, which we typically do with structural representations and verbally (e.g., written or spoken explanations), to engage both channels. Both channels have limited capacity, meaning that learners can only process a limited amount of information at a time. In the context of case comparisons, it is important not to overwhelm students with too much information at once and to guide their attention to the relevant aspect in the visual and verbal channel (Rodemer et al., 2020; Eckhard et al., 2022). Thus, both theories, the CLT as well as the CTML, support the same instructional design principles: guiding students visually and conceptually through a task, to make a task accessible for actual learning.

5.1 Visual attention guidance

Guiding learners to attend to the relevant features, i.e., important functional groups involved in a reaction, can be achieved by multiple means, such as simply signaling or highlighting the relevant areas of the representation [i.e., signaling principle as described by Mayer (2021)], e.g., by zooming in or out, spotlights, coloring, added on-screen text or symbols. Others used experts' eye gaze as a model for the learner, as used in the context of medicine (Jarodzka et al., 2012; Gegenfurtner et al., 2017), whereas transferring this idea to learning organic reaction mechanisms has not yet been convincing (Graulich et al., 2022). By "signaling" (highlighting key structural features in a static or dynamic fashion) students can focus on these key features of the representation and reduce their attentional focus to the rest of the structure, thus, reducing their extraneous cognitive load, if they are not attending to everything all at once (Richter et al., 2016; Schneider et al., 2018). It can also allow us to model a certain sequence of comparing by highlighting, for example, a starting point of comparison and then the sequential decoding process. Although attending to the relevant features is a key step. Implicit chemical properties cannot be read out of the functional group but need to be linked to it. When the attention of the learner is on the relevant features of a representation, the respective implicit information needs to be added, either in terms of verbal or written information. This is in line with the dual channel assumption of the CTML, providing highlighting for the visual features and chemical information for the verbal channel, as well as presenting it at the same time, i.e., the contiguity principle (Mayer and Fiorella, 2014). Some instructors might intuitively use highlighting techniques by pointing toward the representational features on the blackboard and explaining simultaneously or by adding conceptual information, such as pK_a or partial charges on the board. Redirecting a learner's attention to the relevant aspects, thus, can be complex, as decisions have to be made that cannot just be guided by the salience of a functional group, and conceptual information needs to be linked to make a purposeful selection.

In a quantitative study, we tested if a highlighting technique actually supports students to attend to relevant areas of organic chemistry case comparisons and solve them more successfully. Thus, we created tutorial videos with case comparisons and used a dynamic moving dot highlighting representational features, which was synchronized with the information given as a verbal explanation in parallel (Rodemer et al., 2020; Eckhard et al., 2022). The study could document that all students in the study were profiting from the given verbal explanation, but especially low performing students profited from the highlighting. Following students while watching the videos with highlighting with the help of eye-tracking could show that the attention to relevant areas is focused over the entire time of the video, and the perceived extraneous cognitive load is decreased (Rodemer et al., 2022). These overall results illustrated that beginners need more support in decoding the molecular structures that we use in organic chemistry, and guiding their attention is key for a decreased extraneous cognitive load. Besides using eye-tracking as an analytical lens to track students' attention, using it in instruction might help students understand their own viewing behavior. In an eye-tracking study conducted by Hansen et al. (2019), they investigated how students view and critique different animations of redox reactions and precipitation reactions. After their reasoning process, students received visual feedback on their own viewing behavior. Hansen et al. (2019) revealed that viewing this feedback helped the students to be critical about their own viewing behavior and to deepen the critique regarding the animations shown.

5.2 Conceptual guidance

Further breaking down the reasoning process with case comparisons into manageable parts can help students process the information more effectively (Belland, 2017). A simple nucleophilic substitution, taught in an introductory organic chemistry course, for instance, requires the consideration of three main influential factors, i.e., leaving group ability, nucleophilicity, substrate effects, and the cause-effect relationships that determine the reactivity in this type of mechanism. Thus, a lot needs to be considered by the learners. Using case comparison can have positive effects on students' engagement with the conceptual knowledge, as it shifts the focus onto implicit and influential factors of the organic reaction mechanism (Watts et al., 2021). However, if we expect students to reason in a particular way, i.e., building cause-effect relationships, and connect different concepts and properties, we need to be explicit how students should integrate these multiple pieces of knowledge. Developing mastery requires explicit learning of how to create those mechanistic explanations (Cooper, 2015). Thus, supporting students in solving case comparisons should acknowledge the complexity and reasoning steps required and ideally make these steps transparent through a scaffold (Caspari et al., 2018; Kranz et al., 2023). Scaffolding is a known technique widely used as an instruction in science education (*cf.* Lin et al., 2012; Wilson and Devereux, 2014) and helps students to slow down the decision-making process and gives students the opportunity to activate necessary conceptual and procedural knowledge (Rittle-Johnson and Star, 2007; Rittle-Johnson and Star, 2009; Shemwell et al., 2015; Chin et al., 2016). A scaffold for the case comparisons illustrated therein thus can guide the learner through the different considerations necessary to make a claim about the outcome of a case: (1) describing the chemical changes in the given cases; (2) explicitly stating the overall goal of comparison (task prompt); (3) naming the similarities and differences; (4) stating the role of the influential factors (i.e., implicit properties); (5) explaining and contrasting the influences of the implicit properties; (6) stating how the transition state is affected to refer to the energetic account and (7) making a final claim about the reactivity of both reactions (Bernholt et al., 2023).

Various studies already documented the positive effect of using scaffolding with case comparisons on students' reasoning. In prior studies, we used a scaffold grid, represented by a worksheet with empty boxes, which visually connects the structural differences, changes, and cause-effect relations (Caspari et al., 2018). By utilizing this grid, students can systematically relate each structural difference to each ongoing change, verbalizing the influence of the structural difference on the change. We compared how students are reasoning through contrasting cases with and without a scaffold and could observe that students' reasoning is more guided and includes the consideration of more implicit properties and influential effects when solving a contrasting case with a scaffold (Caspari et al., 2018). This structured approach helps students avoid jumping to the final answer without considering the underlying reasons. A mixed-methods study could confirm that especially students with a low prior knowledge profited from working with a scaffold and had a higher learning gain, whereas it does also not harm those with higher prior knowledge (Kranz et al., 2023). Lieber et al. (2022) advanced a scaffold further by acknowledging students' individual needs when arguing about alternative reaction pathways. Those adaptive scaffolds could show that more individualized instruction when using different cases in organic chemistry might be a new avenue to improve teaching.

6 Conclusion

Comparing the outcome of organic reactions, the strength of nucleophiles, or the reaction rate is at the core of organic chemistry. Through asking comparative questions, we gain insight into reaction processes and reactivity patterns, which allow us to predict and explain novel ones. Learning a collection of seemingly unrelated reactions, or even name reactions in organic chemistry, as often the practice in organic chemistry classes, does not allow learners or make it more difficult to understand and derive the underlying principles that govern reactions. Structure mapping theory tells us, that our cognitive structure is barely made to extract with ease a conceptual similarity just by looking at reactions. An explicit surface similarity will always be more salient for an inexperienced learner. The limited capacity of our working memory additionally affects how much effort we can put into learning and understanding. Purposefully comparing and reasoning through case comparisons can help regain the focus on conceptual understanding in organic chemistry but has not yet been fully explored in instructional design as well as assessments. Multiple studies have documented the potential of using case comparisons compared to more traditional task formats, characterized the type of reasoning that can be elicited from learners, and integrated case comparisons into laboratory experiments. We illustrated therein how, based on various theories of cognition and instruction, comparing can serve as a valuable process for selecting attention, limiting the extraneous cognitive load as well as focusing on implicit and explicit properties and cause-effect relationships. This process of comparing can further be supported, following the principles of the Cognitive Theory of Multimedia Learning, by highlighting relevant features of representations through cueing techniques or providing scaffolding by sequentially guiding students through solving a case comparison. This perspective was meant to consolidate the current state of the art around the use of case comparison to provide instructors with a theory-informed basis for changing their practice and exploring comparing.

Author contributions

NG: Conceptualization, Funding acquisition, Visualization, Writing – original draft, Writing – review & editing. LL: Conceptualization, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. NG would like to thank the German Research Foundation DFG (Deutsche Forschungsgemeinschaft) for funding (project number: 446349713).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Afagh, N. A., and Yudin, A. K. (2010). Chemoselectivity and the curious reactivity preferences of functional groups. *Angew. Chem. Int. Ed.* 49, 262–310. doi: 10.1002/anie.200901317

Alfieri, L., Nokes-Malach, T. J., and Schunn, C. D. (2013). Learning through case comparisons: a Meta-analytic review. *Educ. Psychol.* 48, 87–113. doi: 10.1080/00461520.2013.775712

Anzovino, M. E., and Bretz, S. L. (2015). Organic chemistry students' ideas about nucleophiles and electrophiles: the role of charges and mechanisms. *Chem. Educ. Res. Pract.* 16, 797–810. doi: 10.1039/C5RP00113G

Anzovino, M. E., and Bretz, S. L. (2016). Organic chemistry students' fragmented ideas about the structure and function of nucleophiles and electrophiles: a concept map analysis. *Chem. Educ. Res. Pract.* 17, 1019–1029. doi: 10.1039/C6RP00111D

Baddeley, A. (2010). Working memory. *Curr. Biol.* 20, R136–R140. doi: 10.1016/j.cub.2009.12.014

Bego, C. R., Chastain, R. J., and DeCaro, M. S. (2023). Designing novel activities before instruction: use of contrasting cases and a rich dataset. *Br. J. Educ. Psychol.* 93, 299–317. doi: 10.1111/bjep.12555

Belland, B. R., (2017), *In instructional scaffolding in STEM education: strategies and efficacy evidence*, Cham: Springer International Publishing.

Bernholt, S., Eckhard, J., Rodemer, M., Langner, A., Asmussen, G., and Graulich, N. (2023). *In digital learning and teaching in chemistry* The Royal Society of Chemistry.

Bhattacharyya, G. (2014). Trials and tribulations: student approaches and difficulties with proposing mechanisms using the electron-pushing formalism. *Chem. Educ. Res. Pract.* 15, 594–609. doi: 10.1039/C3RP00127J

Bhattacharyya, G. (2023) "Assessment of assessment in organic chemistry - Review and analysis of predominant problem types related to reactions and mechanisms", in *Student reasoning in organic chemistry*. eds. N. Graulich and G. Shultz (Cambridge (UK): Roy. Soc. Chemistry), 269–284.

Bodé, N. E., Deng, J. M., and Flynn, A. B. (2019). Getting past the rules and to the WHY: causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *J. Chem. Educ.* 96, 1068–1082. doi: 10.1021/acs.jchemed.8b00719

Bussey, T. J., Orgill, M., and Crippen, K. J. (2013). Variation theory: a theory of learning and a useful theoretical framework for chemical education research. *Chem. Educ. Res. Pract.* 14, 9–22. doi: 10.1039/C2RP20145C

Caspari, I., Kranz, D., and Graulich, N. (2018). Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework. *Chem. Educ. Res. Pract.* 19, 1117–1141. doi: 10.1039/C8RP00131F

Chin, D. B., Chi, M., and Schwartz, D. L. (2016). A comparison of two methods of active learning in physics: inventing a general solution versus compare and contrast. *Instr. Sci.* 44, 177–195. doi: 10.1007/s11251-016-9374-0

Cooper, M. M. (2015). Why ask why? *J. Chem. Educ.* 92, 1273–1279. doi: 10.1021/acs.jchemed.5b00203

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, 699–721. doi: 10.1002/tea.21093

DeFever, R. S., Bruce, H., and Bhattacharyya, G. (2015). Mental Rolodexing: senior chemistry Majors' understanding of chemical and physical properties. *J. Chem. Educ.* 92, 415–426. doi: 10.1021/ed500360g

Deng, J. M., and Flynn, A. B. (2021). Reasoning, granularity, and comparisons in students' arguments on two organic chemistry items. *Chem. Educ. Res. Pract.* 22, 749–771. doi: 10.1039/D0RP00320D

Domin, D. S., Al-Masum, M., and Mensah, J. (2008). Students' categorizations of organic compounds. *Chem. Educ. Res. Pract.* 9, 114–121. doi: 10.1039/B806226A

Eckhard, J., Rodemer, M., Bernholt, S., and Graulich, N. (2022). What do University students truly learn when watching tutorial videos in organic chemistry? An exploratory study focusing on mechanistic reasoning. *J. Chem. Educ.* 99, 2231–2244. doi: 10.1021/acs.jchemed.2c00076

Galloway, K. R., Leung, M. W., and Flynn, A. B. (2018). A comparison of how undergraduates, graduate students, and professors organize organic chemistry reactions. *J. Chem. Educ.* 95, 355–365. doi: 10.1021/acs.jchemed.7b00743

Gegenfurtner, A., Lehtinen, E., Jarodzka, H., and Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Comput. Educ.* 113, 212–225. doi: 10.1016/j.compedu.2017.06.001

Gentner, D., (1989), *Similarity and analogical reasoning*, New York: Cambridge University Press.

Gentner, D., Loewenstein, J., and Thompson, L. (2003). Learning and transfer: a general role for analogical encoding. *J. Educ. Psychol.* 95, 393–408. doi: 10.1037/0022-0663.95.2.393

Gentner, D., and Markman, A. B. (1997). Structure mapping in analogy and similarity. *Am. Psychol.* 52, 45–56. doi: 10.1037/0003-066X.52.1.45

Gick, M. L., and Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cogn. Psychol.* 15, 1–38. doi: 10.1016/0010-0285(83)90002-6

Goodwin, W. (2008). Structural formulas and explanation in organic chemistry. *Found. Chem.* 10, 117–127. doi: 10.1007/s10698-007-9033-2

Goodwin, W. (2010). How do structural formulas embody the theory of organic chemistry? *Br. Soc. Philos. Sci.* 61, 621–633. doi: 10.1093/bjps/axp052

Graulich, N., and Bhattacharyya, G. (2017). Investigating students' similarity judgments in organic chemistry. *Chem. Educ. Res. Pract.* 18, 774–784. doi: 10.1039/C7RP00055C

Graulich, N., Hedtrich, S., and Harzenetter, R. (2019). Explicit versus implicit similarity - exploring relational conceptual understanding in organic chemistry. *Chem. Educ. Res. Pract.* 20, 924–936. doi: 10.1039/C9RP00054B

Graulich, N., Rodemer, M., Eckhard, J., and Bernholt, S., (2022), *Eye-Tracking in der Mathematik- und Naturwissenschaftsdidaktik: Forschung und Praxis*, Berlin, Heidelberg: Springer Berlin Heidelberg.

Graulich, N., and Schween, M. (2018). Concept-oriented task design: making purposeful case comparisons in organic chemistry. *J. Chem. Educ.* 95, 376–383. doi: 10.1021/acs.jchemed.7b00672

Grove, N. P., Cooper, M. M., and Cox, E. L. (2012). Does mechanistic thinking improve student success in organic chemistry? *J. Chem. Educ.* 89, 850–853. doi: 10.1021/ed200394d

Hansen, S., Hu, B., Riedlova, D., Kelly, R., Akaygun, S., and Villalta-Cerdas, A. (2019). Critical consumption of chemistry visuals: eye tracking structured variation and visual feedback of redox and precipitation reactions. *Chem. Educ. Res. Pract.* 20, 837–850. doi: 10.1039/C9RP00015A

Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., et al. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instr. Sci.* 40, 813–827. doi: 10.1007/s11251-012-9218-5

Kalyuga, S., Chandler, P., and Sweller, J. (1998). Levels of expertise and instructional design. *Hum. Factors* 40, 1–17. doi: 10.1518/001872098779480587

Keith, J. A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Müller, K.-R., et al. (2021). Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* 121, 9816–9872. doi: 10.1021/acs.chemrev.1c00107

Kranz, D., Schween, M., and Graulich, N. (2023). Patterns of reasoning - exploring the interplay of students' work with a scaffold and their conceptual knowledge in organic chemistry. *Chem. Educ. Res. Pract.* 24, 453–477. doi: 10.1039/D2RP00132B

Lapierre, K. R., and Flynn, A. B. (2020). An online categorization task to investigate changes in students' interpretations of organic chemistry reactions. *J. Res. Sci. Teach.* 57, 87–111. doi: 10.1002/tea.21586

Lapierre, K. R., Streja, N., and Flynn, A. B. (2022). Investigating the role of multiple categorization tasks in a curriculum designed around mechanistic patterns and principles. *Chem. Educ. Res. Pract.* 23, 545–559. doi: 10.1039/D1RP00267H

Laszlo, P. (2002). Describing reactivity with structural formulas, or when push comes to shove. *Chem. Educ. Res. Pract.* 3, 113–118. doi: 10.1039/B2RP90009B

Lieber, L., and Graulich, N. (2022). Investigating Students' argumentation when judging the plausibility of alternative reaction pathways in organic chemistry. *Chem. Educ. Res. Pract.* 23, 38–54. doi: 10.1039/D1RP00145K

Lieber, L., Ibraj, K., Caspari-Gnann, I., and Graulich, N. (2022). Closing the gap of organic chemistry Students' performance with an adaptive scaffold for argumentation patterns. *Chem. Educ. Res. Pract.* 23, 811–828. doi: 10.1039/D2RP00016D

Lin, T.-C., Hsu, Y.-S., Lin, S.-S., Changlai, M.-L., Yang, K.-Y., and Lai, T.-L. (2012). A review of empirical evidence on scaffolding for science education. *Int. J. Sci. Math. Educ.* 10, 437–455. doi: 10.1007/s10763-011-9322-z

Lo, M. L., and Marton, F. (2011). Towards a science of the art of teaching: using variation theory as a guiding principle of pedagogical design. *Int. J. Lesson Learn. Stud.* 1, 7–22. doi: 10.1108/20468251211179678

Marton, F. (1981). Phenomenography—describing conceptions of the world around us. *Instr. Sci.* 10, 177–200. doi: 10.1007/BF00132516

Mason, B., Rau, M. A., and Nowak, R. (2019). Cognitive task analysis for implicit knowledge about visual representations with similarity learning methods. *Cogn. Sci.* 43:e12744. doi: 10.1111/cogs.12744

Mayer, R. E. (2021). *Multimedia learning.* Cambridge: Cambridge University Press.

Mayer, R. E., and Fiorella, L., (2014). *The Cambridge handbook of multimedia learning*, New York: Cambridge University Press.

McNeill, K. L., and Krajcik, J., (2012), *Book study facilitator's guide: Supporting grade 5–8 students in constructing explanations in science: the claim, evidence and reasoning framework for talk and writing*, New York: Pearson Allyn & Bacon.

Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educ. Psychol.* 38, 1–4. doi: 10.1207/S15326985EP3801_1

Paas, F., and Sweller, J., (2014). "Implications of Cognitive Load Theory for Multimedia Learning", in *The Cambridge handbook of multimedia learning, 2nd Edition*. Ed. R. Mayer (New York: Cambridge University Press), 27–42.

Pölloth, B., Häfner, M., and Schwarzer, S. (2022). At the same time or one after the other?–exploring reaction paths of nucleophilic substitution reactions using historic insights and experiments. *Chemkon* 29, 77–83. doi: 10.1002/ckon.202100060

Popova, M., and Bretz, S. L. (2018). Organic chemistry Students' understandings of what makes a good leaving group. *J. Chem. Educ.* 95, 1094–1101. doi: 10.1021/acs.jchemed.8b00198

Richter, J., Scheiter, K., and Eitel, A. (2016). Signaling text-picture relations in multimedia learning: a comprehensive meta-analysis. *Educ. Res. Rev.* 17, 19–36. doi: 10.1016/j.edurev.2015.12.003

Rittle-Johnson, B., and Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *J. Educ. Psychol.* 99, 561–574. doi: 10.1037/0022-0663.99.3.561

Rittle-Johnson, B., and Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *J. Chem. Educ.* 101:529. doi: 10.1037/a0014224

Rodemer, M., Eckhard, J., Graulich, N., and Bernholt, S. (2020). Decoding case comparisons in organic chemistry: eye-tracking Students' visual behavior. *J. Chem. Educ.* 97, 3530–3539. doi: 10.1021/acs.jchemed.0c00418

Rodemer, M., Lindner, M. A., Eckhard, J., Graulich, N., and Bernholt, S. (2022). Dynamic signals in instructional videos support students to navigate through complex representations: an eye-tracking study. *Appl. Cogn. Psychol.* 36, 852–863. doi: 10.1002/acp.3973

Roelle, J., and Berthold, K. (2015). Effects of comparing contrasting cases on learning from subsequent explanations. *Cogn. Instr.* 33, 199–225. doi: 10.1080/07370008.2015.1063636

Schmitt, C., Bender, M., Trabert, A., and Schween, M. (2018). What's the effect of steric hindrance? Experimental comparison of reaction rates of primary and secondary alkyl halides in competing SN2 reactions. *Chemkon* 25, 231–237. doi: 10.1002/ckon.201800012

Schmitt, C., Wißner, O., and Schween, M. (2013). Carbenium ions as reactive intermediates – an (experimental) access to a deeper understanding of organic reactions. *Chemkon* 20, 59–65. doi: 10.1002/ckon.201310195

Schneider, S., Beege, M., Nebel, S., and Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educ. Res. Rev.* 23, 1–24. doi: 10.1016/j.edurev.2017.11.001

Schwartz, D. L., and Bransford, J. D. (1998). A time for telling. *Cogn. Instr.* 16, 475–5223. doi: 10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., and Chin, D. B. (2011). Practicing versus inventing with contrasting cases: the effects of telling first on learning and transfer. *J. Educ. Psychol.* 103, 759–775. doi: 10.1037/a0025140

Shemwell, J. T., Chase, C. C., and Schwartz, D. L. (2015). Seeking the general explanation: a test of inductive activities for learning and transfer. *J. Res. Sci. Teach.* 52, 58–83. doi: 10.1002/tea.21185

Stains, M., and Talanquer, V. (2007). Classification of chemical substances using particulate representations of matter: an analysis of student thinking. *Int. J. Sci. Educ.* 29, 643–661. doi: 10.1080/09500690600931129

Stains, M., and Talanquer, V. (2008). Classification of chemical reactions: stages of expertise. *J. Res. Sci. Teach.* 45, 771–793. doi: 10.1002/tea.20221

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* 4, 295–312. doi: 10.1016/0959-4752(94)90003-5

Sweller, J., (2003), *In psychology of learning and motivation: Advances in research and theory*, San Diego, USA: Elsevier Science, pp. 215–266.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* 22, 123–138. doi: 10.1007/s10648-010-9128-5

Sweller, J., and Chandler, P. (1994). Why some material is difficult to learn. *Cogn. Instr.* 12, 185–233. doi: 10.1207/s1532690xci1203_1

Sweller, J., van Merriënboer, J. J., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 261–292. doi: 10.1007/s10648-019-09465-5

Talanquer, V. (2008). Students' predictions about the sensory properties of chemical compounds: additive versus emergent frameworks. *Sci. Educ.* 92, 96–114. doi: 10.1002/sce.20235

Talanquer, V. (2017). Concept inventories: predicting the wrong answer may boost performance. *J. Chem. Educ.* 94, 1805–1810. doi: 10.1021/acs.jchemed.7b00427

Trabert, A., Schmitt, C., and Schween, M. (2023). "Building bridges between tasks and flasks - Design of a coherent experiment-supported learning environment for deep reasoning in organic chemistry", in *Student reasoning in organic chemistry*. Eds. N. Graulich and G. Shultz (Cambridge (UK): Roy. Soc. Chemistry), 248–266.

Trabert, A., and Schween, M. (2018). How do electronic substituent effects work?-design of a concept-based approach applying inventing with contrasting cases to the example of alkaline ester hydrolysis. *Chemkon* 25, 334–342. doi: 10.1002/ckon.201800010

Trabert, A., and Schween, M. (2020). How do electronic substituent effects work?–additional contrasting cases for a differentiated inquiry illustrated by the example of alkaline ester hydrolysis. *Chemkon* 27, 22–33. doi: 10.1002/ckon.201800076

Vorwerk, N., Schmitt, C., and Schween, M. (2015). Understanding electrophilic aromatic substitutions– sigma-complexes as (experimental) key structures. *Chemkon* 22, 59–68. doi: 10.1002/ckon.201410237

Watts, F. M., Dood, A. J., Shultz, G. V., and Rodriguez, J.-M. G. (2023). Comparing student and generative artificial intelligence Chatbot responses to organic chemistry writing-to-learn assignments. *J. Chem. Educ.* 100, 3806–3817. doi: 10.1021/acs.jchemed.3c00664

Watts, F. M., Zaimi, I., Kranz, D., Graulich, N., and Shultz, G. V. (2021). Investigating students' reasoning over time for case comparisons of acyl transfer reaction mechanisms. *Chem. Educ. Res. Pract.* 22, 364–381. doi: 10.1039/D0RP00298D

Weinrich, M. L., and Sevian, H. (2017). Capturing students' abstraction while solving organic reaction mechanism problems across a semester. *Chem. Educ. Res. Pract.* 18, 169–190. doi: 10.1039/C6RP00120C

White, R., and Gunstone, R., (2014), *Probing Understanding*, New York: Routledge.

Wilson, K., and Devereux, L. (2014). Scaffolding theory: high challenge, high support in academic language and learning (ALL) contexts. *J. Acad. Lang. Learn.* 8, A91–A100.

frontiers | Frontiers in Education

TYPE Perspective
PUBLISHED 02 May 2024
DOI 10.3389/feduc.2024.1401835

Check for updates

OPEN ACCESS

EDITED BY
Brett McCollum,
Thompson Rivers University, Canada

REVIEWED BY
Riley Petillion,
University of British Columbia, Canada
Jennifer MacDonald,
Dalhousie University, Canada

*CORRESPONDENCE
Maia Popova
m_popova@uncg.edu

RECEIVED 16 March 2024
ACCEPTED 23 April 2024
PUBLISHED 02 May 2024

CITATION
Popova M (2024) Bridging chemistry education research and practice through research-practice partnerships. *Front. Educ.* 9:1401835. doi: 10.3389/feduc.2024.1401835

Bridging chemistry education research and practice through research-practice partnerships

Maia Popova*

Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro, NC, United States

This perspective article is a call to establish research-practice partnerships (RPPs) to foster collaborations between instructors and education researchers to tune into the needs of practice, share evidence-based practices, and solve modern organic chemistry education problems. I begin the article by discussing some limitations of the traditional approach of "translating" research *into* practice and suggest RPPs as an alternative model for "bridging" research *and* practice. Importantly, RPPs have been shown to address persistent problems of practice and improve educational outcomes. While more common at the secondary level, RPPs are rarely leveraged in post-secondary chemistry education. The article goes on to provide a concrete and relevant context for potential future RPP efforts to improve aspects of organic chemistry education—RPPs between education researchers and organic chemistry instructors to work toward designing, administering, and testing interventions to support learners' representational competence (RC). RC is a set of skills that allow for the reflective use of a variety of representations to think about, communicate, and act on chemical phenomena. Current instruction often falls short of effectively supporting learners in developing RC. It is often tacitly assumed that learners will develop RC without explicit instruction that scaffolds the development of the RC skills. While it is important to improve the teaching about and with representations, implementing innovative pedagogical approaches can be challenging, particularly when instructors feel isolated in their efforts within their work environments. The RPP model could catalyze solutions to these challenges by pooling diverse expertise, thus enabling more robust and sustainable educational innovations.

KEYWORDS

post-secondary education, organic chemistry, representational competence, professional development, research-practice partnerships

1 The problem with traditional approaches to translating research into practice

In the ever-evolving landscape of higher education, the importance of implementing innovative, evidence-based pedagogical approaches cannot be overstated. Educators must continuously seek out and apply the best available evidence to their teaching practices, thereby enhancing the quality of education they provide to improve student learning outcomes. At the same time, implementing innovative pedagogical approaches can be challenging, particularly when instructors feel isolated in their efforts within their work environments. Even if instructors recognize the potential benefits of a particular innovation, the absence of support can impede their well-intentioned curricular reform efforts or efforts to improve their teaching practices (Fairweather, 2008). These barriers can result in a lack of motivation needed to initiate change within their instruction (Shadle et al., 2017). To mitigate these barriers, educators and curriculum developers require knowledge, skills, resources, and support. This need has been previously emphasized by the National Research Council:

> "*The translation of research findings into forms useful for educational practice… will require large-scale, systematic experimentation and demonstration to transform knowledge about human learning and the development of competence into the working vocabulary of teachers…*" (Committee on a Feasibility Study for a Strategic Education Research Program, National Research Council, 1999, p. 3).

Today, one of the most common mechanisms to provide such support is faculty professional development initiatives that typically involve structured programs, workshops, or seminars, where education researchers directly impart knowledge to instructors in a one-directional manner. A significant limitation of such traditional professional development lies in the assumption that knowledge transfer is straightforward and that teaching practices can be directly informed and transformed by simply exposing instructors to new educational theories or research findings. The conventional "translation" metaphor of research into practice provides an impoverished way of understanding the complex relationship between research and practice (Penuel et al., 2015). While effective to a degree, the "translation" approach overlooks several critical aspects that are essential for meaningful pedagogical growth as it fails to account for the specific contextual challenges and opportunities within different teaching environments and results in a lack of instructors' engagement and ownership, missed opportunities for collaboration between educational researchers and instructors, inadequate attention to implementation challenges, and issues with the sustainability of instructional innovations (Burbank and Kauchak, 2003; Chicoine, 2004; Webster-Wright, 2009; Coburn et al., 2013; Penuel et al., 2015; Coburn and Penuel, 2016; Rodriguez and Towns, 2019; Johnson, 2022).

Major task force reports have called for changes in how we conceptualize the "translation" of research into practice (National Research Council, 1999; Donovan et al., 2013). Researchers, administrators, and stakeholders need to better account for the complex challenges researchers and practitioners face when using research to drive educational improvement (National Research Council, 1999; Coburn and Stein, 2010; Donovan et al., 2013). These efforts should go beyond the predominantly used "teaching as telling," "one-directional," or "one-shot" professional development models, especially given the evidence of the failure of some of these professional development approaches (Lovitt and Clarke, 1988; Fullan and Stiegelbauer, 1991; Johnson, 1998). The evolving landscape of higher education and the diverse needs of modern learners call for more innovative, flexible, and inclusive professional development approaches, as well as for a broader range of mechanisms through which faculty professional development can emerge. This perspective article is a call for the chemistry education community to consider an alternative faculty professional development model that has the potential to address the limitations described above—research-practice partnerships (Coburn et al., 2013; Penuel et al., 2015).

2 Research-practice partnerships as an alternative model for bridging research and practice

Research-practice partnerships (RPPs) are long-term collaborations, often between educators and education researchers, that have been shown to address persistent problems of practice, and improve educational outcomes (Tseng, 2012; Coburn et al., 2013; Donovan et al., 2013; Fishman et al., 2013). RPP members establish an external community of like-minded colleagues outside of one's institution which becomes an important force in promoting educational innovation. RPP members engage in processes of collaboration and exchange that are both messier and often more transformative than the traditional "one-directional translation" of research into practice. RPPs are characterized by mutual goals that focus on problems of practice rather than gaps in research and theory. Importantly, these goals and efforts often evolve through interactions between RPP members, rather than being defined fully ahead of time (Penuel et al., 2015).

RPPs offer opportunities for growth and development not only for instructors but also for education researchers. By working closely with practitioners, researchers gain direct insights into the problems of practice. This deep engagement with the realities of educational settings fosters a richer understanding of the context-dependent nature of learning and teaching, enhancing the relevance and applicability of research findings. Effective RPPs establish shared authority where goals, work, and interactions are jointly negotiated, with carefully elaborated roles, routines, and protocols for engagement (Coburn and Penuel, 2016). Ultimately, this symbiosis between research and practice enriches the academic and research communities, fostering an environment where knowledge creation and pedagogical excellence coalesce more effectively.

While there is promising evidence on the impact of RPPs' interventions on student learning at the secondary level (Fishman et al., 2003; Yarnall et al., 2006; Geier et al., 2008; Snow et al., 2009; Barab et al., 2010; Booth et al., 2015), RPPs are uncommon in post-secondary chemistry education. Below I propose a potential direction for future RPPs to improve aspects of post-secondary organic chemistry education.

3 Representational competence as a potential context for future research-practice partnerships to improve aspects of organic chemistry education

Organic Chemistry is known for its high attrition rates (Lovecchio and Dundes, 2002; Jones and Gellene, 2005; Grove et al., 2008). It is imperative to rethink and redesign the curriculum, instructional, and assessment strategies used to teach this course. To effectively transform organic chemistry education, it is necessary to forge effective professional development programs to equip instructors with literature-based resources and support that could meaningfully affect instructors' pedagogical knowledge and teaching practices (Gess-Newsome et al., 2003; Henderson et al., 2011; Talanquer, 2014). One critical direction for these initiatives is in the area of representational competence (RC), as learning and communicating with visualizations is an essential component of chemistry instruction (Kozma and Russell, 2005; Ainsworth, 2006; Gilbert, 2007; Stieff, 2007; Keehner et al., 2008).

The development of RC has been positioned as one of the chief goals for STEM education by the National Research Council (2012). Chemistry is one of the main STEM disciplines where learner success is significantly impacted by RC because many of the fundamental concepts and processes in chemistry cannot be directly observed or experienced in the physical world (Kozma and Russell, 2005; Ainsworth, 2006; Gilbert, 2007). Kozma and Russell (2005), p. 131 define RC as "a set of skills and practices that allow a person to reflectively use a variety of representations, singly and together, to think about, communicate, and act on chemical phenomena in terms of underlying aperceptual physical entities and processes." This set of skills includes the ability to interpret, translate, generate, and use representations, among others. At the same time, one can develop these foundational RC skills without understanding the 'why' behind engaging in tasks that require these skills. This is problematic because instruction and tasks that do not make sense to learners undermine their motivation to continue work in science (diSessa, 2004). Meta-RC "may be precisely what makes learning representations sensible to students" (diSessa, 2004). Meta-RC is a subset of RC that allows for the reflective and purposeful use of representations and includes skills such as the ability to describe affordances and limitations of various representations or select an optimal representation for a particular purpose. Therefore, effective support for developing RC requires a comprehensive approach that targets both the foundational RC skills and meta-RC skills.

Although a wide body of literature has focused on exploring organic chemistry students' learning about and with representations (Bodner and Domin, 2000; Cooper et al., 2010; Grove et al., 2012; Stull et al., 2012; Popova and Bretz, 2018a,b,c), less is known about organic chemistry instructors' approaches toward developing learner RC. Literature in this domain shows that even though conventional teaching approaches in chemistry incorporate a wide variety of representations, they are not frequently guided by learning objectives that explicitly target RC (Talanquer, 2022). It is often tacitly assumed that learners will develop RC without explicit instruction that intentionally scaffolds the development of the various RC skills. Recent studies from our research group found that neither chemistry instructors (Popova and Jones, 2021; Jones et al., 2022) nor chemistry textbooks (Gurung et al., 2022) support learners in developing higher-level meta-RC skills that allow for the reflective and purposeful use of representations. Moreover, currently, no professional development opportunities exist focused on improving how instructors support learner RC, despite chemistry instructors reporting wanting to learn about (a) finding quality representations to use in their teaching (e.g., animations, simulations), (b) effective teaching about representations, (c) proper assessment of student mastery of representations, and (d) expert-novice differences in understanding representations (Popova and Jones, 2021). The lack of such domain-specific professional development is problematic, as learners with developed RC are better set for building conceptual understanding and acquiring scientific practices (Lansangan and Orleans, 2007; Sim and Daniel, 2014; Lansangan et al., 2018; Dickmann et al., 2019; Herunata et al., 2021).

4 Hypothetical five-year RPP between organic chemistry instructors and education researchers

RPPs could be established to fill this gap in domain-specific professional development. Here, I present an outline for a potential five-year RPP between organic chemistry instructors and education researchers, aiming to improve organic chemistry students' RC. Importantly, should RPP work be externally funded, it is imperative to ensure that appropriate funds are allocated to compensate for the work of not only the researchers but also the practitioners. Securing external funding is crucial not only for the operational needs of the RPP but also to honor the significant contributions of each participant. Appropriately allocated funds will ensure that instructors are compensated for their time, particularly for tasks that extend beyond their usual teaching responsibilities. Proper compensation would acknowledge instructors' essential role in the partnership.

The RPP work could be partitioned into six main stages (Table 1) following the application process to express an interest in participating. The application process for the RPP should be intentionally designed

TABLE 1 Timeline and the main stages for the proposed five-year research-practice partnership.

Year	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
0										⏱ application process		
1	Teaching pre-RPP						✹			Stage 1		
2			Stage 2								✹	
3	Stage 3				Stage 4		✹					
4	Stage 5				Stage 6						✹	
5	Teaching post-RPP						✹		⏲			
Legend ⏱ start ✹ f2f meeting ⏲ finish	Stage 1: Discussion of readings to generate ideas for RC interventions based on research and theory			Stages 2 and 4: Design, development, and refinement of RC interventions; CITI, IRB, and plans for collecting and analyzing data			Stages 3 and 5: Administering interventions in Organic Chemistry I and collecting student outcomes data			Stage 6: Dissemination via publications, including journal author guidelines, organization of a manuscript, etc.		

to select instructors who demonstrate a clear commitment to engage in a long-term collaboration. The teaching pre-RPP and post-RPP periods in Years 1 and 5 (Table 1) could be used for collecting various data (e.g., video observations, course artifacts, interviews) to evaluate the impact of RPP on the Pedagogical Content Knowledge (PCK) and practices of instructors over time. Recognizing that RPPs generally involve establishing an external community of like-minded colleagues outside of one's institution, ideally, the RPP work should start at an in-person gathering to help the RPP members more effectively build rapport with each other. Additionally, the RPP members should have an opportunity to come together at least yearly (e.g., at conferences). At the same time, consideration should be made for allowing for hybrid attendance of the in-person meetings, recognizing potential diverse needs and constraints of RPP members (e.g., funding limitations, teaching commitments, and personal health and well-being). Outside the occasional in-person meetings, the RPP members should meet periodically during remote meetings (below, I propose a monthly schedule of meetings).

The RPP work could start with presenting the instructors with the "representational dilemma." The representational dilemma refers to the idea that learners are expected to understand abstract concepts and reaction mechanisms using various representations before they have fully learned how to interpret and use these representations. This is analogous to using a complex map to navigate a new city without the necessary practice in map reading. Just as learning to read a map requires practice, learners also need support in using representations to understand new concepts and mechanisms. This dilemma could be used to outline a very general and broad goal for the RPP—to collectively brainstorm and generate practical solutions for this challenge, which could include designing and testing interventions in authentic classroom settings. At this first meeting, it is also critical to jointly discuss and negotiate roles, routines, and protocols for engagement as an RPP (Stage 1).

At the subsequent monthly remote meetings, RPP members should discuss literature related to theory and research around supporting learner RC (Stage 1). This literature could include (1) primers on RC and meta-RC (diSessa, 2004; Kozma and Russell, 2005), (2) relevant theories of learning including constructivism (Novak, 1993), dual-coding theory (Paivio, 1986), and the information processing model (Baddeley, 2003), (3) Johnstone's chemistry triplet (Johnstone, 2006; Taber, 2013), (4) multimedia and visualization principles (Mayer, 2002; Tasker, 2015), (5) the key factors affecting learners' ability to reason with representation (Schönborn and Anderson, 2006, 2010), (6) assessment including the use of representations to promote equitable assessment design (Towns, 2014; Ralph and Lewis, 2020); and (7) design frameworks for developing educational interventions (Branch, 2010; Rau, 2017). At each remote monthly meeting, the discussion should focus on deriving insights for interventions aimed at improving RC.

Next, the RPP should leverage their knowledge of the (1) needs of their learners, (2) relevant research and theory, and (3) design frameworks to develop interventions to support learner RC (Stage 2). RPP can work collaboratively to review and refine the developed interventions (including student and instructor implementation guides). At this stage, the RPP should also make concrete plans for gathering and analyzing any necessary student data, including, when applicable, completing the Collaborative Institutional Training Initiative (CITI) and submitting Institutional Review Board (IRB) applications to ensure the ethical administration, testing, and evaluation of developed interventions.

The next stage would involve administering interventions in Organic Chemistry I and collecting student outcomes data (Stage 3). Each RPP member administering interventions in their classroom should be able to select a set of newly developed interventions that are best aligned with their learning objectives and the needs of their students. During this stage, RPP members should collect, analyze, and reflect on student data to evaluate the success of the administered interventions. In Stages 4 and 5, RPP members refine, develop (if needed), and readminister interventions in organic chemistry classrooms, emulating Stages 2 and 3. Finally, the RPP members should work on disseminating their work at conferences and via publications (Stage 6). Time should be set aside for planning, writing, and revising publications at the end of the project.

To mitigate the potential burden on instructors, the RPP should be structured to foster a collaborative environment where responsibilities are shared among all partners. This should include substantial support from education researchers in managing the collection and analysis of data, and the administrative tasks associated with IRB approval and funding applications. Leveraging the diverse expertise within the RPP could help enhance the sustainability and effectiveness of the interventions without overburdening any of the members.

5 Discussion, implications, and concluding remarks

While the section above outlines the timeline and main stages for a hypothetical RPP, it is important to tailor this plan to the unique context and needs of each specific project or collective of RPP participants. For example, while Table 1 proposes that in-person meetings occur in March and July, to coincide with the National Meeting and Exposition of the American Chemical Society (ACS) and the Biennial Conference on Chemical Education (BCCE), RPPs based outside of the United States should consider holding in-person meetings in accordance with the timelines of local chemistry education conferences.

Additionally, while this perspective article presents a hypothetical RPP for supporting learner RC, RPPs could be leveraged to address other important problems of organic chemistry education, such as developing student mechanistic reasoning in lecture courses or scientific practices in laboratory courses—other areas identified as particularly difficult for both teaching and learning (National Research Council, 2012; Graulich, 2015).

Finally, research on RPPs in education is very sparse, especially at the postsecondary level. More research is needed to understand the influences of RPPs on instructors and researchers (Coburn and Penuel, 2016). For example, there are several studies that provide evidence that RPP participation is associated with increased access to research (Bickel and Cooley, 1985; Kerr et al., 2006), but little is known about whether it is also associated with greater use of research for making instructional decisions (Coburn and Penuel, 2016). Comprehensive evaluations are necessary to assess RPP contributions to educational improvement. For example, studies are needed to characterize and evaluate (a) the dynamics of RPPs and the mechanisms by which they may foster educational improvement, (b) challenges associated with RPPs at the postsecondary level, (c) how RPP participation might influence instructors' PCK and practices and researchers' insights, (d) whether RPPs foster greater capacity and uptake of research in classroom decision-making, and (e) the extent to which RPPs help improve student learning and RC. By diving deeper into these aspects, we can pave the way for RPPs to not just be a theoretical ideal but a practical vehicle for driving improvements in the quality of chemistry education.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

MP: Writing – original draft, Writing – review & editing.

Funding

The author declares that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ainsworth, S. (2006). DeFT: a conceptual framework for considering learning with multiple representations. *Learn. Instr.* 16, 183–198. doi: 10.1016/j.learninstruc.2006.03.001

Baddeley, A. (2003). Working memory and language: an overview. *J. Commun. Dis.* 36, 189–208. doi: 10.1016/S0021-9924(03)00019-4

Barab, S. A., Gresalfi, M. S., and Ingram-Goble, A. (2010). Transformational play: using games to position person, content, and context. *Educ. Res.* 39, 525–536. doi: 10.3102/0013189X10386593

Bickel, W. E., and Cooley, W. W. (1985). Decision-oriented educational research in school districts: the role of dissemination processes. *Stud. Educ. Eval.* 11, 183–203. doi: 10.1016/0191-491X(85)90022-7

Bodner, G. M., and Domin, D. S. (2000). Mental models: the role of representations in problem solving in chemistry. *Univ. Chem. Educ.* 4, 24–30,

Booth, J. L., Cooper, L. A., Donovan, M. S., Huyghe, A., Koedinger, K., and Pare-Blagoev, E. J. (2015). Design-based research within the constraints of practice: algebra by example. *J. Educ. Stud. Placed Risk* 20, 79–100. doi: 10.1080/10824669.2014.986674

Branch, R. M. (2010). *Instructional design: the ADDIE approach*. Athens, GA: Springer.

Burbank, M. D., and Kauchak, D. (2003). An alternative model for professional development: investigations into effective collaboration. *Teach. Teach. Educ.* 19, 499–514. doi: 10.1016/S0742-051X(03)00048-9

Chicoine, D. (2004). Ignoring the obvious: a constructivist critique of a traditional teacher education program. *Educ. Stud.* 36, 245–263. doi: 10.1207/s15326993es3603_4

Coburn, C. E., and Penuel, W. R. (2016). Research–practice partnerships in education: outcomes, dynamics, and open questions, educational researcher. *Educ. Res.* 45, 48–54. doi: 10.3102/0013189X16631750

Coburn, C. E., Penuel, W. R., and Geil, K. (2013). *Research–practice partnerships at the district level: A new strategy for leveraging research for educational improvement*. Berkeley, CA and Boulder, CO: University of California and University of Colorado.

Coburn, C. E., and Stein, M. K. (2010). *Research and practice in education: Building alliances, bridging the divide*. Lanham, MD: Rowman & Littlefield.

Cooper, M. M., Grove, N., Underwood, S. M., and Klymkowsky, M. W. (2010). Lost in Lewis structures: an investigation of student difficulties in developing representational competence. *J. Chem. Educ.* 87, 869–874. doi: 10.1021/ed900004y

Dickmann, T., Opfermann, M., Dammann, E., Lang, M., and Rumann, S. (2019). What you see is what you learn? The role of visual model comprehension for academic success in chemistry. *Chem. Educ. Res. Pr.* 20, 804–820. doi: 10.1039/C9RP00016J

diSessa, A. A. (2004). Metarepresentation: native competence and targets for instruction. *Cogn. Instr.* 22, 293–331. doi: 10.1207/s1532690xci2203_2

Donovan, M. S., Snow, C. E., and Daro, P. (2013). "The SERP approach to problem-solving research, development, and implementation" in *Design-based implementation research. National Society for the study of education yearbook*. eds. B. J. Fishman, W. R. Penuel, A.-R. Allen and B. H. Cheng (New York, NY: Teachers College Record), 400–425.

Fairweather, J. (2008). Linking evidence and promising practices in STEM undergraduate education. NRC Workshop on Evidence on Selected Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education. Available at:http://www7.nationalacademies.org/bose/Fairweather_CommissionedPaper.

Fishman, B. J., Marx, R. W., Best, S., and Tal, R. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teach. Teach. Educ.* 19, 643–658. doi: 10.1016/S0742-051X(03)00059-3

Fishman, B. J., Penuel, W. R., Allen, A.-R., and Cheng, B. H. (2013). *Design-based implementation research: Theories, methods, and exemplars. National Society for the study of education yearbook*. New York, NY: Teachers College Press.

Fullan, M., and Stiegelbauer, S. (1991). *The new meaning of educational change*. New York: Teachers College Press.

Geier, R., Blumenfeld, P., Marx, R. W., Krajcik, J., Fishman, B. J., and Soloway, E. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *J. Res. Sci. Teach.* 45, 922–939. doi: 10.1002/tea.20248

Gess-Newsome, J., Southerland, S. A., Johnston, A., and Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: the anatomy of change in college science teaching. *Am. Educ. Res. J.* 40, 731–767. doi: 10.3102/00028312040003731

Gilbert, J. K. (2007). "Visualization: a metacognitive skill in science and science education" in *Visualization in science education*. ed. J. K. Gilbert (Dordrecht: Springer), 9–27.

Graulich, N. (2015). The tip of the iceberg in organic chemistry classes: how do students deal with the invisible? *Chem. Educ. Res. Pr.* 16, 9–21. doi: 10.1039/C4RP00165F

Grove, N. P., Cooper, M. M., and Rush, K. M. (2012). Decorating with arrows: toward the development of representational competence in organic chemistry. *J. Chem. Educ.* 89, 844–849. doi: 10.1021/ed2003934

Grove, N. P., Hershberger, J. W., and Bretz, S. L. (2008). Impact of a spiral organic curriculum on student attrition and learning. *Chem. Educ. Res. Pr.* 9, 157–162. doi: 10.1039/B806232N

Gurung, E., Jacob, R., Bunch, Z., Thompson, B., and Popova, M. (2022). Evaluating the effectiveness of organic chemistry textbooks for promoting representational competence. *J. Chem. Educ.* 99, 2044–2054. doi: 10.1021/acs.jchemed.1c01054

Henderson, C., Beach, A., and Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature. *J. Res. Sci. Teach.* 48, 952–984. doi: 10.1002/tea.20439

Herunata, H., Rosyida, I., Sulistina, O., and Wijaya, H. W. (2021). Correlational analysis of conceptual understanding, chemical representation, and representational competence on chemistry equilibrium. *AIP Conf. Proc.* 2330:020016. doi: 10.1063/5.0043116

Johnson, J. (1998). Embracing change: a leadership model for the learning organization. *Int. J. Train. Dev.* 2, 141–150. doi: 10.1111/1468-2419.00042

Johnson, R. (2022). Bridging the divide between chemistry educators and chemistry education researchers. *J. Chem. Educ.* 99, 3631–3632. doi: 10.1021/acs.jchemed.2c01035

Johnstone, A. H. (2006). Chemical education research in Glasgow in perspective. *Chem. Educ. Res. Pract.* 7, 49–63. doi: 10.1039/B5RP90021B

Jones, K. B., and Gellene, G. I. (2005). Understanding attrition in an introductory chemistry sequence following successful completion of a remedial course. *J. Chem. Educ.* 82, 1241–1245. doi: 10.1021/ed082p1241

Jones, T., Romanov, A., Pratt, J. M., and Popova, M. (2022). Multi-framework case study characterizing organic chemistry instructors' approaches toward teaching about representations. *Chem. Educ. Res. Pract.* 23, 930–947. doi: 10.1039/D2RP00173J

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., and Montello, D. R. (2008). Spatial reasoning with external visualizations: what matters is what you see, not whether you interact. *Cogn. Sci.* 32, 1099–1132. doi: 10.1080/03640210801898177

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., and Barney, H. (2006). Strategies to promote data use for instructional improvement: actions, outcomes, and lessons from three urban districts. *Am. J. Educ.* 112, 496–520. doi: 10.1086/505057

Kozma, R., and Russell, J. (2005). "Students becoming chemists: developing representational competence" in *Visualization in science education*. ed. J. K. Gilbert (Dordrecht: Springer), 121–145.

Lansangan, R., and Orleans, A. (2007). Representational competence of secondary chemistry students in understanding selected chemical principles. *Int. J. Sci. Educ.* 24, 7930–7934,

Lansangan, R. V., Orleans, A. V., and Camacho, V. M. I. (2018). Assessing conceptual understanding in chemistry using representation. *Adv. Sci. Lett.* 24, 7930–7934. doi: 10.1166/asl.2018.12459

Lovecchio, K., and Dundes, L. (2002). Premed survival: understanding the culling process in premedical undergraduate education. *Acad. Med.* 77, 719–724. doi: 10.1097/00001888-200207000-00016

Lovitt, C., and Clarke, D. (1988). *The mathematics curriculum and teaching program: professional development package*. Australia: Curriculum Development Centre.

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: an example of the two-way street between cognition and instruction. *New Dir. Teach. Learn.* 2002, 55–71. doi: 10.1002/tl.47

National Research Council (1999). *Improving student learning: a strategic plan for education research and its utilization*. Washington, DC: National Academies Press.

National Research Council (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.

Novak, J. D. (1993). Human constructivism: a unification of psychological and epistemological phenomena in meaning-making. *Int. J. Pers. Constr. Psychol.* 6, 167–193. doi: 10.1080/08936039308404338

Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford: Oxford University Press.

Penuel, W. R., Allen, A., Coburn, C. E., and Farrell, C. (2015). Conceptualizing research–practice partnerships as joint work at boundaries. *J. Educ. Stud. Placed Risk.* 20, 182–197. doi: 10.1080/10824669.2014.988334

Popova, M., and Bretz, S. L. (2018a). It's only the major product that we care about in organic chemistry: an analysis of students' annotations of reaction coordinate diagrams. *J. Chem. Educ.* 95, 1086–1093. doi: 10.1021/acs.jchemed.8b00153

Popova, M., and Bretz, S. L. (2018b). Organic chemistry students' challenges with coherence formation between reactions and reaction coordinate diagrams. *Chem. Educ. Res. Pract.* 19, 732–745. doi: 10.1039/C8RP00064F

Popova, M., and Bretz, S. L. (2018c). Organic chemistry students' interpretations of the surface features of reaction coordinate diagrams. *Chem. Educ. Res. Pract.* 19, 919–931. doi: 10.1039/C8RP00063H

Popova, M., and Jones, T. (2021). Chemistry instructors' intentions toward developing, teaching, and assessing student representational competence skills. *Chem. Educ. Res. Pract.* 22, 733–748. doi: 10.1039/D0RP00329H

Ralph, V. R., and Lewis, S. E. (2020). Impact of representations in assessments on student performance and equity. *J. Chem. Educ.* 97, 603–615. doi: 10.1021/acs.jchemed.9b01058

Rau, M. A. (2017). A framework for educational technologies that support representational competencies. *IEEE Trans. Learn. Technol.* 10, 290–305. doi: 10.1109/TLT.2016.2623303

Rodriguez, J. G., and Towns, M. H. (2019). Alternative use for the refined consensus model of pedagogical content knowledge: suggestions for contextualizing chemistry education research. *J. Chem. Educ.* 96, 1797–1803. doi: 10.1021/acs.jchemed.9b00415

Schönborn, K. J., and Anderson, T. R. (2006). The importance of visual literacy in the education of biochemists. *Biochem. Mol. Biol. Educ.* 34, 94–102. doi: 10.1002/bmb.2006.49403402094

Schönborn, K. J., and Anderson, T. R. (2010). Bridging the educational research-teaching practice gap: foundations for assessing and developing biochemistry students' visual literacy. *Biochem. Mol. Biol. Educ.* 38, 347–354. doi: 10.1002/bmb.20436

Shadle, S. E., Marker, A., and Earl, B. (2017). Faculty drivers and barriers: laying the groundwork for undergraduate STEM education reform in academic departments. *Int. J. STEM Educ.* 4:8. doi: 10.1186/s40594-017-0062-7

Sim, J. H., and Daniel, E. G. S. (2014). Representational competence in chemistry: a comparison between students with different levels of understanding of basic chemical concepts and chemical representations. *Cogent Educ.* 1:991180. doi: 10.1080/2331186X.2014.991180

Snow, C. E., Lawrence, J., and White, C. (2009). Generating knowledge of academic language among urban middle school students. *J. Res. Educ. Effect.* 2, 325–344. doi: 10.1080/19345740903167042

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learn. Instr.* 17, 219–234. doi: 10.1016/j.learninstruc.2007.01.012

Stull, A. T., Hegarty, M., Dixon, B., and Stieff, M. (2012). Representational translation with concrete models in organic chemistry. *Cogn. Instr.* 30, 404–434. doi: 10.1080/07370008.2012.719956

Taber, K. S. (2013). Revisiting the chemistry triplet: drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chem. Educ. Res. Pract.* 14, 156–168. doi: 10.1039/C3RP00012E

Talanquer, V. (2014). DBER and STEM education reform: are we up to the challenge? *J. Res. Sci. Teach.* 51, 809–819. doi: 10.1002/tea.21162

Talanquer, V. (2022). The complexity of reasoning about and with chemical representations. *JACS Au* 2, 2658–2669. doi: 10.1021/jacsau.2c00498

Tasker, R. (2015). "Using multimedia to visualize the molecular world: educational theory into practice" in *Chemists' guide to effective teaching*. eds. N. J. Pienta, M. M. Cooper and T. J. Greenbowe (Hoboken: Pearson Prentice Hall).

Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. *J. Chem. Educ.* 91, 1426–1431. doi: 10.1021/ed500076x

Tseng, V. (2012). *Partnerships: shifting the dynamics between research and practice*. New York, NY: William T. Grant Foundation.

Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Rev. Educ. Res.* 79, 702–739. doi: 10.3102/0034654308330970

Yarnall, L., Shechtman, N., and Penuel, W. R. (2006). Using handheld computers to support improved classroom assessment in science: results from a field trial. *J. Sci. Educ. Technol.* 15, 142–158. doi: 10.1007/s10956-006-9008-4

frontiers | Frontiers in Education

TYPE Curriculum, Instruction, and Pedagogy
PUBLISHED 16 May 2024
DOI 10.3389/feduc.2024.1379886

OPEN ACCESS

EDITED BY
Jay Wackerly,
Central College, United States

REVIEWED BY
Grace Ferris,
Lesley University, United States
Melonie Teichert,
United States Naval Academy, United States

*CORRESPONDENCE
Theresa Gaines
✉ tgaines@haverford.edu

†These authors have contributed equally to this work

RECEIVED 31 January 2024
ACCEPTED 29 April 2024
PUBLISHED 16 May 2024

CITATION
Gaines T and Burrows NL (2024) Exploring alternative assessments during COVID: Instructor experiences using oral exams. *Front. Educ.* 9:1379886.
doi: 10.3389/feduc.2024.1379886

Exploring alternative assessments during COVID: Instructor experiences using oral exams

Theresa Gaines[1*†] and Nikita Lauren Burrows[2†]

[1]Department of Chemistry, Haverford College, Haverford, PA, United States, [2]Department of Chemistry and Physics, Monmouth University, West Long Branch, NJ, United States

Research into conventional summative assessment methods, such as written exams, has been extensively documented in the literature. However, as academia evolves in response to changing dynamics, there is a growing demand for more robust summative assessment approaches. Oral exams have emerged as a noteworthy form of summative assessment with intrinsic formative qualities, enabling instructors to delve deeply into students' comprehension within a meaningful learning framework. Considering the constraints imposed by the traditional written examination format during the COVID pandemic, two educators implemented oral assessments in their chemistry courses, one in general chemistry and the other in organic chemistry. This article presents a comprehensive account of their approach, course structure, rubrics, documentation procedures, and the challenges associated with implementing oral exams. Furthermore, the authors offer insights derived from perceived course outcomes, experiences, collaborative efforts, and reflections from this transformative process. Through candid exploration, this article delves into both the potential advantages and the hurdles associated with the adoption of oral exams in chemistry education. It serves as a valuable resource for educators seeking innovative assessment strategies.

KEYWORDS

oral exam, alternative assessments, organic chemistry, general chemistry, instructor experience

1 Introduction

The use of oral exams as an alternative assessment method in chemistry education has gained some attention in recent years (Ramella, 2019). Traditionally, chemistry courses rely on written examinations to evaluate students' understanding of concepts, problem-solving abilities, and growing expertise with course material; however, it has become increasingly evident that written exams alone may not fully capture the breadth and depth of students' knowledge and skills in chemistry (Tienson-Tseng, 2019). Oral exams have been used as a means to provide a more comprehensive evaluation of students' learning for a century (Muldoon, 1926), and have the potential to offer unique advantages, such as the opportunity for direct interaction between the examiner and student, the ability to assess critical thinking and communication skills, and the potential to create a more engaging and inclusive assessment environment. Herein, we delve into the use of oral exams in chemistry education, their benefits, challenges, and potential impacts on student learning outcomes. By exploring the practical aspects, pedagogical implications, and empirical evidence surrounding the use of oral exams, we aim to shed light on the significance of this alternative assessment method as a means to leverage research to inform practice.

1.1 Background and rationale

While oral examinations are not a novel method of assessing students (Muldoon, 1926), their integration into larger courses (twenty or more students) is not commonly practiced (Lubarda et al., 2021); most learners encounter formal oral evaluations in graduate school, professional school or in a job interview for the first time (Dicks et al., 2012). Recent discussions in the American Chemical Society (ACS) symposium series have highlighted the benefits of oral examinations, emphasizing their ability to thoroughly examine understanding and facilitate meaningful student-faculty discussions through leading questions (Ramella, 2019). Literature suggests these assessments have the potential to offer valuable insights into students' problem-solving skills and information processing abilities (Bowen, 1994; Orgill and Sutherland, 2008; Walker and Sampson, 2013). Students are often more prepared for oral discussions due to the face-to-face nature of the examination, which motivates them to engage with the material at a deeper level—the assessment requires students do more than just memorize and regurgitate—resulting in better responses (Hambrecht, 2003; Boedigheimer et al., 2015; Iannone and Simpson, 2015). Similarly, instructors are able to engage with students and guide them in their critical thinking on a deep level in a way that written exams cannot facilitate (Gent et al., 1999).

Although there are only a few documented examples of oral assessments in the chemical education literature (e.g., Muldoon, 1926; Roecker, 2007; Dicks et al., 2012; Crawford and Kloepper, 2019), existing reports indicate positive student experiences with this evaluation method in both inorganic and organic chemistry classes (Roecker, 2007; Crawford and Kloepper, 2019). Burrows et al. (2021) found that students had a positive overall experience with oral assessments in laboratories, which is consistent with previous research on oral assessments in lecture courses (Roecker, 2007). Students specifically felt a sense of accountability for their preparation during lab interviews, which motivated them to avoid exposing gaps in their knowledge. This individual accountability led to active participation and knowledge construction during the lab session (Burrows et al., 2021).

When teaching, we encourage students to think in different ways with different tools. We think with our bodies and physical models (Gilbert and Treagust, 2009; Flood et al., 2015; Kiste et al., 2016); we think with the physical spaces we occupy and the digital spaces we can manipulate (Kozma and Russell, 2005). We also think by utilizing the people around us as tools. As educators, we encourage students to think with their peers collaboratively, and as they continue to grow into chemists, teach them to think in conjunction with the broader community to tackle big problems by writing, sharing, and discussing our work to build our collective knowledge—shared attention, in collaboration broadly (Gregory and Jackson, 2017), or in the same physical space (Shteynberg and Apfelbaum, 2013) increases our capacity to learn.

When thinking by utilizing the people around us, students learn to think by leveraging experts around them as tools (Paul, 2021). Novices can build expertise by gaining insights into the internal thought processes of experts (Collins et al., 1991); additionally, these expert-novice interactions are places where learners and educators can make the invisible parts of learning visible (Paul, 2021), and figure out next steps to take to support the learner's growing expertise—which is in line with constructivist teaching practices (Ausubel et al., 1978; Bodner, 1986) and not dissimilar to cognitive apprenticeships (Collins et al., 1991). Oral exams are a way to intentionally allocate time to facilitate these novice-expert interactions that use a variety of modes of thinking.

1.2 Concerns about oral exams

Two concerns are frequently cited in education literature regarding oral exams: (1) student anxiety and (2) fairness (Iannone and Simpson, 2015). In Huxham et al.'s (2012) study, focus group testimonies revealed some students felt anxious because they recognized that they needed a deep understanding to explain concepts to their examiner in a way that was unique to oral exams. A similar sentiment was expressed by students who felt exposed during the oral assessments; this was also expressed by students in other studies (Iannone and Simpson, 2015). Within social work education literature, there's some evidence that increasing a students' experiences and familiarity with this type of assessment can lessen some of the anxiety they experience on subsequent oral assessments (Henderson et al., 2002; Huxham et al., 2012; Iannone and Simpson, 2015). This is especially noteworthy for STEM fields as most students do not encounter oral assessments in their undergraduate careers (Goodman, 2020). Despite how anxiety-inducing oral exams might be (Díaz et al., 2001; Huxham et al., 2012; Iannone and Simpson, 2015), literature suggests that some anxious learners still prefer oral assessments over traditional, written exams (Huxham et al., 2012).

Unlike written exams where students gather together in a room to take it, the one-on-one nature of oral exams gives opportunity for bias or prejudice to occur in a way that isn't there for regular exams (Heyneman et al., 2008). Oral examinations are not anonymous in the way a written exam can be to reduce bias. There are concerns that this leaves oral exams especially prone to bias and unfairness (Davis and Karunathilake, 2005). But as Iannone and Simpson (2015)express in their publication, bias is not always eliminated in written assessments, and it will be hard to compare bias remaining in each assessment type.

Due to the conversational nature of an oral exam, each conversation an examiner has with a student will be different—even if they are discussing the same questions. The examiner explores student knowledge so gaps in knowledge can be noted; however, gaps in knowledge don't hinder a student from demonstrating other knowledge. In an oral session, a student can move through, or around, stumbling blocks in the discussion to demonstrate other knowledge and insight in a way that's not possible on a written exam. In these scenarios, one student may receive a guiding question in one session that another student doesn't in theirs, but their overall outcomes might be the same. Some students have reported feeling this is an issue of fairness, but others saw this as a mode to get targeted feedback and correct course in progress (Iannone and Simpson, 2015).

Oral exams might also provide another mode of equitable assessment as a learning tool. For learners who are studying in languages that aren't their native language, an oral assessment may allow them to use their strongest mode of communication (speech) to their advantage in a way they might not be able to on a timed, written exam (Huxham et al., 2012; Ramella, 2019). Oral exams might offer more equitable modes of communicating and assessing knowledge for students who experience dyslexia, dyscalculia or dysgraphia (Waterfield and West, 2005; Huxham et al., 2012); however, there are other factors that must be considered when conducting oral exams such as the types of power dynamics and spaces are being used to facilitate the assessment (Theobold, 2021).

In a study with biology students, students who took the oral exam scored significantly higher on their assessment when compared to students who took the written assessment (Huxham et al., 2012). Students expressed that there was an authenticity to the format that made them feel like a professional (Joughin, 1998). While students take many written examinations in their undergraduate careers, once they move onto their professional lives, they will be expected to communicate in both written and oral formats to convey their ideas. Oral exams might be an important key to helping them build an aesthetic (fitting) and authentic professional identity.

Oral examinations provide an opportunity to rigorously probe understanding as leading questions can be asked that facilitate novice-expert or student-faculty discussions and interactions. The benefits of student–faculty interaction are numerous, including objective increases in grade point averages and in matriculation to post-graduate studies (Cotten and Wilson, 2006). Dr. Nikita Burrows and Dr. Theresa Gaines implemented oral exams in their chemistry courses, general chemistry, and organic chemistry, respectively, in the 2020–2021 academic year. In this work we will discuss our experiences and how we designed our courses to facilitate oral assessments in chemistry classrooms as evidence-based practices.

For more advice on implementing oral exams, we recommend looking at *A Short Guide to Oral Assessments* (Joughin, 2010) and *Oral Exams: A More Meaningful Assessment of Students' Understanding* (Theobold, 2021).

2 Pedagogical frameworks

The primary theory guiding our educational practice is the theory of constructivism—a theory that describes how we gain knowledge. Constructivism purports that we actively piece together knowledge from our experiences (Bodner, 1986)—and that learning is meaningful when we integrate the new knowledge into our previously assembled knowledge structures (Ausubel et al., 1978).

Ausubel et al. (1978) and Bodner (1986) both highlight the advantages of probing student knowledge to subsequently guide learners in the next steps. Knowledge isn't transferred from educator to learner as a complete and organized set, the learners themselves build their knowledge in their minds as they have new experiences and are introduced to new information; their meaningful learning is dependent on their ability to put the pieces together and order them in a way that makes sense with the world around them (Ausubel et al., 1978; Bodner, 1986).

As educators we strive to orchestrate environments, tasks and scenarios that assist learners in building their own knowledge based on their current and prior learning experiences. In addition to constructing knowledge, we also hope that the classroom environments and activities we facilitate promote students to make rich, multi-layered connections between different pieces of knowledge they've constructed. While we can create assessments to probe if learner knowledge constructs map on to ours satisfactorily or not—both the assembly process and the whole interconnected knowledge construct are invisible (Collins et al., 1991). Engaging with learners regularly and deeply in dialog about their thought processes and how they reason when problem-solving might afford space to encourage learners to find places where gaps in knowledge can be filled, or important connections within their knowledge web can be made.

In previously published studies where students shared about their oral exam experiences, the idea of the assessment feeling authentic emerged as a theme (Joughin, 1998). Kharkhurin (2014) developed four criteria for a work to be considered creative: (1) novelty, (2) utility, (3) aesthetic, and (4) authentic. While learners are (often) not yet using chemistry as tool to iterate and generate new knowledge—flexing their expertise and creativity—they are building their expertise in chemistry and identity as a chemist by exploring these dimensions independently through their courses, assessments and practice. Kharkhurin (2014) defines authenticity as honesty to the process and that is perceived in the output. Students recognized that having a conversation about their results was similar to how they might convey them as a professional chemist working somewhere else. It felt authentic.

Aesthetic deals with intentions and how those intentions are conveyed. When giving a presentation or interview, did you intend to appear as a chemist, were you perceived that way? As stated previously, neither written assessments nor scrutiny through timed, written assessments are not common beyond school; however, oral dialogs are extremely common—aesthetically chemistry. Oral exams may give an opportunity for students to explore their identities in an authentic and aesthetic manner as they continue to build their expertise and think creatively within their craft.

3 Learning environment

3.1 Course set-up

Due to the COVID-19 pandemic, both courses were online only. Dr. Burrows' general chemistry course was modeled after a flipped classroom, students were provided videos with which to engage before synchronous class time. The videos were hosted on Play-Posit, which prompted students with questions to answer as they progressed through them to probe understanding. Synchronous meeting time consisted of targeted mini-lectures and problem-solving sessions. This class had weekly quizzes, and three oral exams.

Dr. Gaines' organic chemistry course was asynchronous. Students were provided a series of video lectures and guided notes to fill in while watching the videos each week. In place of synchronous problem-solving sessions, Dr. Gaines hosted daily, elective, problem-solving sessions that students could attend synchronously if they wanted to; these sessions were recorded for students to access later. Students submitted weekly problem sets and had three oral exams. Dr. Gaines' class had thirty students in the fall semester and twenty in the spring semester where oral exams were implemented. Dr. Burrows had twenty students in both semesters.

3.2 Design and structure

Both instructors designed the oral exams collaboratively. Each exam consisted of five multi-step, scaffolded questions that covered specific learning outcomes. Questions 1 and 2 covered the same learning outcomes as each other. Questions 3 and 4 also covered the same learning outcomes as each other. Question 5 covered unique learning outcomes from the previous questions—giving three sets of learning outcomes that can be covered in the exam. These exams were initially given as an open-resource (textbook, internet resources, notes), take-home assessment. Students had a week to work them before they were asked to sit the oral exam.

The take-home portions of the exams were distributed through each instructor's learning management system (LMS). Dr. Burrows' institution used Desire2Learn (D2L) and Dr. Gaines' institution used Canvas. Students were required to upload a PDF copy of their worked exam to the LMS before they sat their exam. This PDF was not assessed; but was held for record-keeping purposes. This also afforded an opportunity for the instructor to prepare for the session by previewing how the student solved the problems. In Dr. Gaines' course, students were encouraged to collaborate on the take-home portion of the assessment. Students in Dr. Burrows' courses were not.

Students signed up for times they wanted to take their oral exam either via their LMS or via Calendly based on the thirty-minute timeslots the instructor created. The exams were hosted on Zoom during the scheduled time. Either the student could share their screen with the PDF of their exam, or the professor would share their screen showing the student's PDF. This helped to facilitate conversation and gave a method to draw if needed.

During the oral exam, discussion centered around three of the five questions, and students selected the first question they wanted to go over. They were prompted to identify which one they felt most confident about; the instructor would pick the other two questions for discussion. Due to the structure of the learning outcomes in each exam question, all sets of learning outcomes could still be assessed, no matter which question students elected to go through first. The instructor would pick two questions to discuss that did not match their initial choice. On average, students finished their oral exam within twenty minutes, where prepared students tended to finish faster than students who were less prepared.

3.3 Rubrics

Rubrics were assessed on eight categories for each question in the oral portion for the organic chemistry course:

1. Overall understanding of the content
2. Communication
3. Valid structures (in their PDF)
4. Argument
5. Evidence
6. Reasonable answers
7. Calculations (Burrows) or curved arrows (Gaines)
8. Prompting (mistakes)

The students were scored in each category as exemplary (5 points), competent (4 points), developing (3 points) or emerging (0–2 points). This gave a maximum of 40 points per question, and 120 points for the exam. In Dr. Gaines' course, this was modified after students expressed concern that the scale was too punishing. After discussing with the class, the following was implemented: exemplary (5 points), competent (4.25 points), developing (3.5 points) or emerging (0–2 points).

Another modification implemented in Dr. Gaines' course was a "Redo" system. Where students could choose to retake any single oral assessment during the semester to replace the initial grade. All aspects of the assessment were the same as the first attempt.

3.4 Keeping receipts/documentation

McCloud (2023) wrote a great piece on keeping evidence in an ungraded class; but, this advice can be applied more broadly when it comes to implementing alternative, non-traditional assessments. Grades are the currency by which students get internships, jobs, scholarships, placements in medical schools or graduate schools. Alternative assessments like oral exams may scare students; they may be anxious as to whether they can successfully collect the currency they want to leverage for later (McCloud, 2023).

For underrepresented or historically excluded faculty, this student anxiety and uncertainty may reflect negatively upon us in teaching evaluations; therefore, keeping evidence is necessary protective documentation when implementing non-traditional pedagogical practices as we may not always receive the benefit of the doubt in extenuating circumstances (McCloud, 2023).

Both instructors kept record of evidence by:

1. Using a standardized rubric for every oral exam that students had access to before the assessment opened (permanently posted in the LMS)
2. Assigning students to complete the assessment as a take-home assessment first (PDF uploaded to the LMS)
3. Requiring students to upload their finished take-home assessment as a PDF into the course LMS before sitting the oral portion
4. Explicitly permitting and encouraging students to refer to their PDF to guide the explanation in the oral exam (instructions in LMS)
5. Using the rubric as the scorecard and giving it to students as part of their feedback (also through the LMS).

For both instructors, it was imperative to use the LMS as much as possible as that created a record of what materials were made available, what emails were sent, and what appointments were created. It is an independent record that also tracks changes, which could be used in case of any dispute. All details regarding the design of the course and rubrics use can be found in the **Supplementary material**.

4 Results to date and assessment

4.1 Course outcomes

The information expressed in this section is anecdotal in the absence of IRB approval. Overall, in both general chemistry and organic chemistry, students performed well on the oral exams compared to traditional teaching methods pre-COVID. However, some students struggled in ways that could be unique to the oral exam format. Without further study, it is impossible to parse if these difficulties were due to the format, or if they were, in part, due to remote instruction during the pandemic. During the oral exams, students initially struggled with how to prepare effectively; students typically required considerable prompting to articulate mechanistic details of reactions, and this aspect saw slow improvement over the course of the year. Although many students were able to arrive at answers through conversation, some prompting was necessary to elicit the relevant information from them. Overall, while the oral exams yielded better course outcomes than the prior teaching methods, we would be willing to investigate how they boost understanding or how they magnify areas of struggle.

4.2 Academic integrity

During the oral exam, students are responsible for demonstrating their own knowledge by walking the instructor through how they solved the exam problems. In such instances, it is hard for a student to pretend to know something they don't understand when an expert is asking questions. There were no instances of academic dishonesty in either set of courses.

4.3 Future work and areas of interest

In the face of remote instruction due to COVID-19, the method of oral examination was appealing as it provided a structure to interact one-on-one with learners in our online chemistry courses. This format was implemented without the intention to study the impacts of oral exams on learners. Now, post-implementation, and post-reflection, we would like to discuss our plans to collect data when implementing these assessments again.

To contrast each cohort between traditional and oral assessments, we would collect scores on comparable assessments for both. This would be useful to support our observation that students performed well. In addition to grades on assessments, investigating how students prepare for their assessment would be valuable to understanding what is fueling any differences between oral and traditional performance. Similarly, probing into student anxiety pre- and post-assessment could be collected. The latter two items via survey. Lastly, if time permits, interviewing students to find out their experiences in a class that uses oral assessments and about their experiences during the assessment would be valuable to document.

5 Discussion on the practical implications, objectives and lessons learned

5.1 Professor experience—Nikita

Implementing oral exams in my chemistry courses was a rewarding experience that brought several advantages. Collaborating with another professor to troubleshoot and strategize created a supportive community of practice, where we could bounce ideas off each other and tackle any issues that arose. This collaborative effort eased the process of incorporating oral exams into the curriculum and reinforced our commitment to this assessment method. Moreover, the collaborative student environment fostered through the Zoom classroom was highly interactive and responsive. My familiarity with breakout rooms encouraged active engagement, and my students embraced the opportunity to work on oral exams independently, seeking my assistance during office hours when needed. The one-on-one interactions during the oral exams allowed for deeper insights into students' knowledge and comprehension levels. In the context of general chemistry, the assessment format compelled students to confront the concepts directly, as they couldn't rely solely on mathematical calculations to mask their understanding. It provided a comprehensive evaluation of their communication abilities across various domains.

The formative and summative aspects of the oral exams complemented each other, creating a holistic assessment approach. Although the exams were summative in nature, the immediate feedback provided during the oral sessions was invaluable to students. Unlike traditional written exams where feedback might go unread, students appreciated understanding why points were deducted and how they could improve. This enhanced their learning experience and motivated them to actively engage with their feedback. While the time investment in conducting oral exams was significant, the payoff was substantial. Students demonstrated a deeper grasp of the material, and some even reached out to me after moving to other schools, seeking my guidance based on their positive oral exam experience.

As for the continuation of this assessment method, I regrettably had to discontinue it for my multi-section classes. The consistency and fairness in assessments across sections demanded uniformity, making it impractical to implement oral exams in this context. However, I firmly believe that for smaller classes or online formats, oral exams can continue to be an effective means of assessing student comprehension and enhancing their learning journey. The interactive and personalized nature of oral exams not only provides valuable insights into student understanding but also instills a sense of accountability and responsibility for their own learning. As educators, we should continuously explore and adapt assessment approaches to best serve our students' academic and intellectual growth.

5.2 Professor experience—Theresa

In my experience, I would be willing to try oral exams again. I think the struggle of overcoming logistical issues is worth the growth and conversations that oral exams afford. My organic classes at the time were online and asynchronous. The oral exam format provided a synchronous, face-to-face candid experience with each of my students that my traditional, in-person classrooms would still benefit from. Increasingly, students in my classrooms value collaboration. For them and their peers, willingness to help others who might not understand the material feels integral to their integrity and their character. Most instances of academic dishonesty that I've experienced in my classrooms stems from collaboration as a core value. My solution, through oral exams, was to give space for that collaboration.

In the oral exam setting, students had a chance to show what they've learned, but it was not at the cost of preventing collaboration with others if that's what they wanted to do. For students, this process illuminated if their collaborative efforts were lacking or counter-productive. In study groups, it is easy to hide behind someone else's explanation, but in the oral exam, each student leads the conversation. Several of my students realized through this experience that they weren't leaving their collaborative sessions with anything but answers; that wasn't sufficient for our oral exam which asked them to re-trace the process of deducing an answer. In written exams, it is difficult to give someone the benefit of the doubt about how much they understand if the answer is ambiguous. In oral exams, we can explore to find if and where a gap in knowledge exists and address it. It was a convenient place to apply just-in-time teaching—as there would be future assessments that our conversation would be applicable to.

As the instructor, it was beneficial to have Dr. Burrows with me. When I was stuck or if I ran into roadblocks, it was nice to talk to her and discuss if there was a bigger issue in our implementation or if there was something specific to my class and my students. Currently, I am not using oral exams. Initially when we went back to in-person classes, figuring out when and how to schedule the oral exams was an issue that I couldn't resolve before the semester started. Now I'm at a new institution and I am teaching a brand-new set of courses. I would prefer to become more familiar with these courses before changing how the exams are conducted.

5.3 Perceived student experiences

In the initial stages of the oral assessment implementation, students exhibited apprehension and nervousness, as many had never encountered such a form of evaluation before. However, as they experienced the process firsthand, their perceptions evolved positively. Students found that the oral assessment was not as intimidating as they initially thought and realized that it provided a unique opportunity to showcase their understanding of the subject matter. This phenomenon was previously described in the literature (Burrows et al., 2021). The shift in students' questions during the assessment was notable, moving from concerns about whether they had done the task correctly to inquiries about their conceptual comprehension and problem-solving approach. Students also felt comforted in their ability to choose the question they were most confident in as the first question we discussed. This change in focus indicated a deeper engagement with the material and a desire to demonstrate their true understanding rather than regurgitation of facts. To mitigate some of the initial student apprehension, several strategies were employed, such as offering the exam as a take-home version initially to acclimatize students to the format, allowing for redos to encourage learning from mistakes, and implementing policies such as dropping the lowest test score to alleviate pressure and foster a growth-oriented mindset. These measures helped create a supportive environment, fostering student confidence and active participation during the oral assessment.

5.4 Overall take-aways

One of the largest benefits that oral exams afforded was the ability and place to have a candid conversation with a student about where they're excelling in their understanding and where they need extra support. This requires a particular faculty mindset, but with the option to revise their work and retake a limited number of oral exams, this can be a powerful iterative process founded in care. This also tended to be a reality check with students who were working with their classmates to realize that they didn't fully understand the material. They could rethink how they want to study and if there's an individual component that works well to supplement the collaborative portions of preparing for the oral exams.

Both professors found that engaging in a community of practice related to oral exams while implementing them was invaluable but agree that in a face-to-face course, as it would be difficult to replicate for classes larger than fifteen students without extra support. The biggest limitation to this type of assessment is the time required. While remote due to COVID-19, we had a lot of flexibility in our assignment times; reproducing this in a face-to-face class isn't impossible but would require greater flexibility and attention to detail with the time commitments and scheduling.

6 Acknowledgment of constraints

6.1 Institutional regulations

We would like to acknowledge upfront that there are significant limitations that might preclude others from implementing oral exams in their courses. Firstly, some institutions or departments might have regulations that dictate how exams and assessments must be conducted. This style of assessment might be out of compliance with those regulations. Class size and structure might be another limitation to if this type of assessment.

6.2 COVID-19 pandemic

One constraint faced by Dr. Gaines was that their institution prohibited the use of synchronous formats for courses conducted remotely. Their institution was in a remote, rural area and while remote, most students did not have reliable internet access at home. Problem-solving sessions were available every day, but students were not required to attend.

6.3 Logistics and time considerations

Setting up one-on-one meeting times was done either through the Canvas' scheduling feature or through Calendly. To find the time to schedule these meetings, synchronous sessions were dismissed for the exam period; similarly, lab periods were used to schedule sessions. If a student wanted to meet outside those times, that could also be arranged depending on the instructor's schedule. These exams were conducted in online courses, so time was more flexible than in traditionally in-person synchronous classes.

Data availability statement

The original contributions presented in this study are included in this article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

TG: Conceptualization, Methodology, Project administration, Writing – original draft, Writing – review & editing. NB: Conceptualization, Methodology, Project administration, Writing – original draft, Writing – review & editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to acknowledge and thank our students who were gracious and worked with us as we tried something new while navigating a pandemic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1379886/full#supplementary-material

References

Ausubel, D. P., Novak, J. D., and Hanesian, H. (1978). *Educational psychology: A cognitive view*, 2d Edn. New York, NY: Holt, Rinehart and Winston.

Bodner, G. M. (1986). Constructivism: A theory of knowledge. *J. Chem. Educ.* 63:873. doi: 10.1021/ed063p873

Boedigheimer, R., Ghrist, M., Peterson, D., and Kallemyn, B. (2015). Individual oral exams in mathematics courses: 10 years of experience at the air force academy. *PRIMUS* 25, 99–120. doi: 10.1080/10511970.2014.906008

Bowen, C. W. (1994). Think-aloud methods in chemistry education: Understanding student thinking. *J. Chem. Educ.* 71:184. doi: 10.1021/ed071p184

Burrows, N. L., Ouellet, J., Joji, J., and Man, J. (2021). Alternative assessment to lab reports: A phenomenology study of undergraduate biochemistry students' perceptions of interview assessment. *J. Chem. Educ.* 98, 1518–1528. doi: 10.1021/acs.jchemed.1c00150

Collins, A., Brown, J., and Holum, A. (1991). COGNITIVE APPRENTICESHIP: MAKING THINKING VISIBLE. *Am. Educ. Prof. J. Am. Fed. Teach.* 15, 6–46.

Cotten, S. R., and Wilson, B. (2006). Student–faculty interactions: Dynamics and determinants. *High. Educ.* 51, 487–519. doi: 10.1007/s10734-004-1705-4

Crawford, G. L., and Kloepper, K. D. (2019). Exit interviews: Laboratory assessment incorporating written and oral communication. *J. Chem. Educ.* 96, 880–887. doi: 10.1021/acs.jchemed.8b00950

Davis, M. H., and Karunathilake, I. (2005). The place of the oral examination in today's assessment systems. *Med. Teach.* 27, 294–297. doi: 10.1080/01421590500126437

Díaz, R. J., Glass, C. R., Arnkoff, D. B., and Tanofsky-Kraff, M. (2001). Cognition, anxiety, and prediction of performance in 1st-year law students. *J. Educ. Psychol.* 93, 420–429. doi: 10.1037/0022-0663.93.2.420

Dicks, A. P., Lautens, M., Koroluk, K. J., atnd Skonieczny, S. (2012). Undergraduate oral examinations in a university organic chemistry curriculum. *J. Chem. Educ.* 89, 1506–1510. doi: 10.1021/ed200782c

Flood, V. J., Amar, F. G., Nemirovsky, R., Harrer, B. W., Bruce, M. R. M., and Wittmann, M. C. (2015). Paying attention to gesture when students talk chemistry: Interactional resources for responsive teaching. *J. Chem. Educ.* 92, 11–22. doi: 10.1021/ed400477b

Gent, I., Johnston, B., and Prosser, P. (1999). Thinking on your feet in undergraduate computer science: A constructivist approach to developing and assessing critical thinking. *Teach. High. Educ.* 4, 511–522. doi: 10.1080/1356251990040407

Gilbert, J. K., and Treagust, D. F. (2009). "Introduction: Macro, submicro and symbolic representations and the relationship between them: Key models in chemical education," in *Multiple representations in chemical education*, eds J. K. Gilbert and D. Treagust (Dordrecht: Springer Netherlands), 1–8. doi: 10.1007/978-1-4020-8872-8_1

Goodman, A. L. (2020). Can group oral exams and team assignments help create a supportive student community in a biochemistry course for nonmajors? *J. Chem. Educ.* 97, 3441–3445. doi: 10.1021/acs.jchemed.0c00815

Gregory, S. E. A., and Jackson, M. C. (2017). Joint attention enhances visual working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 237–249. doi: 10.1037/xlm0000294

Hambrecht, G. (2003). Oral examinations: A new measure of learner success. *Delta Kappa Gamma Bull.* 69, 31–32. doi: 10.1002/ase.1252

Henderson, J., Lloyd, P., and Scott, H. (2002). "In the real world we're all put on the spot at some time or other, so you need to be prepared for it": An exploratory study of an oral method of assessing knowledge of mental health law. *Soc. Work Educ.* 21, 91–103. doi: 10.1080/02615470120107040

Heyneman, S. P., Anderson, K. H., and Nuraliyeva, N. (2008). The cost of corruption in higher education. *Comp. Educ. Rev.* 52, 1–25. doi: 10.1086/524367

Huxham, M., Campbell, F., and Westwood, J. (2012). Oral versus written assessments: A test of student performance and attitudes. *Assess. Eval. High. Educ.* 37, 125–136. doi: 10.1080/02602938.2010.515012

Iannone, P., and Simpson, A. (2015). Students' views of oral performance assessment in mathematics: Straddling the 'assessment of' and 'assessment for' learning divide. *Assess. Eval. High. Educ.* 40, 971–987. doi: 10.1080/02602938.2014.961124

Joughin, G. (1998). Dimensions of oral assessment. *Assess. Eval. High. Educ.* 23, 367–378.

Joughin, G. (2010). *A short guide to oral assessment*. Leeds: Leeds Met Press in association with University of Wollongong.

Kharkhurin, A. V. (2014). Creativity.4in1: Four-criterion construct of creativity. *Creat. Res. J.* 26, 338–352. doi: 10.1080/10400419.2014.929424

Kiste, A. L., Hooper, R. G., Scott, G. E., and Bush, S. D. (2016). Atomic tiles: Manipulative resources for exploring bonding and molecular structure. *J. Chem. Educ.* 93, 1900–1903. doi: 10.1021/acs.jchemed.6b00361

Kozma, R., and Russell, J. (2005). "Students becoming chemists: Developing representationl competence," in *Visualization in science education*, ed. J. K. Gilbert (Dordrecht: Springer Netherlands), 121–145.

Lubarda, M., Delson, N., Schurgers, C., Ghazinejad, M., Baghdadchi, S., Phan, A., et al. (2021). "Oral exams for large-enrollment engineering courses to promote academic integrity and student engagement during remote instruction," in *Proceedings of the IEEE Frontiers education conference FIE*, (Lincoln, NE), doi: 10.1109/fie49875.2021.9637124

McCloud, L. I. (2023). Keeping receipts: Thoughts on ungrading from a black woman professor. *Zeal J. Lib. Arts* 1, 101–105.

Muldoon, J. A. (1926). The value of oral examinations. *J. Chem. Educ.* 3:773. doi: 10.1021/ed003p773

Orgill, M., and Sutherland, A. (2008). Undergraduate chemistry students' perceptions of and misconceptions about buffers and buffer problems. *Chem. Educ. Res. Pract.* 9, 131–143. doi: 10.1039/B806229N

Paul, A. M. (2021). *The extended mind: The power of thinking outside the brain*. Boston, MA: Houghton Mifflin Harcourt.

Ramella, D. (2019). "Oral exams: A deeply neglected tool for formative assessment in chemistry," in *Proceedings of the active learning in general chemistry: Specific interventions ACS symposium series*, (Washington, DC: American Chemical Society), 79–89. doi: 10.1021/bk-2019-1340.ch006

Roecker, L. (2007). Using oral examination as a technique to assess student understanding and teaching effectiveness. *J. Chem. Educ.* 84:1663. doi: 10.1021/ed084p1663

Shteynberg, G., and Apfelbaum, E. P. (2013). The power of shared experience: Simultaneous observation with similar others facilitates social learning. *Soc. Psychol. Pers. Sci.* 4, 738–744. doi: 10.1177/1948550613479807

Theobold, A. S. (2021). Oral exams: A more meaningful assessment of students' understanding. *J. Stat. Data Sci. Educ.* 29, 156–159. doi: 10.1080/26939169.2021.1914527

Tienson-Tseng, H. L. (2019). "Best practices in summative assessment," in *Proceedings of the biochemistry education: From theory to practice ACS symposium series*, (Washington, DC: American Chemical Society), 219–243. doi: 10.1021/bk-2019-1337.ch010

Walker, J. P., and Sampson, V. (2013). Learning to argue and arguing to learn: Argument-driven inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. *J. Res. Sci. Teach.* 50, 561–596. doi: 10.1002/tea.21082

Waterfield, J., and West, B. (2005). *Inclusive assessment in higher education: A resource for change*. Plymouth: University of Plymouth.

frontiers | Frontiers in Education

TYPE Curriculum, Instruction, and Pedagogy
PUBLISHED 31 May 2024
DOI 10.3389/feduc.2024.1400058

OPEN ACCESS

EDITED BY
Michael Wentzel,
Augsburg University, United States

REVIEWED BY
Brian Woods,
Saint Louis University, United States
Brandon J. Yik,
University of Virginia, United States

*CORRESPONDENCE
Matthew J. Mio
✉ miomj@udmercy.edu

RECEIVED 13 March 2024
ACCEPTED 20 May 2024
PUBLISHED 31 May 2024

CITATION
Mio MJ (2024) Alternative grading strategies in organic chemistry: a journey.
Front. Educ. 9:1400058.
doi: 10.3389/feduc.2024.1400058

Alternative grading strategies in organic chemistry: a journey

Matthew J. Mio*

Department of Chemistry and Biochemistry, University of Detroit Mercy, Detroit, MI, United States

The global pandemic forced educators at all levels to re-evaluate how they would engage content, generate relevance, and assess the development of their students. While alternative grading strategies were not necessarily new to the world of chemical education research (CER) in 2020, the pandemic accelerated the examination of such principles for instructors who wished to prioritize learning over compliance with course policies and eschew points for grading. This article describes the transformation of a traditional lecture organic chemistry course under new, standards-based principles. Working from a discrete set of grouped learning outcomes, these courses aim to clearly define standards, give helpful feedback, indicate semester-long student progress, and allow for reattempts without penalty using a token system. Students explore both core (all of which must be passed in order to show progress) and non-core (students can choose a majority of these to pass) learning outcomes in a variety of formative and summative assessment approaches. Unique to the reformatting of these courses is the use of collaborative, take-home assessments, integration of multiple-choice questions, ungrading of student-submitted summary notes, live solving of select problems with peer feedback, and a learning check system that reinforces the flipped nature of content delivery. Another distinctive feature of the courses is the requirement for students who reach a certain threshold of repeating learning outcomes to perform in-person problem solving for the instructor. Overall, students report that the structure of the courses reduces their general and specific anxiety, lowers the temptations to challenge their academic integrity, and increases their own learning self-monitoring, reflecting known pedagogies of metacognition. The instructor reports that alternative grading strategies take far less time, generates more meaningful feedback, and shifts student attitudes away from final grading and toward genuine learning.

KEYWORDS

alternative grading, ungrading, organic chemistry, chemistry education research, metacognition, specifications grading

1 Introduction

"The most difficult academic year in the history of American higher education." This was the social media post that stared back at me in early May 2022. I thought for a moment: how could any academic year be more difficult than 2019–2020, when we all had to pivot our teaching so severely that we were unsure what the nature of learning was anymore? And what about 2020–2021? That was the year of extreme social upheaval and a continuance of uncertainty. Semester after semester, the highly trained professors of the Academy found themselves completely unprepared. Student pushback against tried-and-true methods made the most experienced educator question their abilities, wherewithal, and dedication to the art of teaching and learning. Amid this, after 20 years of teaching organic chemistry, I had my worst semester ever.

Largely, I had the same students during Winter 2022 (Detroit Mercy refers to the January to April semester as "Winter") that I'd had during Fall 2021. What had gone wrong? Student behavior had devolved into raised voices in public, and direct confrontation of my teaching methods. Problems which would have barely challenged students only a few years back were labeled "impossible," and students outright refused to do the work of the class. Students ignored official course announcements, even those posted online and through texting apps, while verbal assurances mattered less and less as daily attendance plummeted to less than 50% of enrollment. In lieu of course content, the majority of my job became explaining, over and over, how grades were calculated in the course. All of this occurred despite a fill-in-the-blank grade calculation worksheet from my syllabus that other colleagues had termed "babying" my students. I survived to the end of the term, but at the expense of my emotional and intellectual wellbeing. Over two decades of teaching, a personal metric for the success of a course was the number of students that "followed-up" after grades were posted (emailed, called, or texted to complain about their final grade). By this measure, the course was a nearcomplete failure – 27 of my 115 students demanded 30 min meetings and a recount of every point gained and lost over the 16 week period.

A regular semester would have had only two or three such contacts. As the warm weather of spring took over, I was exhausted, and the prospect of licking my wounds and doing it all over again in fall seemed more than daunting.

"The most difficult academic year in the history of American higher education." It was happening to other professors, as well. Maybe it wasn't me? Certainly, the pandemic had affected students and their learning. They seemed ill-prepared, discourteous, and refusing to adapt to the rigors of higher education. I retreated into the world of reading and reflecting, hoping that I could research my way out of the most serious existential crisis of my career as an organic chemical educator. As previous confidence in my ability to do this job well faded, my beliefs that a better way must exist grew.

Contemplation brought me quickly to a number of conclusions: (1) of course the Covid-19 pandemic was affecting students, and it was my job, as the more experienced professional in the room, to help them navigate the rough waters; (2) many in higher education knew that the pandemic had not brought these issues anew, but had exacerbated issues that were already fomenting long before 2020 (McMurtrie, 2023); (3) any lost zeal I was projecting to students was being projected back to me, a dark cycle where good learning was breaking down. In the end, my reflections always brought me back to that fundamental, but unwritten rule of the classroom: that the professor is in charge of the content, methods, and environment of learning. Pandemic or not, I had let my students down by not responding appropriately to the trials of the period.

During the pandemic, it seemed certain that every semester had its own unique challenges so that each was different than the last. While responding nimbly to these ever-changing issues was easier said than done, I sensed that the time for change had arrived. It was time to fully overhaul my courses, and to do so in such a way that I stayed true to my core values of engaging content, generation of relevance, and assessment of student development over the length of the course. I discovered the *Grading for Growth* blog of Talbert and Clark (n.d.) and was immediately entranced by their no-nonsense take on cutting-edge pedagogical research surrounding alternative grading methods. At the heart of my reflection after a tough semester was finding a way to prioritize learning over compliance with course policies and eschew points for grading in my courses. I knew I had found what I was looking for when I read the story of the first-year calculus student.

In short, Talbert and Clark talked about an eager, first-year college calculus student who started the semester strong, but slowly ebbed in their work until, after failing a number of assessments and the final exam, earned an F grade for the course. Upon examination of the student's work over time, the professors noted that the student had a perfect mastery of the concepts of calculus that happened to always lag a few days or weeks behind the current topic coverage. Suddenly, the F grade was a metric of how well the student performed at the current topics, not a measure of 16 weeks of progress. In asking other faculty about this student as a hypothetical, Talbert and Clark found that many thought she deserved an A for her efforts – the grade farthest from that which she earned! After all, she was performing admirably, just a few lessons behind the rest of the class. This story spoke directly to me as I thought about how my own students earned their letter grades and how the assumption of learning connected to these grades has historically been viewed.

This article describes the transformation of a traditional lecture organic chemistry course under new, standards-based principles. First, a brief background on ABCDF grading will be described, followed by foundational principles for alternative grading strategies (AGS). Specific incorporation of these principles into an organic chemistry lecture and lab course will be explained, in conjunction with a number of preliminary outcomes from AGS for both student and instructor.

1.1 Background – a brief history of letter grades

According to Feldman (2018), the history of ABCDF grades in the United States is best viewed through the lens of the 20th century, even though the earliest recorded use of grading goes back to 1785 at Yale University. It was not until approximately 100 years later that ABCDF-style grades would become more widespread in the United States, and not until the 1940s that institutions of higher education began ABCDF grade usage extensively. Since 67% of US primary and secondary schools used letter grades by 1970, overall the concept is not a particularly old one; and for members of Generation X (like the author), grades fit comfortably into our lifetimes.

Arguably then, letter grading and students being motivated in academic studies by such grading (and its close cousin percentages) is a construct of the last 50 years of American educational history. Since the unwritten rule of education is that most teachers will teach the way they were taught (Oleson and Hora, 2014), it is not unexpected that these ideas have been passed down to following generations. Furthermore, when letter grading is married to concepts as powerful as admission, school ratings, faculty salaries, and district funding, instances of decoupling learning from grades are bound to increase. This brings us to Goodhart's Law, first written about in 1975 (Goodhart, 1975): "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes." In other words, "When a measure becomes a target, it ceases to be a good measure." While written for modern-age British economics, the application of Goodhart's Law to the 20th century US phenomenon of standardized testing has been written about extensively. As funding for public schools, in particular, became more and more tied to student performance on mandated tests, a corollary of Goodhart's Law came into view – Campbell's Law (Campbell, 1979). Again, tied to the area of economics, this law states: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor." Taken together, Goodhart and Campbell paint a dark, but inevitable start to the 21st century in education. When important metrics like student test scores and grades count for decisions, especially funding and admissions, the more those processes will be susceptible to fraud for the sake of success.

Moreover, pondering these economic philosophies toward education cannot be complete without considering arguments about the meaning of academic rigor, a "sign of the times" in teaching for the first 20 years of the 21st century. A hard concept to define, many educators and students alike know it exists, but do not know how to apply it or what it is actually good for. Student and faculty attitudes toward academic rigor have been studied since the early 2010s with both parties agreeing that rigor is multifaceted, but with students caring most about grades and faculty caring more for learning (Draeger et al., 2013, 2015). Nelson (2010) refers to the muddle of these conflicting viewpoints as "dysfunctional illusions of rigor" and chooses to recast the negativity of the fixed and deficit mindsets with challenges to the Academy. For example, someone might claim that hard courses "weeding out" students helps society rid itself of students with poor preparatory skills or lack of motivation as in the widely-publicized and debated organic chemistry classes of Prof. M. Jones (Supiano, 2022). A more realistic view is that poor or ineffective pedagogy is more likely to blame for student failure. At its core, argues Nelson, are two additional, dysfunctional illusions: (1) that traditional methods of instruction offer effective ways to teach undergrads, and (2) that massive grade inflation is a corruption of standards. He counter-argues that it is more realistic to view lecturing as considerably less effective than other methods and distinguishing between unjustifiably high grades and more effective pedagogy giving better student achievement. A turn toward the positive could lead any educator toward the central illusion of academic rigor: that faculty know enough in the modern era to revise their courses and curricula for the betterment of student learning. Nelson posits that teaching and curricular revision should be driven by pedagogical research and best practices, or DBER (discipline-based education research). In short, the burden has shifted – using what amounts to ancient methods now requires more rationalization than the existing data on best practices in education, including all that we know about ABCDF grading.

Adding issues of diversity, equity, and inclusion (DEI) to this conversation only bolsters the argument for a new philosophy of grading. Article after article finds traditional grading methods to be ineffective, harmful, and unjust (Ko, 2019). We must be purposeful about our instructional methods to reach as many students as possible, but could not possibly tailor individual instruction plans for each student. At my home institution, University of Detroit Mercy, we learn to place contemplatives into action as part of our dual charism Mission (University of Detroit Mercy Mission and Vision, n.d.). This translates into taking care of the whole person ("cura personalis") and all the people ("cura apostolica"). In a recapitulation of the literature findings about rigor, we know that it exists, and faculty wish for our students to learn, so we must search for the most inclusive and equitable ways to do so.

As a final item of focus, many educators exploring alternate grading modes find "the question of the C student" to be the definitive threshold: when thinking about students who earn a C grade in your courses, would you rather they be able to do a small number things very well, or do everything in the course average? There is no correct response to this question – it is meant to be a frame from which educators can plan for what tasks or standards their students should be able to complete with mastery at the end of a lesson, unit, or course. As I considered this fundamental question, I realized I had taken the first steps on a journey of alternative grading, but I needed more input from experts to define my own, organic chemistry version of these varied principles.

1.2 Background – core philosophies for alternative grading

In preparation for the redesign of my course from an alternative grading perspective, I delved into the pedagogical philosophies of purposeful instruction in metacognition, specifications grading, ungrading, matters of grading equity, and evaluating overall student progress.

The work of McGuire (2015), a researcher in the discipline of chemistry, centers around her enthusiasm for metacognition and instructing students in the ways of their own learning. Such an element would be vital to convincing students of the value of a nontraditional evaluation scheme. I have used the "Study Cycle" graphic from her work to help my students plan for their daily and weekly work for over 10 years, and was eager to learn more about the learning habits of my students (Louisiana State University Center for Academic Success, n.d.). McGuire's theories on how to aid students in thinking about their thinking was pivotal in my building an introductory module for my future students, especially in light of the graphical and conceptual nature of organic chemistry. McGuire's discourse on treating separate subjects/disciplines as different when studying became the foundation for my own opening remarks to students at the beginning of the semester.

Nilson's (2014) work on specifications grading was also an inspiration. Even before 2020, I had investigated the use of "all-or-nothing" style problems on my unit-ending assessments to emphasize the difference between full and partial skill mastery. Again, there are as many takes on "specs grading" as there are disciplines of study, but Nilson's theories on connections among learning outcomes, grading criteria, and agreed-upon standards helped me converge on what was right for my students. Additionally, Prof. Susan D. Blum's work in the area of "ungrading" and the disconnect between learning and schooling also brought insight to my ongoing course redesign (Blum, 2020). Her interpretation of the changing educational landscape and how to prepare students for the working world through timely, constructive feedback with no grade attached appealed to my affective side of teaching – why did I always feel guilty when grades were poor for a student? Making the correlation between these emotions and the role that I play in the classroom reinforced my confidence in being able to provide students meaningful feedback and set them on a trajectory of growth. Joe Feldman's writing in the area of grading for equity also helped me understand my students' previous experiences and how they shape their attitudes toward learning in higher education (Feldman, 2018). With knowledge stemming from K-12 experiences, Feldman differentiates compliance with teacher demands and learning to tear down systemic achievement and opportunity gaps for students.

The main motivation for my journey was Talbert and Clark's *Grading for Growth* blog and monograph (Feldman, 2018; Clark and Talbert, 2023). The sensible prose of their writing, in light of the theory and practice of alternative grading, drove home the need for an overhaul in my lecture and lab courses. Talbert and Clark's work made the seemingly insurmountable task of course redesign feel more feasible and, in post-pandemic times, necessary for matters of inclusion and equity. My research was complete, and I was ready to decide the core values that would drive the overarching change in my organic chemistry courses.

1.3 Foundational principles of course redesign with AGS

Even in its nascent form, the amount of research literature surrounding the concepts of alternative grading can be overwhelming (Clark, n.d.; Townsley, n.d.). There have been a number of discipline-based articles, as well as writings on central concepts in the field. In addition, nearly every corner of modern education has been examined, from pre-school to graduate-level instruction. After my personal survey of the alternative grading landscape, I attempted to synthesize a smaller number of essential concepts for my courses, some more fundamental to my pedagogy, some more specific to the content of organic chemistry. This section will detail the "non-negotiables" which served as the guideposts for my course redesign.

Paramount amongst the ideas I was to explore was a commitment to experimentation in the classroom. A common criticism in the world of chemical education research (CER) is that scientific educators are only too eager, mostly from their training, to research in the laboratory, but loathe to do so in the lecture hall. Such inertia is understandable; however, I have found that an enthusiasm for new methods, when explained to students at the start of the experiment, can re-invigorate the post-pandemic classroom environment. While some worry that altering content delivery and assessment methods negatively affects student learning, numerous studies [some in chemistry (Houchlei et al., 2023)] have shown the ability for student metacognition and resilience in the face of shifting pedagogies, even in the same subject area.

In addition, I aimed to hold fast to the following principles in my course redesign: eliminating points for assessment, grading major and minor assessments using specifications grading, monitoring placeholders in the gradebook for work that has not yet met specifications, and enacting all changes with an eye on inclusivity and equity for all students.

For over two (2) decades, I have awarded my students numerical points for correct responses on major and minor assessments. At the end of the course, one would simply need to divide the points earned by the points possible to determine a percentage, and therefore, letter grade. On paper, this is a simple calculation, especially in light of the worksheet I would attach to the last page of my syllabus: complete all assignments, record points earned, divide by maximum possible points. In the years leading up to and after the pandemic, I noticed a trend that I was spending more of the last month of the semester walking my students through this calculation than discussing content or exercises. In addition, I had long wondered what awarding 12/15 versus 13/15 meant for my students, even with a detailed rubric. What good was this rubric if students did not read it and I could not quickly summarize it? During the first two (2) years of the pandemic, I experimented with specifications grading on one (1) problem per major assessment. Students were intrigued by the "all-or-nothing" nature of the grading associated with this problem, but in the end disliked it because there was no way to demonstrate partial knowledge. I liked it because it was quick to grade and allowed me to give more rich feedback. Not allowing students to earn traditional "zeroes" for missed assignments came to me after reading Feldman's treatise on the message it sends students: earning a zero for work not done gets them off the hook and gives license to not learn the skill at hand (Feldman, 2018).

This concept fits into a cornerstone of alternative grading: recognition of the difference between qualitative and quantitative numeration. In other words, a zero can mean something else besides a fraction like 0/100 – it can mean "not done yet." Lastly, I wanted to make all the changes in my courses reflect a deep interlace of inclusivity and equity for all my students. I was particularly concerned for neurodivergent students whose learning preferences so drastically vary from the accepted norm and are often excluded from the areas where neurotypical students are given access.

In summary, I appreciated the practical nature of Talbert and Clark's "Four Pillars of Grading for Growth" (Clark and Talbert, 2023) and modified them to include the important points mentioned in this section:

1. Clearly defined standards (by way of pinpointed learning outcomes).
2. Give helpful feedback (verbal/written, not numeric).
3. Allow for reattempts without penalty (use a token system).
4. Indicate semester-long student progress (growth from unit to unit defines final grade).

Using these four values to guide my course re-design, I began to re-evaluate every course policy, exercise, assignment and major/minor assessment. The only thing that did not change in my courses was the content.

2 Planning the lecture course – organic chemistry I

Organic chemistry affects every person at every moment of their lives in ways that have only been scientifically examined for the last 150 years – in other words, it is a fascinating subject to learn. The recursive relationship between structure and reactivity drives the very engine of life and is the basis for much of modern materials – clothes, building resources, electronics, and vehicles. There is merit in the study of this subject based on its graphical nature, its logical and creative approach to problem solving, and its vastness. Some have wondered about organic chemistry's longstanding inclusion in the undergraduate pre-health curriculum, but few argue with its unique place in the world of higher education (Dixson et al., 2022). In many ways, I have found that organic chemistry's distinctive place in the undergraduate science curriculum makes it ideally suited to experimentation with alternative grading strategies. The Fall 2022 semester would offer a chance to experiment with my Organic Chemistry I lecture course. This three (3)-credit class would have 75 total enrollees, mostly 2nd- and 3rd year students, and would meet for 75 min twice a week, plus a 50 min recitation at the end of the week. Topic coverage would be traditional, including structure, spectroscopy, and introductory reactivity (acids/bases, substitution, elimination, rearrangement, electrophilic addition). But where to begin?

2.1 Clearly defined standards

First and foremost, there can be no "clearly defined standards" without a discrete set of learning outcomes. There are numerous models for how to map student learning onto a set of outcomes for the purposes of evaluation or grading. Upon first being introduced to Bloom's taxonomy in my early years on the faculty, I was attracted to the order with which epistemological exercises could be categorized as learning (Bloom and Krathwohl, 1956). Moving from lower levels (knowledge, comprehension) to higher levels (synthesis, evaluation) gave students a sense of learning progress and educators a pathway to follow for instructional activities. This learning framework emphasizes what a student possesses at the end of an action and is basically the model for the modern learning outcome – i.e., any statement that begins with, "At the end of this lesson, a student should be able to XYZ if they have mastered the outcome." Anderson and Krathwohl's revision of Bloom's Taxonomy preserves the order of the original, but it reframes the student's role as demonstrator of a skill and rewrites the differing levels as active verbs (Anderson et al., 2001). Perry's (1970) casting of a four-step model aimed to frame undergraduate student learning as growth and Wolcott and Lynch go one step further to specifically map grades to outcomes by way of students' ability to re-envision information (Dixson et al., 2022).

With these "maps" in hand, I began the culling of learning outcomes from a comprehensive list down to the particular learning units of both my Organic I and Organic II lecture courses. Before the pandemic, I had distilled my lecture notes into lists of micro-outcomes that better resembled a litany of individual facts than learning outcomes. In fact, there were over 300 of these, and this made the work untenable for a 16-week course (or two). I decided to group, rewrite, and reduce the total number of outcomes to fit into six (6) two-week parts. In addition, for Organic Chemistry I, I followed the advice of Talbert and Clark to consider "Core" learning outcomes for fundamental concepts in an introductory course. What resulted from this work became the superstructure of my Organic I Lecture course in Fall 2022:

> Part 1 – Convention, Composition, and Constitution Core Learning Outcome – drawing and ranking resonance structures
>
> Part 2 – Constitution, Conformation, and Configuration Core Learning Outcome – identifying isomeric relationships
>
> Part 3 – Spectroscopy, Spectrometry, and Spectrophotometry Core Learning Outcome – interpreting (McMurtrie, 2023) H NMR data (symmetry)
>
> Part 4 – Integrated Spectroscopy and Organic Reaction Basics Core Learning Outcome – solving a structure from NMR, MS, IR data
>
> Part 5 – Substitution and Elimination Pathways Core Learning Outcome – using reaction criteria to determine S_N and/or E pathways
>
> Part 6 – Electrophilic Addition Pathways Core Learning Outcome – solving retrosynthesis problems

The core learning outcomes were borne out of a thought experiment many of us have considered – all details aside, what would we be horrified to find out our students did not learn at minimum in our courses? (Figure 1). I had already decided to make some of the non-Core learning outcomes in these Parts optional by way of what constituted a "Pass" for the main summative assessment at the end of the unit. In my first iteration of Organic II Lecture, I decided to abstain from having Core learning outcomes in lieu of review outcomes that recapitulated the concepts of the first term.

2.2 Give helpful feedback

2.2.1 Take-home problem sets

Fast-forwarding to the end of a course unit (I call them "Parts") from the student perspective, I planned to assign multi-day, collaborative take-home problem sets (THPS) as the main summative assessments of the course. For over two (2) decades, I experimented with recent, literature-based, multi-day, peer-collaborative, free-response problem sets to evaluate student learning. Primarily based on the CER (claim, evidence, reasoning) structure (Brunsell, 2024), students were asked to extend their skills past the elementary level and combine learned principles in new applications of problem solving. I retained this method of evaluation, but now linking each specific learning outcome to a single problem. With six (6) learning outcomes per THPS and one (1) of them being a Core learning outcome, students were challenged to complete the set in 4 days or less. If a student's response did not pass on the Core learning outcome (*vide supra* 2.1) or on more than one (1) of the five (5) non-Core learning outcomes, the entire THPS would be termed "Try Again" and sent back to the student for editing or overhaul, depending on the number and magnitude of the misconception. Using tokens (*vide infra* 2.3), students could revise any problem not passed on the first try in an attempt to pass the THPS and earn a "badge" for that Part of the course. The number of badges accrued over the length of the course reflects their final grade. Badges were used to differentiate the successful completion of a THPS versus an unsuccessful attempt.

This style of assessment is time-consuming to write and evaluate. The second principle of "give helpful feedback" is based on a verbal, non-numeric model for constructive criticism. Therefore, I left numeric grading behind and went to an all-or-nothing evaluation style. Heavily involving the ideology of specifications grading, though, it is relatively easy to write "Pass" or "Try Again" with a sentence or two of review. For example – if no less than three (3) resonance contributors were asked for by the problem, but only two (2) were given by a student, this would immediately merit a "Try Again." More fundamental misunderstandings (an incorrect charge on an atom or

CHM 2270-01
Fall 2023
Learning Outcomes (highlighted LO is the Core outcome for this Part)

Part 0 – Introduction to CHM 2270-01 F23
At the end of this Part, students should be able to describe in detail important elements of the course, professor, and the study of organic chemistry.

At the end of Part 0, students should be able to:
0A1: Detail the daily work, assessments, and grading philosophy of the course.

0A2: Specify one weird thing about Prof. Mio.

0B1: Define organic chemistry.

0B2: Detail the importance of studying organic chemistry for their future career in science.

Part 1 – Structure: Convention, Composition, and Constitution
At the end of this Part, students should be able to connect knowledge from General Chemistry to Organic Chemistry and recognize the first three (3) C's of organic chemistry: convention, composition, and constitution.

At the end of Part 1, students should be able to:
1A1: Integrate existing knowledge of ionic/covalent bonding, lone pairs, formal charge, valence, octet rule, and Lewis structures to generate chemically-correct organic building blocks and line-angle structures.

1A2: Define electron pair domain (EPD), count EPDs on building blocks, and use EPD knowledge to determine 3D geometry and bond angles of a tetrahedral, trigonal planar, linear building block.

1B1: Understand the foundational principle of IUPAC nomenclature (one structure = one name) and apply the three (3) rules of organic IUPAC nomenclature: longest/parent chain, lowest number substituent combination, alphabetical order of substituents.

1B2: Draw multiple, valid resonance structures for line-angle structures using curved arrow notation and rank drawn resonance structures in order of their energy, stability, and contribution to the overall electronic nature of the structure.

1C1: Use Ω and the Darling model kit to determine molecular formula (MF), including an understanding of CPK coloring, bond length/angle, lone pairs, and markers.

1C2: Define and memorize the major functional groups (FGs) and their categories based on carbon atom oxidation number (number of carbon bonds to heteroatom).

FIGURE 1
The part 0 and 1 learning outcomes for the organic chemistry I lecture course.

missing multiple bonds) would also merit "Try Again," but with more direct feedback on what to change to make the response completely correct. In general, students understand these methods, but they do not like them when first introduced. Until they see the style of feedback offered a few times, they view "all-or-nothing" evaluation as punitive. To aid in students' quick adjustment to the alternative grading model, I discuss with students what small errors (i.e., those that did not directly affect a student's achievement of the learning outcome) were present and how to identify them. They quickly learn the best way to plan their responses to earn "Pass" on the first try.

Even if students do not earn a "Pass" on the first try for the THPS, tokens can be earned and used (*vide infra* 2.3) to review and revise their responses to full accuracy. In the spirit of the third and fourth 'pillars' (allow for reattempts without penalty and indicate semester-long student progress) students are allowed one (1) chance to correct on their own, followed by any further chances after the second being performed in person as an oral exam. Speaking to pandemic positives, many of us in higher education have returned to the medieval roots of university education with oral exams, being possibly the last bastion of extemporaneous evaluation in a world of Chegg and domyclasscholarly.com.

2.2.2 Skeleton notes, learning checks, and exercises

If learning outcomes are the first steps of a Part and take-home problem sets are the summative assessments, what occurs in the middle — the domain of formative assessment? First, as a result of the older learning outcome catalog, I transformed my personal lecture notes into "skeleton notes" videos for students to watch for content introduction. In the wake of traditional lecture courses fading away, the clear, concise, engaging, and scrollable content introduction video reigns supreme. To ensure students engage with the videos, they are required to prepare enough to pass the next day's learning check (LC) – a three-question quiz (multiple choice, short response) where the only purpose is for students to take the LC, not give correct responses. The sum result is a flipped classroom (Bergmann and Sams, 2012), wherein the vast majority of class time is spent formatively evaluating progress and running problems, with the vast majority of outside classwork introducing new concepts. The LCs also allow for an easy attendance policy to be enacted.

While in class, students engage in think-pair-share, small-group, and discussion-based active learning exercises to extend their content knowledge beyond the introductory video. Students are always encouraged to attend office hours with more specific questions, but they are trained to bring broader concerns to the whole class during regular meetings. Homework, both online and written, is used to follow-up on new topics after class meetings. In addition, submission of summarized notes at the end of the course Part is also used for content follow-up. Students are asked to fit the main concepts of the last topic set into a certain page space, then their submissions are ungraded (Bergmann and Sams, 2012; Bergmann and Sams, 2014). As per many items in the course, the number of summarized notes turned in on time and meeting a minimum set of criteria affect a student's final grade in the course.

2.2.3 Part summary quizzes and problem days

For the last class meeting of a course Part, problem day and a Part summary quiz are administered. Problem Day is a chance for students to think on their feet and solve smaller problems on the whiteboard. At the start of the term, students are pre-assigned multiple problem dates to reduce anxiety about this very out-in-the-open work. On the day of, students are randomly assigned a partner and are set to a problem, either as a "writer" or a "reviewer." Writer pairs respond directly to the question after consultation with peers and the professor. They must draw out a detailed response next to the original problem on the board. Reviewer pairs work together to evaluate the response on the board and make edits as needed. They will make a 20- to 30-s presentation at the end to talk about the original problem, the writer response, and any changes they could make to the response. Again, the only tracked quantity for problem day is whether students participate based on their randomly assigned role. The activity gives a sense of finality to the set of learning outcomes in the past unit and prepares students for the collaborative nature of the Take-Home Problem Sets.

Part Summary Quizzes (PSQs) were generated as a response to the number of pre-health students at Detroit Mercy (essentially the majority of all science majors) who will take multiple choice, organic chemistry-based admissions exams. An online course response system (CRS) is used to administer a multiple-choice quiz with one question per learning outcome at the very end of problem day. Students who participate in the PSQ earn credit toward their final grade, and those that "pass" with the same rubric as the Take-Home Problem Set (students must pass the Core learning outcome and 4/5 of the non-Core learning outcomes) can earn a token as a reward for keeping up with the material. Detroit Mercy also has a well-established tradition of recitation sessions for science lecture courses. These non-required sessions are independent problem-solving class meetings where new material cannot be covered, but skills can be practiced. The CRS is used to practice new material with multiple choice mini quizzes that can earn students tokens for simply being present and taking them, as well as additional tokens for passing with a number of questions responded to correctly.

2.3 Allow for reattempts without penalty

The next pillar of the alternative grading scheme requires what is commonly known as a token system. While the administration of the token system may seem like busywork for the educator, I assure you it's no more complicated than the multiple weighted-point calculations that we are all used to with percentage grades! As a baseline, tokens are introduced to students as a system that will encourage students to become proficient in as many course learning outcomes as they choose. Students can redeem tokens: to extend a deadline for submission, to replace a "Try Again" mark to "Pass" for a lower-level item (non-summative assessment, THPS), to replace an online homework submission, to allow for a revision of a Take-Home Problem Set where a badge was not earned, or on some other bending of the course rules agreed upon by the student and the professor.

Every student in the course starts with one (1) token. Students can earn more tokens throughout the semester by: (1) attending and participating in recitation, (2) getting a minimum of 2 out of 3 RQuiz questions correct at the end of recitation, (3) attending and writing a one-page reflection for a Detroit Mercy event approved of ahead of time, (4) uploading complete THPS responses early, (5) attending in-person office hours once in the first two (2) weeks of the term, (6) enrolling in the course messaging app by Friday of the first week of classes, (7) passing the Part Summary Quiz, and (8) responding in a timely manner with regard to grade check-in. Some token-earning events are one-time-only, others can be completed multiple times during the semester. Some events earn partial tokens, while others

FINAL GRADE BUNDLING TABLE

Assessment → Grade↓	Learning Checks (out of 19)	Notes/ODB (out of 6)	Achieve (out of 7)	Problem Day (out of 1)	Part Summary Quiz (out of 6)	THPS Badges (out of 7)
A	19	6	7	1	6	7
A-	18	6	7	1	6	7
B+	17	6	6	1	6	6
B	16	6	6	1	6	6
B-	15	6	6	1	6	6
C+	14	5	5	1	5	5
C	13	5	5	1	5	5
C-	12	5	5	0	5	4 or less
D+	11	4	4	0	4	4 or less
D	19	4	4	0	4	4 or less
F	9 or less	3 or less	3 or less	0	3 or less	4 or less

FIGURE 2
The final grade bundling table for the organic chemistry I lecture course.

earn full ones. In my courses, there was no maximum set on token earning and the average earned was approximately twelve (12).

Redeeming tokens is done through an online form, the link to which is displayed prominently on the course Learning Management System page. Once students submit this form, the student and the instructor receive an email receipt. This receipt serves as official approval, allowing students to immediately do the thing they redeemed the token for. Some items in the class require one (1) token to redeem (lower-level item replacement or deadline extension), while others need more than one (1) token (homework replacement, revision of a Take-Home Problem Set). The token total is updated in the professor's gradebook.

As a matter of record, tokens cannot be shared, and unused tokens will be discarded at the end of the semester. The earning and redeeming of tokens have been shown to reduce student and instructor anxiety, as they allow for a universal correction coefficient for many of life's unexpected twists and turns (Clark and Talbert, 2023). Tokens make the many iterations needed to master a learning outcome possible and can be used to make due dates flexible. In short, there are various ways to earn tokens and various items students can redeem tokens to effect.

Students are instructed in the first week that this type of pedagogy nearly always generates a "token economy." This term refers to the fact that students are advised to earn as many tokens as they think they will need to use. Some students may be late to class often (missing low-level Learning Checks) or need multiple revisions of their Take-Home Problem Sets. Statistics are supplied to current students based on the last term of token usage; for example, in Fall 2022, students used 77% of their tokens to revise THPS responses and earned 42% of their tokens by choosing to attend optional recitation.

Having used alternative grading in my courses for the past three semesters, I have noticed that tokens afford two (2) major benefits for the instructor. First, students tend to recognize the positive reward of a token early in the term, even though they may not have ever used the system before. In the Fall 2023 iteration of my course, over 89% of my 60 students responded during a week 1 assessment that they did not understand how the token system worked but had a favorable view of the concept. Bringing students on board with a token system may be the easiest part of making a transition to alternative grading. Secondly, tokens allow for a number of "normal" course operations to be incentivized. I have given students a colored notecard to hand to me during the first two (2) weeks of the term at office hours to earn a token reward. I have also asked students to self-determine their midterm and final grades with evidence to earn a token. This practice, in particular, opens up the lines of communication with students who over- or under-estimate their performance. Need students to sign up for your class messaging app? Offer them a token. The possibilities are endless and can apply to multiple modes of instruction.

2.4 Indicate semester-long student progress

From the very start of the course, students are familiarized with the concept of the final grade bundling table (Figure 2). As previously mentioned for the multiple dimensions of the course, a minimum number of "passes" needs to be earned to reach a certain final letter grade level. Several key concepts of alternative grading are reflected directly or indirectly by this key component of the syllabus. First off, students are presented with the fact that the course will not be based on accrued points being divided by possible points and then charted against the percentage grade scale. Practically speaking, this was the largest departure from traditional course operation for my students and needs special emphasis during the first weeks of class. Second, in most cases the final grade bundling table makes clear that students do not need to "do everything" in the course to earn an A final grade or pass with a D final grade. Theoretically, students can chart out a minimum number of passes on specific course dimensions to "dial in" their desired final grade. Of course, thanks to the token system there can always be another chance to pass critical items and bulk up weaker areas. In a world where even before the pandemic, I was struggling to get my students to calculate a simple fraction to know their grade in the course, daily tracking of progress is very uncomplicated with a tracking sheet attached to the last page of the syllabus (Figure 3). Lastly, the final grade bundling table establishes an absolute metric for final grades and performance in the course. No more haggling over decimal places, rounding, or looking for extra credit. If a student earns only five (5) take-home problem set badges in the course, the best

CHM 2270-01 F23 FINAL GRADE TRACKER

Item	Learning Checks																		
	1A	1B	1C	2A	2B	2C	3A	3B	3C	3D	4A	4B	4C	5A	5B	5C	6A	6B	6C
✓ or ×																			

Item	Notes/Organic Data Bank (ODB)					
	1	2	3	4	5	6
✓ or ×						

Item	Achieve Online Homework									
	Training: Drawing Tool	Training: Curved Arrows	Training: Stereochemistry	Nomen-clature	1	2	3	4	5	6
✓ or ×										

Item	Problem Day
	once
✓ or ×	

Item	Take-Home Problem Set Badges						
	1	2	3	4	5	6	Final Exam
✓ or ×							

Item	Tokens					
	Start of Term	Recitations Attended (×⅓)	RQuizzes min. 2 out of 3 (×⅓)	Events Reflected Upon (×¼)	Office Hour Attended In-Person (during first two weeks)	Miscellaneous
#	1					

FIGURE 3
The final grade tracker for the organic chemistry I lecture course.

grade they can earn is a B+. In fact, there is still a little nuance in a final grade bundling table, as different criteria can give different final grades that can be minimized, maximized, or averaged for reported final grades. The specific number chosen for each assessment type is meant to align with a specific fundamental academic skill: learning checks (attendance for engagement), notes/ODB (organic data bank, record-keeping), Achieve and Problem Day (exercises for practice), and THPS badges (summative assessment through collaboration).

2.5 Communicating expectations

During the first iteration of the alternatively graded lecture course, a student information survey on the first day of class showed that zero students out of 75 identified with being a part of an alternatively-graded course in the past. In each consequent iteration, the total number of students who claimed to understand alternative grading never crested higher than 5% of total enrollment. Following a lead from the literature (McGuire, 2015), a large amount of time during the first week of classes was spent training the students in three (3) major areas: (1) the pedagogical principles upon which the course was based, (2) logistics of how to navigate the course, and (3) the general approach of metacognitive learning. In all iterations of the alternatively graded lecture course since the first, these simple investments did a great deal to quell student uneasiness and quickly prove the heart of the course was learning, not grades. Since there is no one recipe for how to change one's teaching (Supiano, 2023) (and even perfectly executed best practices can easily backfire), I decided to vulnerably convey my own anxiety about these relatively untested methods by designing a graphic to guide students through the cycles of the course (Figure 4). On it, I lay bare the "right" and "wrong" ways to proceed through the course, harshly detailing ways to circumvent course policies and therefore, deep learning. In short, my pledge to students was to meet them where they are (on screens, via video content introductions) and to gamify (owing to their love of social media and video games) the course. Anecdotally, students mentioned that this

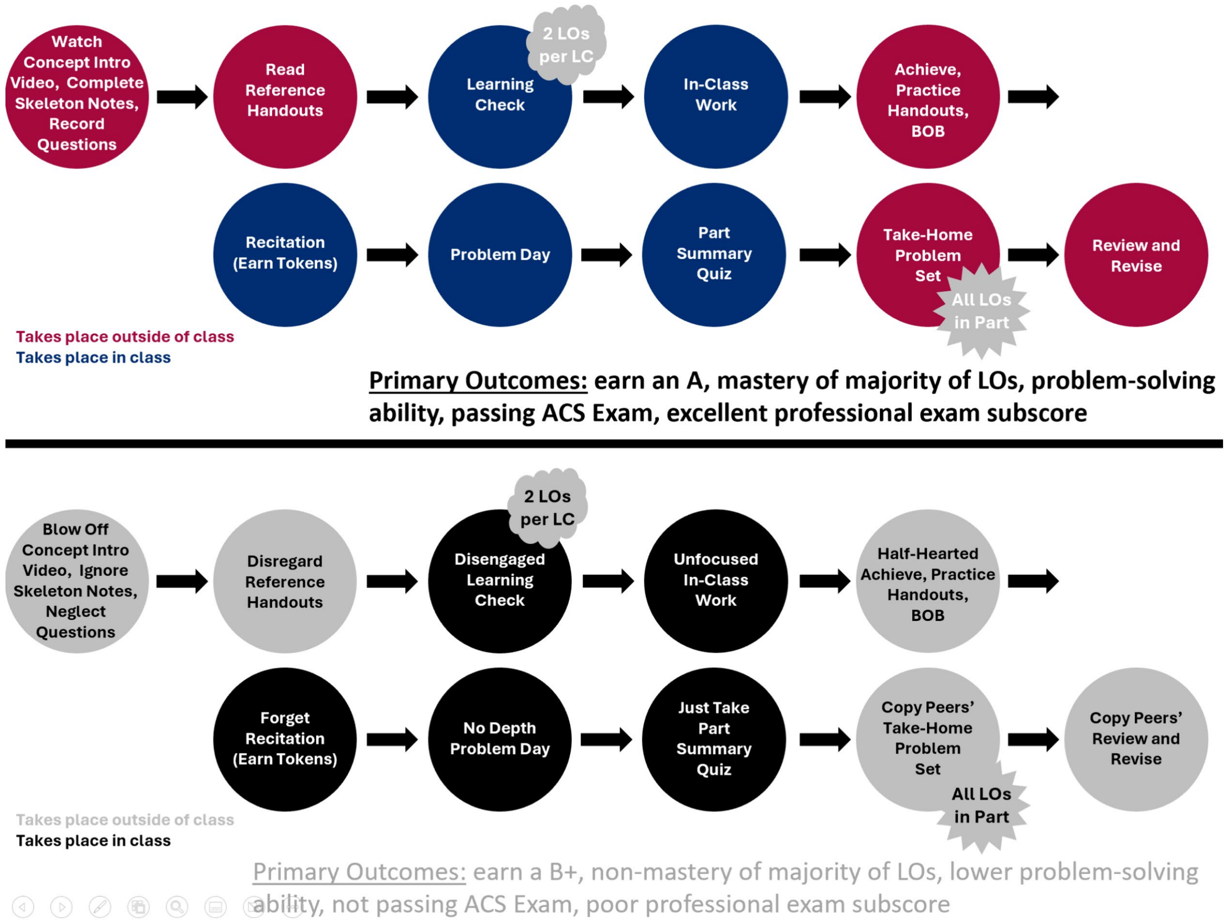


FIGURE 4
Graphic guide to the organic chemistry I lecture course.

may have been the first course they took that cared enough to walk them through the process and expectations that undergirded their learning. I learned that it can be a common mistake to assume your students know even the basics of how to learn – showing up, taking notes, reviewing content, practicing with exercises, taking breaks, appropriate collaboration. Each of these skills is individually incentivized in the alternatively graded system, emphasizing skills beyond those of basic organic chemistry.

The key here is to encourage both a growth mindset and student agency in discovering that mindset for themselves (Torres, 2023). Students need to be placed into environments where they feel safe enough to take calculated risks, especially where livelihoods and lives are not yet on the line. When environmental conditions are changed, students' struggles with new material can be normalized or overcome. We can also ask ourselves if we assign failure as a natural part of revision and, ultimately, learning? Students all too often equate performance with identity and assume that most skills are innate. Since potential can be cultivated through different means, various methods of demonstrating mastery of learning outcomes should be designed for students. Additionally, student work should be responded to with the twin goals of affirmation and challenge. Most educators agree that work falls to the student to do the labor of learning. Instructors, however, can see from the 30,000-foot view and can parse out content and avoid feedback overload. Finally, we need to ask ourselves if students have the chance to digest, interpret, and apply the feedback we provide in our courses. Appropriate time for reflection and discernment is needed when moving from smaller, lower-stakes work to larger, higher-stakes assessment. Alternative grading removes many of the barriers to fulfilling these ideologies and casts the instructor-student relationship in positive, outcome-centered, and self-determining light.

At the very end of the first iteration of the course, I was very pleased to find that my original goal had been achieved: ZERO students contacted me after final grades to discuss, barter, or complain about what they had earned. After all, they were directly interacting with the principles of alternative grading from the first day of class to the conclusion of the cumulative final exam. I observed that students were in touch with their final grades on a nearly day-to-day basis for the entirety of the 16-week course.

2.6 Planning the lab course – organic chemistry I

In the same semester I experimented with alternative grading strategies in my Organic Chemistry I lecture course, I did the same with my Organic Chemistry I laboratory course. This one (1)-credit class would have 24 total enrollees, mostly 2nd- and 3rd year students, and would meet for 180 min once a week, plus a 50 min recitation that preceded the lab session. Because of previous tweaks to the AGS format, and perhaps also due to the more adaptable nature of practical laboratory courses, less overhaul was required. I had already designed a successful CURE (course-based undergraduate research experience) surrounding legal cannabinoids in over-the-counter products (Mio, 2022). Most scientific educators know that laboratory instruction takes a very different tack than theory courses, and in many cases, a much larger amount of prework. The alternative grading methods used for the lab course would have to follow the Four Pillars mentioned previously, in conjunction with a few corollaries: maximization of time in the lab and out-of-lab reflection, as well as ample time for engagement when lab is not center of mind (at Detroit Mercy, a one-hour lab recitation session occurs right before a three-hour lab session). All specific learning outcomes were arranged around the safe handling and bench chemistry of volatile organic compounds.

For the lab course, the focus became (a) what students will do to prepare for the lab session, (b) what students will accomplish during the lab session, and (c) what students will do to follow-up from the lab session. Weekly prep work involved a lab notebook setup, along with CURE-based research into both materials and techniques. Elements of safety were always part of prep, in addition to short videos showing students performing the technique to be practiced. Upon arriving in lab, student pre-work was checked in and partners were assigned. Questioning all aspects of the bench work was encouraged by both instructor and teaching assistant (TA) throughout the session, and a quick partner evaluation was filled out at the end upon exit. After lab, a certain number of days were given for post-lab questions to be responded to, where this task "unlocked" a follow-up assessment on both the theory and the practice of the week's experiment, including simple distillation, thin-layer chromatography (TLC), solubility, extraction, and instrumental (GC–MS, FT-IR, NMR) analysis methods. Any missed or less-than-satisfactory aspect of the course resulted in a call for discussion with the instructor within 48 h. A literature research project with weekly objectives brought about a group presentation at the end of the semester describing both data and social implications of the work. In short, as long as students showed up and completed their assigned work in good faith and a timely manner, they were rewarded with positive feedback. Missed items and unsatisfactory work, as measured against a "minimum expectations" rubric, caused short-term discussion and chances to redo the work. While preliminary outcomes will be discussed in the section of this work, students from two (2) iterations of this alternatively-graded lab course anecdotally reported they experienced reduced pressure from not taking quizzes/exams, they found lab sessions more enjoyable because they felt more prepared, and also thought their learning went deeper with more focused reflection outside of lab.

3 Discussion – preliminary outcomes

Considering the timetable for *innovation* in alternative grading has been relatively short, many educational researchers have published their *findings* for over 30 years (Clark, n.d.). Qualification and quantification of data in this area will serve to legitimize the field and attract more educators as the benefits of alternative grading are explored from pre-elementary all the way up to the graduate level.

Active scholars in the field have clustered four (4) student-centered and positive results of courses run with alternatively-graded activities or completely formatting under the principles of alternative grading: (Clark, n.d.) (1) task-related feedback encourages students' intrinsic motivation and can improve performance; (2) students feel and exhibit less anxiety and more risk-taking; (3) instances of student academic misconduct are less likely; and (4) student who transition from AGS can do well in later AGS or traditional courses. In short, alternatively graded courses in higher education are not harmful; and in fact, students report that the structure of such courses is beneficial to many aspects of their learning in the course.

In my few iterations of alternatively graded courses, I have found all of these initial findings to be true and have discovered many instructor-specific benefits along the way. While it is true that course redesign and preparation are always time-consuming on the order of weeks, the investment pays off greatly in day-to-day activities taking far less effort and nimbleness raised to its highest boundary. Since all activities of the course are fenced by the learning outcomes phrased as direct tasks, the writing of daily/unit assessments can morph around the specific strengths and weaknesses of the current cohort of students. In addition, more meaningful feedback can be generated more quickly because the assumed rubric of grading is "Did the student accomplish the task?"

In general, I have found that shifting student attitudes away from final grade and toward genuine learning is a benefit that nearly all educators would embrace. Again, I can report that this has occurred in my AGS courses in both day-to-day and overarching conversations. There is a sense in my students, especially after the course structure is unveiled on the first few days of classes that having options for what to complete in a course does not equate to setting lower learning goals. As anecdotal evidence, in the Fall 2022 and 2023 versions of the Organic Chemistry I course, nearly all students revised all of their incorrect responses on Take-Home Problem Sets, even the ones not required to earn the "pass." I have witnessed that student self-agency has overall increased in my AGS courses, and anecdotally, students report on end-of-term evaluations that they appreciate being put "in the driver seat" of their learning, not just being told what to do.

Finally, I have discovered that in writing letters of recommendation for former students who have shared AGS courses with me, it is a far more straightforward endeavor. Each task in the lecture course, for example, is paired with a chief, non-organic chemistry learning goal: daily learning checks (keeping up with content), submission of summarized notes (ability to condense major topics), online homework (exercises to extend skills), extemporaneous problem-solving presentations (spontaneous exercise work), and biweekly, skills-based assessments (summative evaluation). Alignment of student work to these metrics allowed for a very simple structure to letters of recommendation that includes many of the non-content skills employers and graduate schools are looking for in undergraduate students. Speaking of final grades, detractors of AGS state that too many high grades will be assigned using these principles. In short, I have found that the number of A and B grades increases with the application of AGS. However, there may be a fallacy conflated in statements like, "massive grade inflation is a corruption of standards." I have found that it is a more realistic view to distinguish between negative grade inflation, or unjustifiably high grades, and positive grade inflation, or more effective pedagogies resulting in higher achievement for more students. I find that I can now effortlessly focus on what my students can and cannot do, down to the number of times it takes them to "pass" an individual learning outcome. In fact, I am beginning to sense that there may be no difference between a student able to pass a learning outcome on the first versus more numerous tries. Accommodating more than one (1) learning preference to a course is always a supreme challenge for the educator and AGS provides one very beneficial way to accomplish that.

3.1 Advice for those starting an alternative grading journey

The early returns on alternative grading delivering on promises of enhanced student learning and equity are strong. However, decades of traditional pedagogical methods will not submit overnight. In addition, the best laid plans often do not work out the way we educators wish they would or intend them to. Herein lies some advice for those looking to take the first steps in their AGS journey.

3.1.1 Matters of scale matter, both in the size of course enrollment and content

It is theoretically possible to incorporate AGS into an entire course, one unit of a course, one lesson of a course, or simply one class period. I have done this for smaller (24 students) and larger (116 students) courses. Moreso than traditional methods, I find AGS balances out in these enrollment realms with little cause for overhaul. I have also been fascinated by experienced faculty's response to which of these is the better place to start. Some think that the only way to begin anew is to cast off all course logistics. Others state that starting small is always the lowest risk. I agree with both statements. As long as you are learning outcome-focused, an entire course or one class period can accommodate change. I would rephrase this as: Do something – anything – and tie its assessment to a discrete learning outcome. The two-point rubric is quite freeing – can student do it or not? What do they need to do it? How can you, as an educator, support that need?

3.1.2 Building the plane while flying it is OK

Welcoming students to your experiment can have a calming effect on all involved—we do not know what will happen next, but we are excited about it! Interestingly, this was the advice given to me as I started my academic career with regard to laboratory research. As an organic chemist, I reflect often on the comment "Why are we so eager to experiment in the lab, but not in the classroom?" Any amount of effort could reap benefits in ways difficult to visualize at the outset. Critics may bring up the tyranny of content and the fact that too much gear-shifting can disorient students. A more positive take on these common critiques might be that exposing students to many different methods of instruction and assessment can bolster their learning toolbox. Ungrading and metacognitive instruction are the low-hanging fruit that can instantly transform instruction. Because all disciplines are distinct in what pedagogy works best for learning, and taking into account the diversity of learners swells the possibilities to incalculable numbers, we educators must simplify, simplify, simplify. Students learn lessons, then can practice to earn feedback unfettered by grading (ungrading). In the same vein, walking students through a "cycle" of the course and how to think about their own thinking (metacognition) in the discipline proves our dedication to reflection on learning, a skill we wish to incorporate in them.

3.1.3 As a cornerstone principle, pledge to decouple learning from fear of evaluation

Post-Covid students in American higher education have endured global pandemic, massive social and political upheaval, the advent of social media and AI, all in an era of near-instant information sharing over less than 10 years. Our best students will recognize that in considering alternative grading, their educators are concerned for both their learning and mental health. When educators attempt to use anxiety as motivation with the current generation of college students, academic performance diminishes. Alternative grading strategies demonstrate, from first principles, that we care about a student's learning and academic success. Paraphrased by the writer and civil rights activist Maya Angelou, "I've learned that people will forget what you have said, people will forget what you did, but they will never forget how you made them feel."

4 Conclusion

I hope that my writing has served to inspire you to pursue a path of alternative grading for your courses, organic chemistry or otherwise. The benefits of the pedagogy far outweigh the uncertainty and time involved in converting aspects of your course, or complete courses. With traditional lecturing hundreds of years old and ABCDF grading still relatively young, time will afford us little chance to await an educational sea change as substantial as the Covid-19 pandemic. I, like many members of the Academy, think that the global pandemic only accelerated and exacerbated issues that were already fomenting. We know our chosen disciplines so very well. The question has always been: How can we help students embrace the confusion of learning a brand-new set of concepts? Alternative grading, with its student-centered goal-setting, timely and applicable feedback, intent of clear learning outcome achievement, and reassessment of revised work both meets and exceeds the needs of a new generation for academic evaluation. The winds have already changed, and the time has come to walk down a different road on the journey of teaching and learning.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

MM: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The author would like to thank his colleagues M. R. Livezey and M. Abdel Latif for inspiring and insightful conversations on the many topics of alternative grading. I would also like to thank my students for always being the best motivators and idea-machines for better pedagogy.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anderson, L. W., and Krathwohl, D. R. (Eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston, MA (Pearson Education Group): Allyn & Bacon.

Bergmann, J., and Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Washington, DC: International Society for Technology in Education.

Bergmann, J., and Sams, A. (2014). Flipping for mastery. *Educ. Leadersh.* 71, 24–29.

Bloom, B. S., and Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook I: Cognitive domain*. NY: Longmans, Green.

Blum, S. D. (2020). *Ungrading: why rating students undermines learning (and what to do instead)*. Morgantown, WV: West Virginia University Press.

Brunsell, E. Designing science inquiry: claim + evidence + reasoning = explanation. Edutopia, Available at: https://www.edutopia.org/blog/science-inquiry-claim-evidence-reasoning-eric-brunsell (Accessed March 1, 2024).

Campbell, D. T. (1979). Assessing the impact of planned social change. *Eval. Program Plann. 2*, 67–90. doi: 10.1016/0149-7189(79)90048-X

Clark, D. Bibliography of alternative grading research. Available at: https://docs.google.com/document/d/1zCpoFgrzN2JUXyTfAm5xWdfLeLWt8P6xu8oqySZflYc/edit#heading=h.34m439eanynn (Accessed March 1, 2024).

Clark, D. Four key research results: a summary of some important research on alternative grading. Available at: https://gradingforgrowth.com/p/four-key-research-results?utm_source=%2Fsearch%2Fbibliography&utm_medium=reader2 (Accessed March 1, 2024).

Clark, D., and Talbert, R. (2023). *Grading for growth*. New York and London: Routledge.

Dixson, L., Pomales, B., Hashemzadeh, M., and Hashemzadeh, M. (2022). Is organic chemistry helpful for basic understanding of disease and medical education? *J. Chem. Educ. 99*, 688–693. doi: 10.1021/acs.jchemed.1c00772

Draeger, J., del Prado Hill, P., Hunter, L. R., and Mahler, R. (2013). The anatomy of academic rigor: the story of one institutional journey. *Innov. High. Educ. 38*, 267–279. doi: 10.1007/s10755-012-9246-8

Draeger, J., del Prado Hill, P., and Mahler, R. (2015). Developing a student conception of academic rigor. *Innov. High. Educ.* 40, 215–228. doi: 10.1007/s10755-014-9308-1

Feldman, J. (2018). *Grading for equity: what it is, why it matters, and how it can transform schools and classrooms*. Thousand Oaks, CA: Corwin, 23.

Goodhart, C. (1975). "Problems of Monetary Management: the U.K. Experience" in *Inflation, depression, and economic policy in the west*. ed. A. S. Courakis (Totowa, New Jersey: Barnes and Noble Books), 116.

Houchlei, S. K., Crandell, O. M., and Cooper, M. M. (2023). "What about the students who switched course type?": an investigation of inconsistent course experience. *J. Chem. Educ. 100*, 4212–4223. doi: 10.1021/acs.jchemed.3c00345

Ko, A. (2019). Grading is ineffective, harmful, and unjust—let's stop doing it. *Medium* 16. Available at: Medium.com.

Louisiana State University Center for Academic Success. Available at: https://www.lsu.edu/cas/ (Accessed March 1, 2024).

McGuire, S. Y. (2015). *Teach students how to learn*. New York and London: Routledge.

McMurtrie, B. (2023). Students crossing boundaries: rudeness, disruptions, unrealistic demands. Where to draw the line? *Chron. High. Educ.* 14.

Mio, M. J. (2022). "Second-year, course-based undergraduate research experience (CURE) in the organic chemistry of cannabinoids" in *Abstract of papers, ACS fall 2022 (American Chemical Society)* (Washington, DC: American Chemical Society).

Nelson, C. E. (2010). Dysfunctional illusions of rigor. *Improve Acad.* 28, 177–192. doi: 10.1002/j.2334-4822.2010.tb00602.x

Nilson, L. B. (2014). *Specifications grading*. New York and London: Stylus Publishing.

Oleson, A., and Hora, M. T. (2014). Teaching the way they were taught? Revisiting the sources of teaching knowledge and the role of prior experience in shaping faculty teaching practices. *High. Educ. 68*, 29–45. doi: 10.1007/s10734-013-9678-9

Perry, W. G. Jr. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart, and Winston.

Supiano, B. (2022). What does it mean when students can't pass your course? *Chron. High. Educ.* 6.

Supiano, B. (2023). Changing your teaching takes more than a recipe. *Chron. High. Educ.* 9.

Talbert, R., and Clark, D., Available at: gradingforgrowth.com (Accessed March 1, 2024).

Torres, J. T. (2023). 4 questions to ask to promote student learning. *Inside High. Ed.* (Accessed March 7, 2023).

Townsley, M. A comprehensive list of standards-based grading articles. Available at: http://mctownsley.net/a-comprehensive-list-of-standards-based-grading-articles/ (Accessed March 1, 2024).

University of Detroit Mercy Mission and Vision. Available at: https://www.udmercy.edu/about/mission-vision (Accessed March 1, 2024).

frontiers | Frontiers in Education

TYPE Curriculum, Instruction, and Pedagogy
PUBLISHED 06 June 2024
DOI 10.3389/feduc.2024.1402599

Check for updates

OPEN ACCESS

EDITED BY
Brett McCollum,
Thompson Rivers University, Canada

REVIEWED BY
Jessica D'Eon,
University of Toronto, Canada
W. Stephen McNeil,
University of British Columbia, Okanagan Campus, Canada

*CORRESPONDENCE
Stephen L. MacNeil
smacneil@wlu.ca

RECEIVED 17 March 2024
ACCEPTED 01 May 2024
PUBLISHED 06 June 2024

CITATION
MacNeil SL, Wood E and Arslantas F (2024) Development of a metacognition co-curriculum for a university course in introductory organic chemistry. *Front. Educ.* 9:1402599.
doi: 10.3389/feduc.2024.1402599

Development of a metacognition co-curriculum for a university course in introductory organic chemistry

Stephen L. MacNeil[1]*, Eileen Wood[2] and Fatma Arslantas[2]

[1]Department of Chemistry & Biochemistry, Wilfrid Laurier University, Waterloo, ON, Canada,
[2]Department of Psychology, Wilfrid Laurier University, Waterloo, ON, Canada

Metacognition is a fundamental skill that allows advanced learners to adapt to diverse learning environments. Metacognition, however, can be domain specific and students may fail to generalize metacognitive skills across domains. Thus, students in higher education may require specific training to acquire relevant metacognitive skills in differing domains or may need cueing to engage their metacognitive skills and knowledge in new domains. The present report describes the development of a co-curricular metacognitive program for chemistry students and suggests how this program could be adopted by other chemistry courses or adapted for other domains in higher education. Several supports were introduced in this program including self-assessment of competence with learning task inventories (LTIs; i.e., detailed lists of learning tasks), self-assessments of confidence regarding in-class content questions, and performance predictions and postdictions on tests. In general, exposure to these supports resulted in overall performance and confidence gains. However, individual differences were evident with some students demonstrating greater learning gains than others. Initial Dunning-Kruger effects associated with pre- and postdictions, with low-performing students overestimating grades and high-performing students underestimating grades, decreased over exposure. A summary of the evolution of this metacognitive co-curricular program, the educational literature that steered it, and the differential impact on students is explained.

KEYWORDS

metacognition, organic chemistry, discipline-based education research, higher education, prediction and postdiction, scaffolded learning

Introduction

Metacognition

Metacognition is a multifaceted, fundamental skill that allows advanced learners to adapt to diverse learning environments effectively and efficiently. Metacognition is the mechanism through which adult learners control their cognitive processes through planning, monitoring, evaluating, and regulating their own learning (Flavell, 1979; Dimmit and McCormick, 2012; Rivas et al., 2022). In addition to cognitive control, three key components of metacognitive knowledge include declarative knowledge (i.e., awareness of what you do and do not know in a particular domain); procedural knowledge (i.e., a repertoire of strategies to utilize for

 10.3389/feduc.2024.1402599

different learning demands), and conditional knowledge (i.e., the ability to use available knowledge, strategies, and other tools when needed; Schraw et al., 2006). Successful learning is evidenced when metacognitive skills are used to identify, map approaches to tackle, and monitor the process and outcomes of a learning task, and, at steps throughout the process, engage in self-reflection and self-assessment regarding decisions made, progress occurring, and outcomes achieved. Typically, more successful students are better able to calibrate their learning and performance (Dimmit and McCormick, 2012; Saks et al., 2021).

In today's higher education classrooms, instructional approaches nested in student-centered learning, constructivist pedagogies, and flexed/hybrid formats draw heavily on students' metacognitive skills to drive their own learning. Students must regularly assess what is new, unknown, or poorly understood to regulate their reading, studying, and preparation within these contexts. Although advanced learners in higher education contexts typically have acquired generalized metacognitive skills that prepare them to approach a broad range of new learning tasks (Geurten et al., 2018), learners may fail to draw upon these skills. More precisely, in some circumstances, application of metacognitive skills differs across domains. This domain specificity means that students are less able or likely to apply their metacognitive skills across domains of study, especially when the domain or tasks are perceived to be more difficult (Scott and Berman, 2013). In addition, some advanced learners may not have the scope of strategic repertoires to engage as independently as expected in today's classrooms (García-Pérez et al., 2021). As a result, students may need explicit instruction or prompts to scaffold and encourage use of metacognitive skills in new or challenging domains.

Effective instructional design acknowledges this student need and necessitates the development of supports to teach or scaffold activation of metacognitive skills. This paper describes the translation of research on metacognition and its role in the learning process through the development of a co-curricular metacognitive program to facilitate learning in a second-year organic chemistry course. We outline both the translation process and the effects of components of this metacognitive program on students' metacognitive skills and performance in the course.

Key to this program were course elements designed to enhance planning, monitoring, and reflection skills as well as explicit instruction regarding metacognition. When students are presented with tasks, they first need to assess whether the task is familiar. They must ask themselves, "Can I recall doing this type of task or a similar task before?" Memory can then serve as a guide for *planning* (i.e., what, when, and how to read, engage in study strategies, and organize learning priorities) depending on the learners' assessment of what they do or do not know about the topic or task. Many types of monitoring activities then follow this initial memory task, for example, checking and assessing performance as a task unfolds. This may require a single simple check for easier tasks but may be an iterative process for challenging tasks. To draw on an organic chemistry example, consider the assignment of *R* vs. *S* stereochemistry to a chirality center in an organic molecule. The student must first assess their readiness for this task, e.g., "Have I done this previously?" "Can I recall the steps involved?," and "Are there particular steps I have struggled with in the past?." After recalling or reviewing steps required, the student must then execute each step and monitor progress. Have they properly assigned priorities to substituents at the chirality center using the Cahn-Ingold-Prelog rules? Is the molecule being viewed from the proper angle? If not, the student must decide on a mental or physical manipulation to reorient the molecule and judge the effectiveness of that manipulation. Finally, the student must assess the direction in which the remaining substituents decrease in priority and, subsequently, assign *R* or *S* stereochemistry. In addition to monitoring the process, the learner must monitor the outcomes. Did the student assign stereochemistry correctly? More precisely, effective monitoring requires that corrections are made after errors are detected. If a mistake in assigning stereochemistry was made, what was the problem? Were substituent priorities assigned correctly? Was the mental/physical manipulation done properly? Was the correct label – *R* vs. *S* – assigned based on the direction of decreasing priorities?

Successful planning and monitoring presume that learners are motivated to learn and have sufficient time to learn. Both motivation and executive processing can enhance use of metacognitive skills (Rivas et al., 2022). For example, a highly motivated, self-regulated learner is more likely to schedule study time in advance, arrange an environment conducive to studying, complete assignments, explore additional examples, read, review and summarize, and engage in diligent assessment of what has or has not been learned (Heikkilä and Lonka, 2006). Less motivated and more challenged students require cues or prompts to scaffold these steps and keep them on task. Metacognitive reflection is an effective learning tool (e.g., Bangert-Downs et al., 2004; Dignath and Büttner, 2008) that involves thinking about what, how, and why one does what one does. Through engaging in critical evaluation learners gain new insights and perspectives (e.g., Grimmett and MacKinnon, 1992). To encourage reflection the learning environment must provide opportunities for learners to 'take stock' of their own approach to learning.

Why metacognition in organic chemistry?

Given the diversity in metacognitive skills among university students and the potential for domain specificity in application of these skills, students in higher education may require specific training, scaffolding, or cueing to acquire or transfer relevant metacognitive skills across differing domains (Zohar and Dori, 2012). Introductory organic chemistry is typically a prerequisite course for subsequent chemistry and other science studies. As such it is both desirable and feared. Consequently, significant research has been directed toward learning and instruction in organic chemistry (e.g., Kranz et al., 2023; Pilcher et al., 2023). The present paper summarizes steps toward establishing course-specific metacognitive training.

What prompted the development of a metacognition co-curriculum for introductory organic chemistry?

Inspired by the potential inherent in blended learning designs (Garrison and Vaughan, 2008), and tenets of adult metacognition, the traditional lecture-based course in introductory organic chemistry was transformed to a blended learning format. As part of the re-design process comprehensive lists of chapter-by-chapter learning tasks were created with specific low-level tasks, e.g., "*define* chirality centre" (Blooms revised taxonomy level 1; Krathwohl, 2002), identified as those to be completed independently by learners before class, and other higher-level tasks, e.g., "given the structure of an organic compound, *identify* all chirality centers" (Blooms revised taxonomy level 4), demarcated as those to be completed, with assistance from other students and the instructor as needed, in class. These task lists served as an organizational scaffold that clearly identified the knowledge and skills required for success. When the blended learning course was launched, these learning task lists were posted as PDFs via the online course management system. Consistent with self-directed models of adult learning (Merriam, 2001), and metacognitive monitoring it was expected that students would access and use this resource to plan and evaluate their progression through the term. However, on average, only 39% of students accessed the PDFs across the term. Given the novelty of these learning task lists, we decided to encourage students to access and review the lists and recognize them as a support for learning and not just extra supplementary materials. In the next course offering, the PDFs were converted to surveys delivered through the course management system and students were instructed to engage in self-assessment of knowledge by indicating on a 5-point scale how well they could do each learning task before they could access online resources for the next chapter. Restricting access to the next week's content improved participation (> 90% across the term), but students' thoughtful engagement with the learning tasks remained problematic. For example, when the following item was added to a learning tasks survey, "Select 1 if you are reading this," only ~50% of students selected 1! This clearly indicated that many students were not engaging with the surveys as anticipated and many were arbitrarily responding just to gain access to the next week's content. It was at this point that a partnership formed between the instructor (an organic chemist) and an educational/developmental psychologist who introduced the concept of metacognition and speculated that the failures of students to access and engage with the provided resources may stem from metacognitive failure. This led to a discipline-based educational research (DBER) project aimed at improving students' metacognitive skills starting with an examination of engagement with the learning task lists. The following narrative is a retroactive summary of the DBER project as it unfolded across four consecutive Fall term iterations of a single section, in-person, introductory organic chemistry course (Organic Chemistry I) taught by the same instructor.

Motivations, methods, and results to date

Phase I

The first step in the DBER project was to determine optimal conditions for having students engage with the LTIs. In total, 293 students of the 311 students registered in the Organic Chemistry I course agreed to participate and were randomly assigned to five treatment groups, each of which was provided nine weekly LTIs across the term and, at the end of the week, asked to rate their ability to complete each task using the same 5-point mastery scale noted above. Students in Condition 1 self-assessed domain knowledge by completing the LTI ratings. To gage the fidelity of the LTI rating measure, students in Condition 2 completed the LTI ratings including a question imbedded in the LTI that assessed how carefully students were completing the ratings (e.g., "Are you reading each task carefully or clicking away without doing so?"). We then added three additional conditions to assess students' performance when tested on their knowledge and provided with varying levels of feedback. To evaluate the effect of testing (Roediger and Karpicke, 2006) and feedback (Pashler et al., 2005; Fazio et al., 2010), participants in Conditions 3–5 completed the LTI ratings and fidelity measure, followed by a 5-question multiple choice quiz. This permitted us to compare perceived mastery to actual performance for five learning tasks per LTI. Participants in Condition 3 received no feedback on their quiz performance; those in Condition 4 were provided the correct answers with no explanations after each quiz; and those in Condition 5 were provided the correct answers with full explanations after each quiz. These students also responded to a 10-item post-quiz survey assessing students' perceptions of quiz difficulty, comparison of LTI ratings to quiz performance, and changes in quiz difficulty, content covered, and student engagement and interest from week to week (see Supplementary material). All students also completed an end-of-term survey to ascertain reactions to the LTIs and provide general chemistry grades and GPAs as measures of prior learning.

The major findings were that (i) treatment condition did not affect final exam grades but (ii) the number of LTIs completed did, with completion of more LTIs leading to higher final exam grades (See Table 2 in MacNeil et al. (2013)). The effect of number of LTIs completed was over and above that which could be explained by differences in prior learning. In addition, 72% of students indicated they would recommend using LTIs in future offerings of the course and most participants felt that LTIs improved awareness of learning tasks they could not do. However, when asked about the impact of using the LTIs on changing study habits or improving grades, most students attributed only a small (56.6%) or no (28.1%) impact on final grades and very few students reported any impact on study habits, with 80.4% indicating no impact. Overall, these results confirmed the value of the LTIs as a tool for supporting learning but clearly students did not perceive differences in metacognitive skills associated with planning, monitoring or evaluation.

Phase II

Given the documented effects of testing (Roediger and Karpicke, 2006) and type of feedback (Pashler et al., 2005; Fazio et al., 2010) on learning, it was surprising to see in Phase I that treatment condition had no effect on final exam grades, but the overall positive effect of number of LTIs completed on final exam grades was encouraging. This result prompted us to assess whether completion of LTIs also improved metacognitive skills more explicitly. Consequently, 211 students of the 310 students registered in the next offering of Organic Chemistry I participated in Phase II of the DBER project. To measure metacognitive skills directly, students completed a metacognitive awareness assessment scale (Schraw and Dennison, 1994) at the beginning and a condensed version at the end of the term. Given the documented benefits of priming (Ratcliff et al., 1997) and distributed practice (Benjamin and Tullis, 2010) on learning and performance, we also wanted to test the effects of 'priming' and timing of LTIs on participants' metacognitive skills. Thus, four treatment conditions were employed. Participants in Conditions 1 and 2 completed weekly LTIs (distributed practice) as in Phase I, but those in Condition 2 were primed with a list of LTI items at the beginning of each week posted as a PDF on the course management system. These listed items served as cues to prime students for the full LTIs that would be required later in the week. Participants in Conditions 3 and 4 completed only two aggregated LTIs (massed practice), one during the week of the midterm test and one during the week of the final exam, with those in Condition 4, like Condition 2, primed to these aggregated LTIs via the corresponding aggregated inventory cue posted to the course management system at the beginning of the week. The "Group 5" LTI conditions from Phase I were employed, i.e., completion of the basic LTI with metacognitive prompts, e.g., "As you read each learning task, do you think of it as a possible question that you might be tested on?," to test the reliability of students' mastery ratings, followed by a quiz with full feedback and post-quiz survey, with one exception – instead of rating their mastery for individual learning tasks, participants rated their abilities for groups of 6 learning tasks so as to reduce the likelihood of survey fatigue. Surprisingly, there were no differences across condition (primed versus not primed, or massed versus distributed) with respect to number of LTIs completed or exam performance, nor were there differences across the term for the three measures of metacognitive awareness: knowledge of cognition, regulation of cognition, and overall metacognitive awareness.

Phase III

Given the lack of observed differences from priming and distributed vs. massed practice with respect to students' self-reported metacognitive skills across the term in Phase II, we decided to make metacognition a more explicit part of the course. Specifically, we made five changes for Phase III including (i) introduction of a lecture on metacognition (ii) reversion to weekly LTIs based on the previously observed finding that completion of more LTIs lead to higher final exam grades, (iii) addition of a weekly in-class question at the beginning of each week to assess how students interacted with the priming inventory cues, (iv) expansion of the end-of-term metacognitive awareness assessment to include *all* items used in the beginning-of-term metacognitive awareness assessment (note that only about *half* of the items were used at the end-of-term in Phase II), and (v) addition of more metacognitive practice activities in the form of midterm test and final exam grade predictions and postdictions.

Consistent with literature regarding effective metacognitive training (Cook et al., 2013), the addition of the lecture on metacognition explicitly cued students to use metacognitive skills (planning, monitoring, self-assessment, etc.) when engaged in their chemistry course. The addition of a weekly in-class question at the beginning of each week to assess how students interacted with the inventory cues served to prime and encourage self-reflection and assessment. Thus, each week at the beginning of the first class, participants were asked via the iClicker® personal response system "What did you do with the Chapter 'X' inventory cue posted to the 'Content' section of the CH202/204 MyLS page?" Possible answers were: A. I did not know, or I forgot it was available. B. I know it was available, but I did not access it. C. I skimmed it briefly. D. I read it carefully, and E. I read it carefully and prepared a physical/mental check list of what I could/could not do. This priming was expected to lead to improved engagement with the inventory cues at the beginning of the week and, consequently, with the LTIs at the end of each week. It was anticipated that this priming and improved engagement should, in turn, lead to improved use of metacognitive skills and performance. In total, 238 students of the 296 registered in the next offering of Organic Chemistry I participated in Phase III. Across the term, students accessed the inventory cues 37.4% of the time, indicating that the inventory cues prompted students to preview the material only some of the time. Analyses indicated that "skimmed" (response C) was the most frequently endorsed choice among "middle" (final course grade: 60–79.99%; 39.1% endorsing) and "bottom" (final course grade: 9.99–59.99%; 39.1% endorsing) performers, and 'no access' (response B) was the most frequently endorsed choice among "top" (final course grade: 80–100%; 38.9% endorsing) performers (See Supplementary Information for a more complete summary of select analyses for Phase III). Although effective use of the inventory cues decreased across the term (See Supplementary Information Phase III summary), regression analysis revealed that the number of inventory cues accessed predicted the number of LTIs completed ($t_{228} = 3.292$, $p = 0.001$) which, in turn, had a significant effect on final course grades, with participants completing more LTIs earning higher course grades ($t_{228} = 2.72$, $p = 0.007$). Noteworthy is that the effect size for LTIs was larger than for all three prior learning variables (Chemistry 1 and 2 grades and overall GPA) (See Supplementary Information Phase III summary). Unfortunately, access to inventory cues had no effect on participants' self-reported metacognitive skills.

Given the importance of self-evaluation as a metacognitive skill (Hacker et al., 2000), one group of participants (Group A) was asked to make midterm test and final exam grade predictions immediately after each test/exam, i.e., postdictions. A second group of participants (Group B) was asked to make midterm test and final exam grade predictions immediately before (same day as) and immediately after completing each test/exam. Finally, a third group of participants (Group C) was asked to make midterm test and final exam grade predictions 2 weeks before, immediately before (same day as), and immediately after completing each test/exam. These test/exam grade predictions and postdictions proved important to students' metacognitive skills in that a Dunning Kruger effect (Kruger and Dunning, 1999) was observed for all pre-and postdictions, with the lower achieving students grossly overestimating their grades and higher achieving students slightly underestimating their grades, but this effect decreased over time. For example, statistical significance was achieved for Group 3's "day of" predictions between midterm test 1 and 2 ($F_{(2,46)} = 10.21$, $p < 0.001$) and all Groups' postdictions between midterm test 1 and 2 ($F_{(1,159)} = 18.03$, $p < 0.001$), with predictions being more accurate at the later time points. A reduction in the Dunning Kruger effect suggests improvements in metacognition. Interestingly, participants reported an overall decrease in metacognition from beginning to end of term, and treatment group did not affect test, exam, or final course grades.

Phase IV

With several positive results realized through Phases I-III of the DBER project, a formal metacognition co-curriculum was implemented during Phase IV. For the first time, all 289 students registered in the course were expected to complete metacognitive practice activities as a required component of the course. Of these 289 students, 259 agreed to have their data analyzed. The metacognition co-curriculum accounted for 10% of the final course grade and was composed of the following activities: (i) an introductory survey including a metacognitive awareness assessment (1%); (ii) an instructor-developed prior learning assessment (2.5%); (iii) 8 weekly inventory cues (1% if at least 6 of 8 were completed); (iv) 8 weekly LTIs (1% if at least 6 of 8 were completed), (v) weekly in-class confidence questions (2.5%); (vi) 9 test grade prediction and postdiction surveys, including 3 one-week predictions, 3 day of predictions, and 3 postdictions immediately following each test (1% if at least 7 of 9 were completed); and (vii) an end-of-term survey including a metacognitive awareness assessment (1%).

The instructor-developed prior learning assessment consisted of 22 multiple-choice questions based on relevant material from the prerequisite general chemistry courses. Students first completed the assessment with no preparation and for no points and could immediately review their results. Students were then granted access to relevant course resources and provided the opportunity to review material before completing the assessment again 1 week later, this time for 2.5% of their final course grade. This assessment served as an introduction to the metacognition co-curriculum, prompting students to reflect on prior learning while completing the first iteration of the assessment, then self-assess and review only the content for which they struggled in preparation for the second iteration.

The addition of weekly in-class confidence questions was inspired by a report on the use of formative assessment and self-regulated learning in mathematics (Hudesman et al., 2014). In each class, whenever a content-based question was posed (typically 3–4 per class), students would first be asked to rate their level of confidence in responding to the question on a 5-point scale. For example, if the content question was to indicate the number of chirality centers in an organic molecule, students would be told what the question is but would be asked to rate their confidence before being shown the structure. Immediately following the collection of responses to the content question, students would be asked to rate their level of confidence in having responded correctly to the question before being shown the correct answer. These confidence questions are similar but distinct in that the first prompts students to summon and assess prior learning, whereas the second asks them to judge performance.

Major findings from Phase IV include (i) a clear Dunning-Kruger effect across the term for test grade pre-and postdictions, with (a) low-performing students overestimating grades and high-performing students underestimating, (b) the overall effect decreasing over time, both within test and between test, (c) low-achieving students becoming more accurate over time but high-achievers becoming less accurate, and (d) day of predictions and postdictions being significant predictors of actual test scores (e.g., for the final exam, $F_{(3,140)} = 45.18$, $p < 0.001$; See Supplementary Information Phase IV findings); and (ii) an increase in confidence for both pre-and post-question confidence questions when average confidence ratings before midterm test 1 were compared to average confidence ratings after midterm test 2 (e.g., the significant main effect for pre-question confidence was $F_{(2,244)} = 105.06$, $p < 0.001$). Interestingly, and consistent with previous research (Hacker et al., 2000), low performing students on the final exam were those who tended to be overly confident on pre-question ratings – even after controlling for GPA (e.g., for average pre-question confidence prior to midterm test 1, $\beta = -4.46$, $t = -2.84$, $p = 0.005$). Overestimations in confidence for post-questions, however, were no longer evident and post-question responses did not predict actual test grades for any of the tests. Finally, and once again, metacognitive awareness assessment scores were observed to decrease from beginning-to end-of-term.

Discussion

The key aim of the present report was to describe the development of a metacognition co-curriculum that was implemented to scaffold and enhance students' use of metacognitive skills in an Organic Chemistry course. An important part of the translation from the literature to practice involved quasi-experimental analyses to better assess change, challenges, and successes. Borrowing from literatures spanning education, cognition and instructional psychology, a series of supports were integrated into ongoing classroom instruction over four offerings of the same course. The translation process was iterative and progressive culminating in a true co-curriculum of metacognition for students in the course. A salient outcome of the process was that supporting students' application of metacognitive skills enhanced their course performance. Perhaps more important is that the metacognitive co-curriculum supports an array of skills tied to metacognition that students may or may not have and that engaging in these skills affirms the importance of these advanced learning behaviors which may also be transferred to other domains they study.

Several supports were introduced in this program including priming for content accompanied by self-assessment of competence with learning task inventories (LTIs), self-assessments of confidence regarding in-class content questions, and performance predictions and postdictions on tests. In general, exposure to these supports resulted in overall performance and confidence gains.

Scaffolding metacognitive skills by providing students with the full inventories of learning tasks by the end of the week and subsequently with LTI cues at the beginning of a week was associated with learning gains as evidenced by improvements in final exam or final course grades. These interventions support students' ability to assess their declarative knowledge which allows them to control and direct subsequent cognitive processes involving planning, monitoring, and evaluating their acquisition of these concepts (Dimmit and McCormick, 2012; Rivas et al., 2022). However, availability of these supports is not sufficient. When they were simply available, students did not access them. When students were required to access them, their grades improved. Interestingly, many students indicated skimming the initial "cuing" lists. At a glance, this may seem to signify minimal engagement. However, advanced learners should be skimming such content to quickly assess familiarity with the concepts as a first step in processing what they do or do not know. To engage students more deeply, they need to be encouraged to try to solve the tasks, evaluate their attempts, and make corrections as needed. This was facilitated through two mechanisms, introducing mastery questions to have students solve problems and subsequently pairing these content questions with confidence questions asked before and after answering the content questions during class. These components encouraged students to engage their procedural and conditional knowledge skills to access strategies and problem solving specific to the question. The confidence questions also encouraged further assessment in these knowledge areas through self-reflection regarding their ability to respond to the questions and their performance once they answered questions (Bangert-Downs et al., 2004; Dignath and Büttner, 2008). These adaptations supported students in the learning process. The test and exam grade predictions and postdictions assisted students in their application of metacognitive skills in the assessment process. Given that calibrations became more accurate over practice is evidence that students were better able to assess their learning. Instating these estimations before, during, and after testing extended use of metacognitive skills throughout the process.

An important consideration when shifting from testing different iterations to translating the metacognitive scaffolds into a full co-curriculum was the intentional assignment of grades for completion of the co-curriculum elements. Although learners may be expected to adopt supports when they are made available, it is clear from our studies that students may not engage in best practice especially in courses where attention to anything but content may be perceived to be costly. Given demands inherent in learning the course content, students may perceive that the co-curriculum presents additional, unsurmountable demands that could cost performance. Evidence of this reluctance was clear through the limited uptake of the LTIs when review was made optional and through the dismissal of the LTI content by students performing most poorly in the course. Thus, to signal to students that the co-curriculum is valuable and worthy of their efforts, aspects were assigned course weight. This allowed students to engage and embrace the co-curriculum. Developing these skills in the context of a required introductory organic chemistry course provides a foundation for generalizing the skills learned through this course to more senior courses in this domain but also for transfer across domains.

As part of the process of creating the metacognitive co-curriculum we identified challenges translating some evidence-based and evidence proven interventions identified as supports in the extant literature. For example, there were no differences across condition for those engaged in massed versus distributed practice with the LTIs, priming worked sometimes but not consistently, and ratings for aspects of metacognitive awareness tended to decrease over time. Despite the lack of support for distributed practice in the one instance where we tested it, we adopted it as part of the co-curriculum given substantial work showing its value (e.g., Donovan and Radosevich, 1999). Our failure to detect differences in these two study approaches may have been an artifact of the many manipulations being tested at the same time which may have masked effects. Similarly, the initial lack of priming effects (when tested in Phase II) was evident in later iterations when cueing and self-evaluation were the greater focus of study. Regardless, priming with the list cues at the beginning of the week was retained in the co-curriculum. Initially, we were concerned to see a drop in students' self-reported metacognitive awareness scores from the beginning to the end of term. However, the reduction in the Dunning-Kruger effect and enhanced calibration in prediction and postdiction measures suggests that the decrease in students' evaluation of their metacognitive skills may reflect a shift from an overly optimistic view of their abilities to a more realistic evaluation as a function of engaging in so many relevant activities.

Conclusion, constraints, and future work

Overall, at the class level, the addition of the metacognition co-curriculum improved students' performance as indicated by final exam and final course scores, and improved metacognitive skills, as indicated by the reduction in the Dunning-Kruger effect. The optimistically high ratings of the in-class confidence question results, particularly when coupled with the negative correlation to final exam scores, suggest a possible self-preservation mechanism that hints at the importance of considering the affective domain of learning. Results also indicate that not all students interacted with the elements of the metacognition co-curriculum to the same extent. This may be expected given individual differences in metacognitive skills and especially with these skills applied to different domains. It may be the case that the co-curriculum benefits some students more than others. For example, more advanced students may experience less benefit, likely because they have already developed and automatically use the skills that the metacognition co-curriculum aims to foster. Examining individual differences as a function of the co-curricular program in general and as a function of different components of the co-curriculum would be ideal to establish what works best for which learners under which circumstances. Being able to assess performance beyond a single term also permits evidence of changing needs over time which would be another important avenue of exploration. Finally, one aspect that the present model of metacognitive co-curriculum has not assessed is social metacognition, the sharing of metacognitive responsibilities when students engage interactively with one another when working in partnership or groups (e.g., Chiu and Kuo, 2009). Given the shift to active and student-centered learning, this additional aspect of metacognition would be interesting to explore and would contribute to a broader understanding of metacognitive skill development in the active classroom.

The present study describes the development of a co-curricular metacognitive program for chemistry students that can be adapted and extended to other chemistry courses or other domains in higher education. Instructors wanting to implement this metacognitive co-curriculum can follow the elements outlined above. Instructors will need to prepare an introduction to metacognition for their students which can be achieved using available resources (see Cook et al., 2013). Instructors will need to create a prior learning assessment based on the prerequisite course(s) which can be drawn from examination of course outlines of these courses. Finally, instructors need to develop detailed lists of learning tasks associated with their course. All other components of the metacognition co-curriculum identified above (having students predict their grades at planned times during the course and assess confidence in answering content-based in-class questions), can be implemented with minimal preparation. In general, over our series of studies, exposure to these supports resulted in overall performance and confidence gains, and these results should transfer to other chemistry courses and domains.

Data availability statement

The data sets presented in this article are not readily or publicly available. However, anonymized versions of some of the data sets may be available upon reasonable written request to the first author. Ethics approval and informed consent for some of the data reported here did not include sharing of data sets. In these situations, those specific aspects of the data and other identifiable information (gleaned from open ended answers where participants did not provide permission for quotes or use) may not be shared. Requests to access the data sets should be directed to SM, smacneil@wlu.ca.

Ethics statement

The studies involving humans were approved by the Research Ethics Review Board, Wilfrid Laurier University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in the studies described here.

Author contributions

SM: Writing – original draft, Writing – review & editing. EW: Writing – original draft, Writing – review & editing. FA: Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We thank the undergraduate students who assisted with data collection and analysis associated with this DBER project, including Patrick Smith, Robyn Glover, Hanna Douglas, Deanna MacNeil, Brandon Mullen, Brittany Doerr, Andrea Debrouwer, Mike Urschel, and Tarique Plummer.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2024.1402599/full#supplementary-material

SUPPLEMENTARY DATA SHEET 1
Representative measures and selected data.

References

Bangert-Downs, R. L., Hurley, M. M., and Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: a meta-analysis. *Rev. Educ. Res.* 74, 29–58. doi: 10.3102/00346543074001029

Benjamin, A. S., and Tullis, J. (2010). What makes distributed practice effective? *Cogn. Psychol.* 61, 228–247. doi: 10.1016/j.cogpsych.2010.05.004

Chiu, M. M., and Kuo, S. W. (2009). From metacognition to social metacognition: similarities, differences, and learning. *J. Educ. Res.* 3, 321–338.

Cook, E., Kennedy, E., and McGuire, S. (2013). Effect of teaching metacognitive learning strategies on performance in general chemistry courses. *J. Chem. Educ.* 90, 961–967. doi: 10.1021/ed300686h

Dignath, C., and Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacogn. Learn.* 3, 231–264. doi: 10.1007/s11409-008-9029-x

Dimmit, C., and McCormick, C. B. (2012). "Metacognition in education" in *APA educational psychology handbook*. eds. K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra and J. Sweller (Washington, DC: American Psychological Association).

Donovan, J. J., and Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: now you see it, now you don't. *J. Appl. Psychol.* 84, 795–805. doi: 10.1037/0021-9010.84.5.795

Fazio, L. K., Huelser, B. J., Johnson, A., and Marsh, E. J. (2010). Receiving right/wrong feedback: consequences for learning. *Memory* 18, 335–350. doi: 10.1080/09658211003652491

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am. Psychol.* 34, 906–911. doi: 10.1037/0003-066X.34.10.906

García-Pérez, D., Fraile, J., and Panadero, E. (2021). Learning strategies and self-regulation in context: how higher education students approach different courses, assessments, and challenges. *Eur. J. Psychol. Educ.* 36, 533–550. doi: 10.1007/s10212-020-00488-z

Garrison, R. D., and Vaughan, N. D. (2008). *Blended learning in higher education: framework, principles, and guidelines*. San Francisco: Jossey-Bass.

Geurten, M., Meulemans, T., and Lemaire, P. (2018). From domain-specific to domain-general? The developmental path of metacognition for strategy selection. *Cogn. Dev.* 48, 62–81. doi: 10.1016/j.cogdev.2018.08.002

Grimmett, P., and MacKinnon, A. (1992). "Craft Knowledge and the education of teachers" in *Review of research education*. ed. C. Grant (Washington, DC: American Educational Research Association), 385–456.

Hacker, D. J., Bol, L., Horgan, D. D., and Rakow, E. A. (2000). Test prediction and performance in a classroom context. *J. Educ. Psychol.* 92, 160–170. doi: 10.1037/0022-0663.92.1.160

Heikkilä, A., and Lonka, K. (2006). Studying in higher education: students' approaches to learning, self-regulation, and cognitive strategies. *Stud. High. Educ.* 31, 99–117. doi: 10.1080/03075070500392433

Hudesman, J., Crosby, S., Ziehmke, N., Everson, H., Isaac, S., Flugman, B., et al. (2014). Using formative assessment and self-regulated learning to help developmental mathematics students achieve: a multi-campus program. *J. Excel. Coll. Teach.* 25, 107–130.

Kranz, D., Schween, M., and Graulich, N. (2023). Patterns of reasoning – exploring the interplay of students' work with a scaffold and their conceptual knowledge in organic chemistry. *Chem. Educ. Res. Pract.* 24, 453–477. doi: 10.1039/D2RP00132B

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory Pract.* 41, 212–218. doi: 10.1207/s15430421tip4104_2

Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134. doi: 10.1037/0022-3514.77.6.1121

MacNeil, S., Wood, E., Zivcakova, L., Glover, R., and Smith, P. (2013). Learning task inventories (LTIs). Exploration of optimal conditions to help students develop, improve and sustain good study and learning practices. *Coll. Essays Learn. Teach.* 2, 1–9. doi: 10.22329/celt.v7i2.3976

Merriam, S. B. (2001). Andragogy and self-directed learning: pillars of adult learning theory. *New Directions for Adult and Continuing Education* 89, 3–14. doi: 10.1002/ace.3

Pashler, H., Cepeda, N. J., Wixted, J. T., and Rohrer, D. (2005). When does feedback facilitate learning of words? *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 3–8. doi: 10.1037/0278-7393.31.1.3

Pilcher, L. A., Potgieter, M., and Fletcher, L. (2023). Blending online homework and large class tutorials to provide learning support for introductory organic chemistry. *Afr. J. Res. Math., Sci. Technol. Educ.* 27, 1–13. doi: 10.1080/18117295.2022.2155771

Ratcliff, R., Allbritton, D., and McKoon, G. (1997). Bias in auditory priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 143–152. doi: 10.1037/0278-7393.23.1.143

Rivas, S. F., Saiz, C., and Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Front. Psychol.* 13:913219. doi: 10.3389/fpsyg.2022.913219

Roediger, H. L., and Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychol. Sci.* 17, 249–255. doi: 10.1111/j.1467-9280.2006.01693.x

Saks, K., Ilves, H., and Noppel, A. (2021). The impact of procedural knowledge on the formation of declarative knowledge: how accomplishing activities designed for developing learning skills impacts teachers' knowledge of learning skills. *Educ. Sci.* 11:598. doi: 10.3390/educsci11100598

Schraw, G., Crippen, K., and Hartley, K. (2006). Promoting self-regulation in science education: metacognition as part of a broader perspective on learning. *Res. Sci. Educ.* 36, 111–139. doi: 10.1007/s11165-005-3917-8

Schraw, G., and Dennison, S. R. (1994). Assessing metacognitive awareness. *Contemp. Educ. Psychol.* 19, 460–475. doi: 10.1006/ceps.1994.1033

Scott, B. M., and Berman, A. F. (2013). Examining the domain-specificity of metacognition using academic domains and task-specific individual differences. *Aust. J. Educ. Dev. Psychol.* 13, 28–43.

Zohar, A., and Dori, Y. J. (2012). "Metacognition in science education" in *Trends in Current Research*. ed. D. Ziedler (Dordrecht: Springer), 1–19.

frontiers | Frontiers in Education

TYPE Hypothesis and Theory
PUBLISHED 06 August 2024
DOI 10.3389/feduc.2024.1412417

Check for updates

OPEN ACCESS

EDITED BY
Jodye I. Selco,
California State Polytechnic University, Pomona, United States

REVIEWED BY
Wahyu Widada,
University of Bengkulu, Indonesia
Sarbjeet Kaur,
Curia Global, United States
Nicole Graulich,
University of Giessen, Germany

*CORRESPONDENCE
Jay Wm. Wackerly
✉ wackerlyj@central.edu

RECEIVED 04 April 2024
ACCEPTED 05 July 2024
PUBLISHED 06 August 2024

CITATION
Wackerly JW, Wentzel MT and Zingales SK (2024) Teaching abductive reasoning for use as a problem-solving tool in organic chemistry and beyond.
Front. Educ. 9:1412417.
doi: 10.3389/feduc.2024.1412417

Teaching abductive reasoning for use as a problem-solving tool in organic chemistry and beyond

Jay Wm. Wackerly[1]*, Michael T. Wentzel[2] and Sarah K. Zingales[3,4]

[1]Chemistry Program, Department of Natural Sciences, Central College, Pella, IA, United States, [2]Department of Chemistry, Augsburg University, Minneapolis, MN, United States, [3]Department of Chemistry, University of Saint Joseph, West Hartford, CT, United States, [4]Department of Chemical and Environmental Sciences, United States Coast Guard Academy, New London, CT, United States

The second-year undergraduate Organic Chemistry course sequence is often cited as one of the most, if not the most, challenging for students in the US. Thus, a persistent question remains: What is it about Organic Chemistry that makes the course so difficult for students? Herein, we put forward the hypothesis that a new mode of thinking and problem solving is expected of the students; these skills have not yet been developed in their prior scientific coursework and are often not deliberately taught in Organic Chemistry. This form of reasoning and problem solving, known as abductive reasoning, is highlighted for its connection to medical diagnosis and scientific thinking. We provide examples to showcase how instructors could explicitly foreground the reasoning process in their classroom. Ultimately, we argue that teaching how to reason using abduction may benefit students in both the short term (in the course) and the long term (in their careers as scientists and medical practitioners).

KEYWORDS

abduction, abductive reasoning, organic chemistry, diagnosis, metacognition, problem solving, pre-health education

> "What changes must be made in the *kind* of science that we teach and the way that we teach it so that the fundamental ideas of our discipline can be used outside the classroom?" – Herron & Greenbowe

1 Introduction

1.1 Background

Organic Chemistry, as traditionally taught in the US as a primarily second-year undergraduate course sequence, is often considered a course for "weeding out pre-meds" (Moran, 2013) that "strik[es] fear in the hearts of students" (Garg, 2019). This socially constructed barrier adds an additional level of pedagogical challenge for instructors. We, the authors, are instructors of Organic Chemistry and also write and review questions for standardized exams that are required for entrance into specialized medical programs;[1] thus, we are at a position in both the content delivery and assessment where we find ourselves continually asking the question: What do we want students to learn in the Organic Chemistry course sequence?

While some students may think the answer to this question is "to know, understand, and recite back the course material," this is an unsatisfying response for a number of reasons. First, such a response would imply that only memorization and algorithmic problem-solving skills are necessary for success in Organic Chemistry (Stowe and Cooper, 2017).[2] However, expert organic chemists recognize that the interconnected complexities within chemical systems means that simply following basic rules (i.e., deductive inference) will not necessarily lead to a set outcome (e.g., bulky bases do not *always* react via E2) (Achet et al., 1986). Second, while the students enter our classrooms as novices, some of them will go on to become practicing, expert organic chemists. We owe it to them, and the future of scientific discovery, to build a sound foundation of both fundamental (e.g., understanding the aldol condensation) and higher order (e.g., performing retrosynthetic analysis) skills within the discipline. Third, most US health professions (e.g., MD, DO, PA, DDS, DMD, OD, PharmD) require this course to be taken as a prerequisite for admission into their graduate programs (Kovac, 2002). These students should be presented, within their undergraduate education, the chance to improve their scientific reasoning and critical thinking skills. We think that these three features, which might not be clear to all students entering the course, illustrate that students are expected to learn and problem solve in new ways—essentially to begin to "think like a chemist" (e.g., Platt, 1964).

While certain ideas within this article were presented in a preceding paper (Wackerly, 2021), we intend to flesh out and expand upon some of those initial assertions in this manuscript and craft a more detailed hypothesis that the use of abductive reasoning is critical in the learning of organic chemistry concepts. Herein we provide support for this hypothesis by viewing it from a few different conceptual angles. First, we provide a science education overview on why learning certain organic chemistry concepts is considered challenging for students. Then, we briefly summarize the medical education viewpoint on the teaching of diagnosis and why this is important to many students in Organic Chemistry. Finally, using the lens of the Organic Chemistry curriculum we provide problem-solving examples of how abductive reasoning can assist in the teaching and learning of organic chemistry.

1 We wish to keep the focus of this manuscript on the relevant student population of the Organic Chemistry course sequence. Students intending to pursue medically relevant careers which require advanced degrees (e.g., medical, dental, optometry, pharmacy, etc.) are a large portion of this population. However, if the reader is curious, we specifically write for the dental and optometric admissions exams.

2 In this manuscript we attempt to provide the reader a broad overview of important chemical education and philosophy of chemistry publications. Since this is not a review article and the scope is quite a bit smaller, all possible relevant literature has not been cited.

1.2 Why is science difficult to learn?

Johnstone asked this titular question in his seminal 1991 paper (Johnstone, 1991). One conclusion that he drew, which has since been supported by a variety of other work (e.g., Graulich, 2015; Tiettmeyer et al., 2017; Reid, 2020; Dood and Watts, 2022), is that the nature and complexity of scientific concepts strain the working memory of students. To assist instructors in conceptualizing the strain of a given concept, he created the "triangle model" which illustrated three levels of thought (Figure 1). He argued that the more levels a concept included the more cognitive load was placed on students.

One feature that might make learning science difficult is that the instructor, or expert, may not be aware of the extent of cognitive load they are placing on students, or novices. When "multicomponent phenomena that are invisible, dynamic, and interdependent" are presented to students, a large demand is placed on the working memory of novices (Hmelo-Silver et al., 2007). However, experts are able to easily connect two or more cognitive components by "chunking several pieces of information together" (Overton and Potter, 2008) and through years of practice (Randles and Overton, 2015). Specialization within a discipline that requires connecting multiple levels will lower cognitive load for such repetitive tasks over time (Tiettmeyer et al., 2017; Price et al., 2021). However, students have typically not been exposed to such tasks, let alone have the opportunity to consistently repeat them, and thus instructors need to disentangle new concepts that might cause cognitive overload for students so they can process and incorporate new material starting from their present knowledge base and scientific models.[3]

> "[R]easoning [is the] knowledge of some facts [which] leads to a belief in others not directly observed." – C. S. Peirce

1.3 Why is organic chemistry so difficult to learn?

Here we argue that it should come as no surprise when former and current students of organic chemistry cite that organic chemistry is difficult to learn, because they are asked to problem solve and reason in new ways utilizing new content without prior exposure to, or repetition of, these scientific tasks.[4] Naturally, when a student enters a course they are expected to be ignorant of the course content since they enroll to learn it. However, students might feel that a bait-and-switch has occurred in Organic Chemistry because not only is the content new, but the logical processes required to be successful are also typically new to the students as well.

In prior scientific courses, which for most pre-health (*vide infra*) US students are two courses in general biology and two in general

3 Cognitive overload could also stem from misconceptions and oversimplified concepts, such as the oft-stated "breaking bonds in ATP releases energy" from introductory biology courses.

4 This can be contrasted with General Chemistry which repeats some of the content of the high school chemistry.

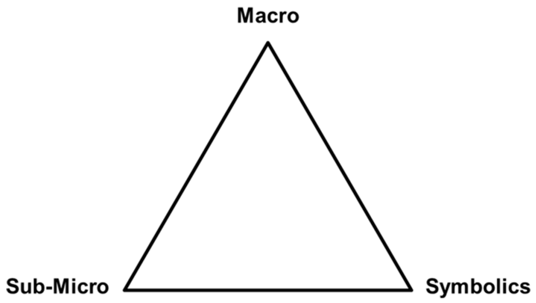


FIGURE 1
Reproduction of Johnstone's model: "Triangle of Levels of Thought".

chemistry, students are typically required to perform recall (memorization) or reason algorithmically on summative assessment items (Raker and Towns, 2010). While these skills hold value in organic chemistry, current organic chemistry education research shows that skills such as multivariate (Kraft et al., 2010; Christian and Talanquer, 2012) and mechanistic reasoning (Bhattacharyya, 2013) are more important.[5] Thus, inspired by the work in chemistry education research, the philosophy of science, and Johnstone's seminal triangle, here we propose a tetrahedron model of layered reasoning strategies that are important for consideration by instructors when teaching novice organic chemistry students.

The bottom-most point of the tetrahedron (Figure 2) was chosen to be memorization because it is not a reasoning skill. However, terms and chemical facts still need to be learned by students, which is often not a problem because they have developed this skill during their general biology and chemistry coursework. Algorithmic reasoning is a skill many students leaving General Chemistry assume they will utilize in Organic Chemistry because it was employed so frequently in that course. For example, if a student knows the pressure, temperature, and number of moles of an ideal gas, these students will likely be able to provide the volume of the gas's container. While these mathematical and deductive reasoning skills remain relevant in the laboratory portion of Organic Chemistry and even for the IUPAC naming of organic molecules (i.e., there is a definitive rule set), they start to break down when chemical systems become more complex and chemical formulas evolve to contain more meaning in the form of chemical structures.

The right corner of the tetrahedron is for the set of competencies required to interpret diagrams in organic chemistry, such as visualization (Gilbert, 2005), visuo-spatial reasoning (Pribyl and Bodner, 1987; Habraken, 1996), and representational competence (Kozma and Russell, 1997). In lieu of individually listing these skills, we designate this corner as perceptual learning, which integrates conceptual knowledge with a broad set of skills, including those related to visualization and representational competence (Van Dantzig et al., 2008; Kellman and Massey, 2013). Perceptual learning "refers, roughly, to the long-lasting changes in perception that result from practice or experience" (Connolly, 2017), and is beginning to be more deeply explored in organic chemistry pedagogy (e.g., Kim et al., 2019).

We briefly illustrate how changes associated with perceptual learning might take place with students. Consider, for example, that in General Chemistry students might be asked to calculate the heat of combustion of hexane (denoted at C_6H_{14}). For most students at that stage, the sole association they would have with the compound's name is its molecular formula, whereas its "zig-zag" structure might represent nothing more than a crooked line. As these students progress into Organic Chemistry and learn about different representational systems and constitutional isomers, the verbal representation "hexane" changes, this is because the term is now associated with five unique isomers each with unique connectivity, properties, and reactivity (e.g., radical reaction with Br_2). Through this process, the students' *perception* for the term "hexane" changes from representing a single molecular formula to representing a family of five constitutional isomers each with a unique bond-line structure. This process continues as students advance to more complex structures (e.g., stereochemistry) and learn additional concepts like three-dimensionality, IMFs, physical properties, etc. We propose that the three corners of the tetrahedron discussed thus far are often directly connected to abductive reasoning which focuses on solving problems by generating the most likely most likely outcome of a chemical situation.

Our hypothesis includes the postulation that abductive reasoning is a complex reasoning skill for students in Organic Chemistry and should explicitly be taught in the classroom. While this idea has been presented by us previously (Wackerly, 2021), here we will just provide a brief overview so we can move on to discuss the relevance of this reasoning skill within the Organic Chemistry classroom and to highlight some examples. Firstly, the term "abduction" (Douven, 2021) is often used interchangeably with the terms "inference to the best explanation" (Lipton, 2017) and "scientific hypothesis"—and below we will argue "diagnosis." All of these terms hold common ground in that they use reasoning that connects various (similar or dissimilar) pieces of evidence/observations together in a way where a plausible conclusion can causally describe the collection of phenomena.[6] For example, say you are inside of grain windmill by the grindstone, and then you begin to see the stone rotating and producing flour. You will abduce that the weather outside has become windy. While this is a simple example only requiring you to understand that outside wind turns the sails and the sails, via a series of machinery, turn the grindstone, it is similar to the reasoning employed by expert organic chemists. Leaving the windmill and heading into your synthetic laboratory, let us say you wish to publish a new compound in the *Journal of Organic Chemistry*. According to the journal, to conclude that you have made this new compound you must "establish both identity and degree of purity." Minimally, this means you will need to obtain a ^{1}H NMR spectrum, ^{13}C NMR spectrum, and HRMS spectrum then interpret the data present in the spectra to abduce the molecular

5 Multivariate and mechanistic reasoning are highlighted as examples because they often require combining features from all four points of the tetrahedron.

6 The conclusion need not explain the entire collection of evidence as some may be irrelevant, and they are unrelated to the conclusion. However, the entire collection may not contain a piece of evidence that refutes the conclusion. Thus, abductive reasoning can be useful in differentiating science from non-science and pseudoscience.

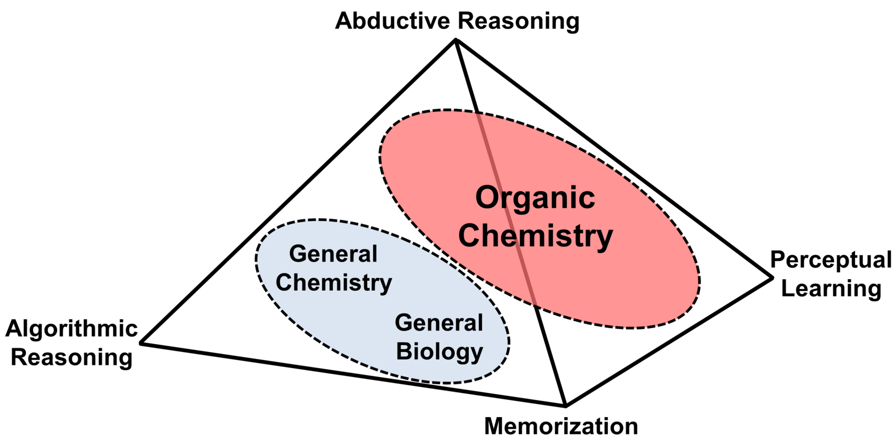


FIGURE 2
Tetrahedron model of problem-solving in Organic Chemistry.

structure of your new compound. This exact same skill that is required of expert organic chemists, is typically required of students in Organic Chemistry (Stowe and Cooper, 2019a). Thus, these students should be taught how to reason like expert scientists in order for them to develop into scientists (Cartrette and Bodner, 2010). Just as the spectroscopic analysis example highlights, instructors of Organic Chemistry often profess a goal is for students to develop critical thinking and scientific problem-solving skills: Our hypothesis presented here is that instructors must explicitly utilize the abductive reasoning process within their teaching and assessment.

Solving problems that require abductive reasoning will also require skills from the three other points of the tetrahedron, which will render them cognitively complex. Teaching abductive reasoning in the classroom should not require additional formal training for instructors/experts since abductive reasoning skills have already been developed over the course of their careers. Further, philosophers have long held (Harman, 1965) that humans utilize abductive reasoning as a matter of course in their day-to-day lives. Paralleling human logic, abductive reasoning has likely been utilized (Pareschi, 2023) and will continue to be (Dai and Muggleton, 2021) an integral part of artificial intelligence. This reasoning skill is particularly important for students required to take Organic Chemistry. It might be obvious that future scientists will need the skills to create new hypotheses and design experiments that could potentially refute current hypotheses, but in our experience, it seems less obvious to pre-health students that using abductive reasoning for problem solving in Organic Chemistry will play a critical role in their desired careers.

2 Framing for pre-health students (diagnosis)

2.1 Why is organic chemistry relevant for pre-health students?

In a post-COVID world where test-optional admissions are on the rise and the future of post-graduate education feels increasingly uncertain, convincing students of the importance of Organic Chemistry goes beyond just passing the course. This is especially true for the majority of students taking Organic Chemistry who are pre-health majors. Instructors need to show students the connection between organic chemistry and the health field.

Thus, problem solving in Organic Chemistry can be framed as a diagnostic problem-solving tool–similar to what medical practitioners do when making a diagnosis (Stowe and Cooper, 2019b). By overtly showing students the parallels between medical diagnosis and organic chemistry problem solving, instructors demonstrate that students are not just being taught a bunch of facts–they are developing critical thinking skills they can use in the real world. Bridging the gap between theory and practice helps students see the bigger picture and gives them the tools they need to succeed in both their studies and future careers.

The parallels between medical diagnosis and organic chemistry problem solving should be readily apparent (Table 1). Both involve analyzing complex systems (human body/chemical reactions) to identify patterns and relationships, emphasizing the importance of critical thinking and logic-based problem-solving skills, as well as using evidence. Both fields rely on the use of abductive reasoning (Wackerly, 2021; Martini, 2023), although typically neither field explicitly states it to students. Table 1 uses simplified language accessible to students that describes the abductive theory of method (ATOM) in clinical diagnosis (Vertue and Haig, 2008), and its parallel to expert thinking in organic chemistry.

For example, to "diagnose" the product of an organic chemistry reaction, first the background information, including structure, reactivity, and stability of the starting materials and reagents must be analyzed, which is similar to how medical professionals take patient history. Abductive reasoning is then used to generate the most likely answer. Finally, the hypothesis is tested through gathering evidence such as utilizing spectroscopic analysis which is similar to a physician ordering lab work or imaging. This is an iterative process, wherein multiple pieces of spectroscopic evidence are needed to point to the same answer. Similarly, a physician may order additional studies or perform physical exams to support or refute their medical diagnosis. Although the goals appear different, the same skills are developed such as drawing hypothesis based on empirical evidence. By explicitly demonstrating how these thought processes are parallel, instructors of Organic Chemistry may help students to appreciate the mental training they are receiving in the course.

TABLE 1 Comparison of medical diagnosis to skills developed in Organic Chemistry.

	Medical diagnosis	Parallel to organic chemistry
Background information	Involves gathering patient history, physical exams, and tests	Requires analysis of structure and reactivity of molecules
Initial hypothesis	Utilizes abductive reasoning to form a preliminary diagnosis	Involves abductive reasoning to propose plausible mechanisms or structures
Iterative nature	Hypotheses are continuously revised based on feedback from other clinicians and tests	Hypotheses are tested and revised through experimentation
Goal	To identify the underlying cause of symptoms and provide treatment	To generate plausible mechanisms and/or structures from observed data
Skills developed	Critical thinking, pattern recognition, hypothesis refinement	Critical thinking, analytical skills, hypothesis refinement
Importance of evidence	Relies on empirical data and patient feedback	Relies on experimental data and spectroscopic analysis
Real-world application	Used by physicians to generate positive health outcomes for patients	Used by chemists to create new medicines, materials, etc.

Organic Chemistry has been deemed essential as a prerequisite for medical school by a panel of medical school professors of biochemistry (Buick, 1995). While many current medical students do not think that the material covered in Organic Chemistry was a valuable part of their undergraduate curriculum, the majority agree that the critical thinking skills learned in the course were valuable (Dixson et al., 2022). While there are those in the field of medicine who think that Organic Chemistry should be de-emphasized in the pre-med curriculum, those that defend Organic Chemistry do so for some of the same reasons we discuss herein, namely that the critical thinking and problem-solving skills in the course directly align with patient diagnosis (Higgins and Reed, 2007).

This process of abductive reasoning coupled with framing for the medical field may serve the students better in both the short term and long term. Students who employ more metacognitive strategies such as the type we are advocating for here are better able to solve problems in Organic Chemistry (Blackford et al., 2023). Connecting course material to students' future career aspirations also leads to better engagement and course performance (Hulleman et al., 2010). Additional benefits of this diagnostic reasoning process include students' ability to apply this metacognitive strategy in other courses in their majors, such as biology (Morris Dye and Dangremond Stanton, 2017), and their future medical careers (Friel and Chandar, 2021). Therefore, diagnostic reasoning should be explicitly modeled and assessed in Organic Chemistry courses.

2.2 Using "diagnosis" in examples for students

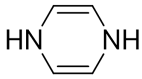
While there are a variety of ways to teach students how to approach organic chemistry problems like an expert, we would like to present how to do this through the lens of "diagnosis." Other ways of describing argumentation and the process of problem solving have been discussed in the chemical education literature (e.g., Cruz-Ramírez De Arellano and Towns, 2014; Stowe and Cooper, 2019a; Walker et al., 2019) as well as the philosophy of chemistry literature (e.g., Kovac, 2002; Goodwin, 2003). While they differ in the number of steps and what those steps are called, the processes have a similar logical flow. First, gather evidence and make observations (*What you see*), link this to previous knowledge (*What you know*), and finally make a reasoned conclusion (*Hypothesis*) which is a logical consequence—often via abductive inference.

The following examples (Figures 3–6) are designed to highlight the use of these three steps to explicitly diagnose problems from across the two semester Organic Chemistry sequence. This process can be used in the classroom as a model to guide students through the abduction process and could be used to explicitly scaffold problems. Moreover, instructors can use this model to ascertain the complexity of their assessments including the required prerequisite factual knowledge and the multiple steps required. The complexity of organic chemistry questions is determined by the number of "subtasks" the student must complete (Raker et al., 2013), factual knowledge required, and facets of perceptual learning (*vide supra*). A number of explicit decisions were made in formulating the below questions. The discussion points are certainly not exhaustive, and practitioners should adapt questions to their own students and situations. The amount of information provided or not provided, such as the exclusion of lone-pairs and inorganic by-products, was chosen to be consistent with the information provided by practicing organic chemists and one goal of teaching organic chemistry is to facilitate the development toward expert-level practice. We intentionally included one example of additional information, Figure 4 entry marked with a *, to highlight that there are many more subtasks that could be utilized to assist with arriving at a probable conclusion, but we tried to exclude all other non-essential explanations. We do not suggest that all students should solve each problem from top to bottom as outlined here; in reality expert chemists often take different routes, based on the same evidence and premises, to reach similar conclusions. Although these problems are multiple-choice, we have modeled how to solve them as either multiple-choice or open format. The complexity of these questions can also be adjusted, for example in Figure 4 the mechanistic arrows could be included in the distractors and answer instead of in the stem. This type of alteration can allow for the assessment of mechanistic thinking (e.g., Bodé et al., 2019; Finkenstaedt-Quinn et al., 2020; Watts et al., 2020; Dood and Watts, 2022). The following examples demonstrate that when the diagnosis/abduction process is utilized, students can develop and enhance their problem-solving skills.

The first example shown in Figure 3 is a case of aromaticity (Jin et al., 2022). Students will typically memorize the requirements and check the structure for being cyclic, planar, containing Huckel's number $(4n+2)$ electrons, and a p orbital at every vertex (i.e., conjugated). However, this problem does not ask for a simple definition of aromaticity, but an application of the ruleset to a structure students would not have typically encountered. The diagnosis requires observations about the structure including

 10.3389/feduc.2024.1412417

Which best explains why this molecule is non-aromatic?



A) The ring has 8 π electrons
B) The molecule is not planar
C) The molecule has no conjugation
D) The molecule is acyclic

Evidence & Observations (What you see)	Premises (What you know)	Reasoned Conclusion (Hypothesis)
A six-membered ring	Cyclic systems are a requirement for aromaticity (low energy)	This molecule meets the cyclic requirements for aromaticity
Two nitrogen atoms adjacent to two π bonds	A lone pair of electrons on a second-row atom adjacent to a π bond will typically be conjugated and sp^2 hybridized (vinylic)	The lone pairs on nitrogen atoms can be conjugated to the π bonds, making a fully conjugated system
There are 2 π bonds and 2 lone pairs in the cyclic system (8 π electrons)	If a cyclic, conjugated, planar system contains an even number of π electron pairs, the molecule is antiaromatic (high energy)	If the molecule is planar, it will be antiaromatic
The problem states the molecule is non-aromatic	Cyclooctatetraene adopts a non-planar "tub" conformation to avoid antiaromaticity	This molecule is not planar to avoid antiaromaticity
Diagnosis: The molecule is non-aromatic because the ring is not planar.		

FIGURE 3
Diagnosis of an aromaticity problem.

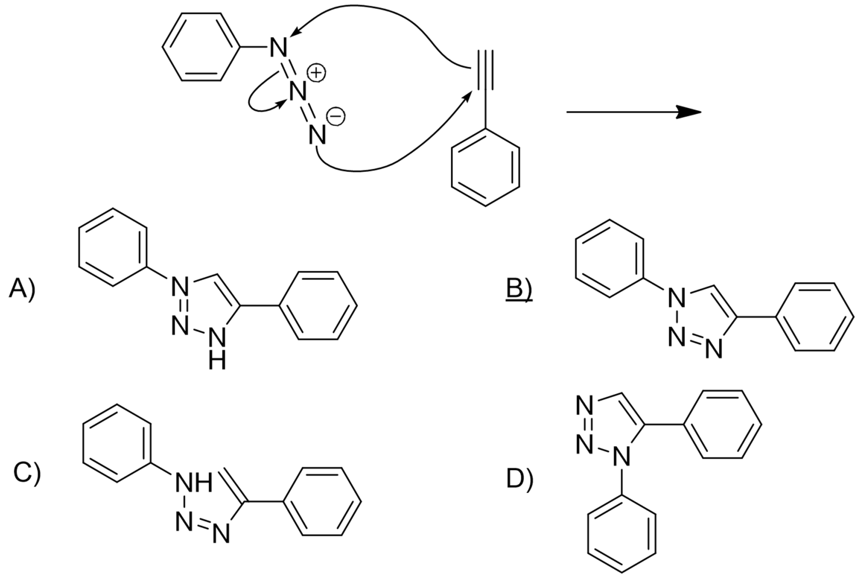
recognition of the implicit lone pairs on the nitrogen atoms and the carbon–carbon π bonds, recall of the requirements of aromaticity, and then application of abductive reasoning to the concepts learned (e.g., in class) and perceived by the structural representation. It is easy to see that the 1,4-dihydropyrazine is cyclic, has 8 π electrons, and a *p* orbital at each vertex. However, this simple analysis would result in the structure being anti-aromatic, so the student must recognize that in order for it to be non-aromatic as the problem states, planarity must be disrupted.

The second example shown in Figure 4 is a curved arrow mechanism problem for a reaction not typically covered in the Organic Chemistry course sequence (Sarode et al., 2016). Students must apply the rules of curved arrows and properly atom map to diagnose the correct product.[7] The pre-existing conditions for a mechanism question with arrows shown include the nature of curved arrows and the examination of the scheme will require atom mapping and keeping track of which bonds are broken and formed.

The third example shown in Figure 5 is a substitution/elimination problem (Brown et al., 1956). Students frequently find these reactions challenging and may employ a variety of heuristic models to approach them. Just as medical diagnosis begins with gathering information (taking patient information), solving this problem begins with direct observation and application of what is known about the structure and reactivity of these molecules. The alkyl halide has a good leaving group and has tertiary electrophilic carbon classification while the *t*-butoxide reagent is electron-rich, bulky, and reactive. Students must reason abductively how these characteristics interact with each other. This iterative process first eliminates S_N2 due to the nature of the alkyl halide, then identifies E2 as the mechanism with the bulky alkoxide. Next, an understanding of thermodynamics vs. kinetics to differentiate the two possible E2 pathways. Finally, a re-examination of the problem indicates the less stable product is formed preferentially; this is *best explained* by steric crowding in the transition state of the reaction between the alkyl halide and alkoxide.

7 While one could argue that the diagnosis/answer to the problem presented in Figure 4 does not require abductive reasoning, we have included it because the skills required here can be applied to more complex problems that, for example, include mechanistic reasoning (*vide infra*).

Which of the following is the product of the reaction mechanism (curved arrows)?

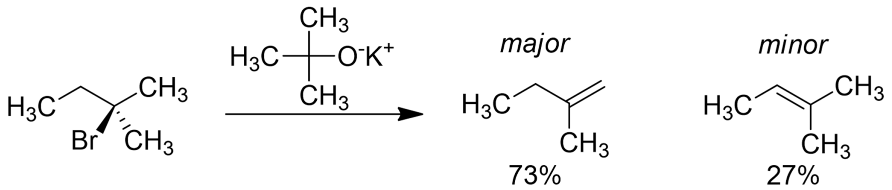


Evidence & Observations (What you see)	Premises (What you know)	Reasoned Conclusion (Hypothesis)
There are 3 curved arrows	Curved arrows represent the movement of a pair of electrons	There should be 3 bonds made/broken
Arrow 1 starts at the negative nitrogen and ends at the alkyne	Negatively charged species are good nucleophiles	The negative nitrogen is the nucleophile and the alkyne is the electrophile
Arrow 1 starts at the negative nitrogen and ends at the alkyne	If an arrow starts at a lone pair and ends at an atom, a new σ bond is formed	A bond is formed between nitrogen and carbon
Arrow 2 starts at the alkyne π bond and ends at the neutral nitrogen	If an arrow starts at a π bond and ends at an atom, the π bond is broken and a new σ bond is formed	A bond is formed between carbon and nitrogen
Arrow 3 starts at the diaza π bond and ends at the positive nitrogen	If an arrow starts at a π bond and ends at an atom, it becomes a lone pair	The central nitrogen is now neutral with a lone pair
There are 5 atoms involved in the 3 arrows	The atoms involved in bond formation need to be connected	A ring is formed
The arrows connect the 5 atoms in the ring in a specific order	Atoms in mechanisms must be mapped	The phenyl groups have a 1,3-relationship
There are no N-H bonds in the starting material*	No arrows involved hydrogen atoms	There are no N-H bonds in the product
Diagnosis: B is the correct product because it has the proper atoms and connectivity		

FIGURE 4
Diagnosis of a mechanism problem.

The final, and most complex, example shown in Figure 6 is a predict the product, addition reaction problem (Inoue and Murata, 1997) that is analogous to halohydrin formation. The problem requires separate diagnoses as it is layered where advancement to the second part is necessitated by the successful completion of the first addition step. Students would need to

Which of the following best explains the reaction?



A) The thermodynamic product predominates due to the stability of the product
B) The thermodynamic product predominates due to steric hinderance in the transition state
C) The kinetic product predominates due to the stability of the product
D) <u>The kinetic product predominates due to steric hinderance in the transition state</u>

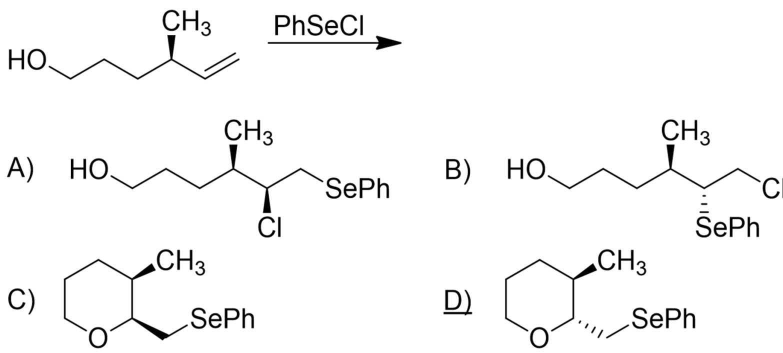
Evidence & Observations (What you see)	Premises (What you know)	Reasoned Conclusion (Hypothesis)
The organic reactant is a haloalkane	Haloalkanes typically react in substitution or elimination reactions	S_N & E reaction analyses will be needed
The haloalkane contains a bromine	S_N & E reactions require a good leaving group (i.e., I, Br, Cl, OTf, OTs, OMs)	The bromine will act as the leaving group in an S_N or E reaction
The leaving group is on a 3°, electrophilic carbon	Reactions at 3° carbons typically proceed via S_N1/E1/E2	An S_N2 reaction cannot occur
The reagent contains a negatively charged oxygen bonded to an alkane (alkoxide)	Alkoxides can react as strong nucleophiles or strong bases	An S_N2 or E2 reaction can occur
The alkane of the alkoxide is a *tert*-butyl group	Sterically bulky nucleophiles/bases will predominantly react as bases	An E2 reaction will occur
The minor product alkene is more substituted than the major product alkene	More substituted alkenes are more stable (thermodynamic product)	The minor product is the thermodynamic product
The haloalkane is tertiary and the alkoxide is bulky	The transition state will be sterically crowded	The lower energy transition state (kinetic product) leads to the major product
Diagnosis: The preferential formation of the less stable, major product proceeds with a lower energy of activation due to the steric bulk of the alkoxide acting as the base, thus the kinetic product is preferentially formed in this reaction.		

FIGURE 5
Diagnosis of a substitution/elimination question.

differentiate between the nucleophilicity of the alcohol and π bond after recognizing them as potential nucleophiles. After using abduction to recognize the higher reactivity of the π bond, students should then reason that selenium is electrophilic, akin to bromine in Br_2 due to being polarizable and bonded to a leaving group. This diagnosis is supported when taking into account the stereospecificity of the transformation, which precludes carbocation intermediates. The second diagnosis requires that students recall the regioselectivity of reactions with 3-membered cationic rings at the more substituted carbon. The remaining nucleophilic oxygen atom can now react with a higher energy seleniranium ion. However, conformational analysis of

10.3389/feduc.2024.1412417

Which of the following is the major product of the reaction?



Evidence & Observations (What you see)	Premises (What you know)	Reasoned Conclusion (Hypothesis)
The organic reactant is an alkene	Alkenes (π bonds) can react as nucleophiles	An alkene addition mechanism may be operative
The organic reactant is an alcohol	Alcohols (lone pair) can react as weak nucleophiles	The alkene is more reactive as a nucleophile than the alcohol
The reagent is an organoselenium compound	Selenium is large/polarizable (akin to bromine) and can form cationic 3-membered rings	A cationic, 3-membered ring containing selenium (seleniranium ion) can be formed
The reagent is a chloroselenium compound	Bonds between large atoms are weak (e.g., Br-Br)	Chloride may be a leaving group in this reaction
Diagnosis: The first step of the reaction involves the formation of a seleniranium ion via nucleophilic attack of the alkene on the chloroselenium electrophile		
The seleniranium ion is formally cationic	Cations are electron poor and electrophilic (e.g., bromonium ion)	The seleniranium ion can react as an electrophile to open the ring
The alcohol remains unreacted	The alcohol is nucleophilic (e.g., halohydrin formation)	The alcohol oxygen attacks the carbon bonded to selenium
This attack will be nucleophilic substitution	Weak nucleophiles attack carbons that bear more partial positive charge	The cationic 3-membered ring will be attacked at the more substituted carbon
This attack will be intramolecular substitution	The alcohol must attack from the backside of the C-Se bond	Attack will occur with less buildup of steric strain in the transition state
Diagnosis: The molecule shown in D will result from the three-step mechanism		

FIGURE 6
Diagnosis of a predict the product reaction.

the transition state is needed to discern the pseudo axial/equatorial approach of the oxygen atom on the seleniranium ion (Figure 6, bottom). Students would then need to apply their knowledge of chair conformations and the lower energy state when having ring substituents equatorial. Thus, the *trans*-oxacyclohexane is formed.

3 Conclusion and future work

While Organic Chemistry is often regarded as the most challenging undergraduate course in the US, we argue it has gotten a "bad rap" because students are not always prepared for the challenges that lie ahead when they enter the course. Students generally perform better on assessments when they employ metacognitive strategies (i.e., "thinking about thinking"). This has been demonstrated in a variety of courses (Arslantas et al., 2018), including Organic Chemistry (e.g., Graulich et al., 2021; Blackford et al., 2023). The consensus is that students who employ more metacognitive strategies in Organic Chemistry are more successful in problem-solving tasks and are better able to use those strategies when they are explicitly modeled and scaffolded. We have argued that instructors of Organic Chemistry should teach and demonstrate how to think and problem solve via "diagnosis" (i.e., abductive reasoning) in their classrooms. We hypothesize that students may score higher on metrics that assess scientific learning when these types of diagnostic models are utilized.

As constructors of nationally standardized exams, we fully acknowledge that a lot of growth on organic chemistry knowledge assessment still remains to be achieved. For Organic Chemistry course instructors, we hope the above insight into abductive reasoning can also be used on the assessment side of teaching requirements. Namely, that the cognitive load placed on students when solving each problem be carefully considered when constructing summative assessment items. Though this point has been frequently made previously (e.g., see Raker et al., 2013), we believe it is worthwhile for all writers of questions in Organic Chemistry to map out, step-by-step, the logic required to solve each question to determine the cognitive load. This can, in turn, help these instructors teach from a novice-focused perspective—as opposed to the "sage on the stage." The prior section provided examples with varying levels of complexity and demonstrated that cognitive load can be approximated by the number of reasoning steps (subtasks) required when the assessment piece is broken down. Further, this process could potentially also help the exam writer identify if items require little to no scientific reasoning (e.g., pure memorization questions).

The above manuscript merely outlines a hypothesis that we have generated over the course of our time teaching Organic Chemistry with this "diagnosis" method of abduction. To fully explore its validity, educational research is needed. This will be a precarious endeavor, because measuring the efficacy of teaching abductive reasoning will require assessment of scientific thinking skills in Organic Chemistry, and, as we just pointed out, there are already strong arguments that we are still quite far away from such valid assessments. However, we can be sure that if you are teaching Organic Chemistry from the perspective of your experience and expertise as an organic chemist, then opening a window for your students into how *you* think and problem solve will benefit your students. Our position is that instructors of Organic Chemistry should not only be explicitly teaching students the abductive reasoning skills to tackle complex problems, but they should also frame it as "diagnosing" the chemical situation.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JW: Writing – review & editing, Writing – original draft, Conceptualization. MW: Writing – review & editing, Writing – original draft. SZ: Writing – review & editing, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank Gautam Bhattacharyya for helpful discussions during revisions of this manuscript, specifically regarding perceptual learning theory. JW would like to acknowledge his undergraduate Organic Chemistry professor, Thomas Nalli, on the recent occasion of his 65th birthday for teaching him scientific problem-solving skills and fostering his interest in the discipline.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Achet, D., Rocrelle, D., Murengezi, I., Delmas, M., and Gaset, A. (1986). Reactions in slightly hydrated solid/liquid heterogeneous media: the methylation reaction with dimethyl sulfoxide. *Synthesis* 1986, 642–643. doi: 10.1055/s-1986-31729

Arslantas, F., Wood, E., and Macneil, S. (2018). "Metacognitive foundations in higher education chemistry" in International perspectives on chemistry education research and practice (American Chemical Society).

Bhattacharyya, G. (2013). From source to sink: mechanistic reasoning using the Electron-pushing formalism. *J. Chem. Educ.* 90, 1282–1289. doi: 10.1021/ed300765k

Blackford, K. A., Greenbaum, J. C., Redkar, N. S., Gaillard, N. T., Helix, M. R., and Baranger, A. M. (2023). Metacognitive regulation in organic chemistry students: how and why students use metacognitive strategies when predicting reactivity. *Chem. Educ. Res. Pract.* 24, 828–851. doi: 10.1039/D2RP00208F

Bodé, N. E., Deng, J. M., and Flynn, A. B. (2019). Getting past the rules and to the why: causal mechanistic arguments when judging the plausibility of organic reaction mechanisms. *J. Chem. Educ.* 96, 1068–1082. doi: 10.1021/acs.jchemed.8b00719

Brown, H. C., Moritani, I., and Okamoto, Y. (1956). Steric effects in elimination reactions. Vii. The effect of the steric requirements of alkoxide bases on the direction of bimolecular elimination. *J. Am. Chem. Soc.* 78, 2193–2197. doi: 10.1021/ja01591a047

Buick, P. E. (1995). An evaluation of organic chemistry as a prerequisite for admission to Florida medical schools. Doctorate, Waldon University.

Cartrette, D. P., and Bodner, G. M. (2010). Non-mathematical problem solving in organic chemistry. *J. Res. Sci. Teach.* 47, 643–660. doi: 10.1002/tea.20306

Christian, K., and Talanquer, V. (2012). Modes of reasoning in self-initiated study groups in chemistry. *Chem. Educ. Res. Pract.* 13, 286–295. doi: 10.1039/C2RP20010D

Connolly, K. (2017). Perceptual learning [online]. Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/sum2017/entries/perceptual-learning/ (Accessed 2024).

Cruz-Ramírez De Arellano, D., and Towns, M. H. (2014). Students' understanding of alkyl halide reactions in undergraduate organic chemistry. *Chem. Educ. Res. Pract.* 15, 501–515. doi: 10.1039/C3RP00089C

Dai, W.-Z., and Muggleton, S. (2021). Abductive knowledge induction from raw data. *arxiv:2010.03514v2*. doi: 10.48550/arXiv.2010.03514

Dixson, L., Pomales, B., Hashemzadeh, M., and Hashemzadeh, M. (2022). Is organic chemistry helpful for basic understanding of disease and medical education? *J. Chem. Educ.* 99, 688–693. doi: 10.1021/acs.jchemed.1c00772

Dood, A. J., and Watts, F. M. (2022). Mechanistic reasoning in organic chemistry: a scoping review of how students describe and explain mechanisms in the chemistry education research literature. *J. Chem. Educ.* 99, 2864–2876. doi: 10.1021/acs.jchemed.2c00313

Douven, I. (2021). How explanation guides belief change. *Trends Cogn. Sci.* 25, 829–830. doi: 10.1016/j.tics.2021.07.009

Finkenstaedt-Quinn, S. A., Watts, F. M., Petterson, M. N., Archer, S. R., Snyder-White, E. P., and Shultz, G. V. (2020). Exploring student thinking about addition reactions. *J. Chem. Educ.* 97, 1852–1862. doi: 10.1021/acs.jchemed.0c00141

Friel, D. D., and Chandar, K. (2021). Teaching diagnostic reasoning to medical students: a four-step approach. *Perspect. Biol. Med.* 64, 557–586. doi: 10.1353/pbm.2021.0041

Garg, N. K. (2019). How organic chemistry became one of Ucla's most popular classes. *J. Biol. Chem.* 294, 17678–17683. doi: 10.1074/jbc.AW119.008141

Gilbert, J. K. (2005). Visualization: A Metacognitive Skill in Science and Science Education. In: GILBERT, J. K. (ed.) *Visualization in Science Education.* Dordrecht: Springer Netherlands.

Goodwin, W. (2003). Explanation in organic chemistry. *Ann. N. Y. Acad. Sci.* 988, 141–153. doi: 10.1111/j.1749-6632.2003.tb06093.x

Graulich, N. (2015). The tip of the iceberg in organic chemistry classes: how do students deal with the invisible? *Chem. Educ. Res. Pract.* 16, 9–21. doi: 10.1039/C4RP00165F

Graulich, N., Langner, A., Vo, K., and Yuriev, E. (2021). "Scaffolding metacognition and resource activation during problem solving: a continuum perspective" in Problems and problem solving in chemistry education: Analysing data, looking for patterns and making deductions. ed. G. Tsaparlis (The Royal Society of Chemistry).

Habraken, C. L. (1996). Perceptions of chemistry: Why is the common perception of chemistry, the most visual of sciences, so distorted?. *J sci educ technol,* 5, 193–201.

Harman, G. H. (1965). The inference to the best explanation. *Philos. Rev.* 74, 88–95.

Higgins, T. S., and Reed, S. F. (2007). Changing premedical requirements. *JAMA* 297, 37–39. doi: 10.1001/jama.297.1.37-b

Hmelo-Silver, C. E., Marathe, S., and Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: expert: novice understanding of complex systems. *J. Learn. Sci.* 16, 307–331. doi: 10.1080/10508400701413401

Hulleman, C. S., Godes, O., Hendricks, B. L., and Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *J. Educ. Psychol.* 102, 880–895. doi: 10.1037/a0019506

Inoue, H., and Murata, S. (1997). Novel cyclization of unsaturated alcohols by phenyl selenocyanate in the presence of copper bis(trifluoromethanesulfonate). *Heterocycles* 45, 847–850.

Jin, X., Li, S., Guo, L., Hua, J., Qu, D.-H., Su, J., et al. (2022). Interplay of steric effects and aromaticity reversals to expand the structural/electronic responses of Dihydrophenazines. *J. Am. Chem. Soc.* 144, 4883–4896. doi: 10.1021/jacs.1c12610

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *J. Comput. Assist. Learn.* 7, 75–83. doi: 10.1111/j.1365-2729.1991.tb00230.x

Kellman, P. J., and Massey, C. M. (2013). "Chapter four - perceptual learning, cognition, and expertise" in Psychology of learning and motivation. ed. B. H. Ross (Academic Press).

Kim, T., Wright, L. K., and Miller, K. (2019). An examination of students' perceptions of the Kekulé resonance representation using a perceptual learning theory lens. *Chem. Educ. Res. Pract.* 20, 659–666. doi: 10.1039/C9RP00009G

Kovac, J. (2002). Theoretical and practical reasoning in chemistry. *Found. Chem.* 4, 163–171. doi: 10.1023/A:1016035726186

Kozma, R. B., and Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching,* 34, 949–968.

Kraft, A., Strickland, A. M., and Bhattacharyya, G. (2010). Reasonable reasoning: multi-variate problem-solving in organic chemistry. *Chem. Educ. Res. Pract.* 11, 281–292. doi: 10.1039/C0RP90003F

Lipton, P. (2017). "Inference to the best explanation" in A companion to the philosophy of science. ed. W. H. Newton-Smith (Blackwell).

Martini, C. (2023). "Abductive reasoning in clinical diagnostics" in Handbook of abductive cognition. ed. L. Magnani (Cham: Springer International Publishing). doi: 10.1007/978-3-031-10135-9_13

Moran, B. (2013). How to get an A- in organic chemistry. *New York Times.*

Morris Dye, K., and Dangremond Stanton, J. (2017). Metacognition in upper-division biology students: awareness does not always Lead to control. *CBE Life Sci. Educ.* 16:ar31. doi: 10.1187/cbe.16-09-0286

Overton, T., and Potter, N. (2008). Solving open-ended problems, and the influence of cognitive factors on student success. *Chem. Educ. Res. Pract.* 9, 65–69. doi: 10.1039/B801307C

Pareschi, R. (2023). Abductive reasoning with the Gpt-4 language model: case studies from criminal investigation, medical practice, scientific research. *Sistemi Intelligenti* 35, 435–444. doi: 10.48550/arXiv.2307.10250

Platt, J. R. (1964). Strong inference. *Science* 146, 347–353. doi: 10.1126/science.146.3642.347

Pribyl, J. R., and Bodner, G. M. (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching,* 24, 229–240.

Price, A. M., Kim, C. J., Burkholder, E. W., Fritz, A. V., and Wieman, C. E. (2021). A detailed characterization of the expert problem-solving process in science and engineering: guidance for teaching and assessment. *CBE Life Sci. Educ.* 20:ar43. doi: 10.1187/cbe.20-12-0276

Raker, J. R., and Towns, M. H. (2010). Benchmarking problems used in second year level organic chemistry instruction. *Chem. Educ. Res. Pract.* 11, 25–32. doi: 10.1039/C001043J

Raker, J. R., Trate, J. M., Holme, T. A., and Murphy, K. (2013). Adaptation of an instrument for measuring the cognitive complexity of organic chemistry exam items. *J. Chem. Educ.* 90, 1290–1295. doi: 10.1021/ed400373c

Randles, C. A., and Overton, T. L. (2015). Expert vs. novice: approaches used by chemists when solving open-ended problems. *Chem. Educ. Res. Pract.* 16, 811–823. doi: 10.1039/C5RP00114E

Reid, N. (2020). The triangle model: the contribution of the late professor Alex H. Johnstone. *J. Sci. Educ.* 2, 47–61.

Sarode, P. B., Bahekar, S. P., and Chandak, H. S. (2016). Dabco/Acoh jointly accelerated copper(I)-catalysed cycloaddition of Azides and alkynes on water at room temperature. *Synlett* 27, 2681–2684. doi: 10.1055/s-0036-1588590

Stowe, R. L., and Cooper, M. M. (2017). Practicing what we preach: assessing "critical thinking" in organic chemistry. *J. Chem. Educ.* 94, 1852–1859. doi: 10.1021/acs.jchemed.7b00335

Stowe, R. L., and Cooper, M. M. (2019a). Arguing from spectroscopic evidence. *J. Chem. Educ.* 96, 2072–2085. doi: 10.1021/acs.jchemed.9b00550

Stowe, R. L., and Cooper, M. M. (2019b). Assessment in chemistry education. *Israel J. Chem.* 59, 598–607. doi: 10.1002/ijch.201900024

Tiettmeyer, J. M., Coleman, A. F., Balok, R. S., Gampp, T. W., Duffy, P. L., Mazzarone, K. M., et al. (2017). Unraveling the complexities: an investigation of the factors that induce load in chemistry students constructing Lewis structures. *J. Chem. Educ.* 94, 282–288. doi: 10.1021/acs.jchemed.6b00363

Van Dantzig, S., Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cogn. Sci.* 32, 579–590. doi: 10.1080/03640210802035365

Vertue, F. M., and Haig, B. D. (2008). An abductive perspective on clinical reasoning and case formulation. *J. Clin. Psychol.* 64, 1046–1068. doi: 10.1002/jclp.20504

Wackerly, J. W. (2021). Abductive reasoning in organic chemistry. *J. Chem. Educ.* 98, 2746–2750. doi: 10.1021/acs.jchemed.1c00295

Walker, J. P., Van Duzor, A. G., and Lower, M. A. (2019). Facilitating argumentation in the laboratory: the challenges of claim change and justification by theory. *J. Chem. Educ.* 96, 435–444. doi: 10.1021/acs.jchemed.8b00745

Watts, F. M., Schmidt-Mccormack, J. A., Wilhelm, C. A., Karlin, A., Sattar, A., Thompson, B. C., et al. (2020). What students write about when students write about mechanisms: analysis of features present in students' written descriptions of an organic reaction mechanism. *Chem. Educ. Res. Pract.* 21, 1148–1172. doi: 10.1039/C9RP00185A

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 14 October 2024
DOI 10.3389/feduc.2024.1355078

Check for updates

OPEN ACCESS

EDITED BY
Gautam Bhattacharyya,
Missouri State University, United States

REVIEWED BY
Pamela Lundin,
High Point University, United States
José Hernández-Ramos,
Metropolitan University of Educational Sciences, Chile

*CORRESPONDENCE
Katrin Schuessler
✉ katrin.schuessler@uni-due.de

RECEIVED 13 December 2023
ACCEPTED 23 September 2024
PUBLISHED 14 October 2024

CITATION
Schuessler K, Striewe M, Pueschner D, Luetzen A, Goedicke M, Giese M and Walpuski M (2024) Developing and evaluating an e-learning and e-assessment tool for organic chemistry in higher education. *Front. Educ.* 9:1355078.
doi: 10.3389/feduc.2024.1355078

Developing and evaluating an e-learning and e-assessment tool for organic chemistry in higher education

Katrin Schuessler[1]*, Michael Striewe[2], Daniel Pueschner[3], Arne Luetzen[3], Michael Goedicke[2], Michael Giese[4] and Maik Walpuski[1]

[1]Department for Chemistry Education, University of Duisburg-Essen, Essen, Germany, [2]Paluno – The Ruhr Institute of Software Technology, University of Duisburg-Essen, Essen, Germany, [3]Kekulé-Institute of Organic Chemistry and Biochemistry, University of Bonn, Bonn, Germany, [4]Department for Organic Chemistry, University of Duisburg-Essen, Essen, Germany

A convincing e-learning system for higher education should offer adequate usability and not add unnecessary (extraneous) cognitive load. It should allow teachers to switch easily from traditional teaching to flipped classrooms to provide students with more opportunities to learn and receive immediate feedback. However, an efficient e-learning and technology-enhanced assessment tool that allows generating digital organic chemistry tasks is yet to be created. The Universities of Bonn and Duisburg-Essen are currently developing and evaluating an e-learning and technology-enhanced assessment tool for organic chemistry. This study compares the effectiveness of traditional paper-pencil-based and digital molecule-drawing tasks in terms of student performance, cognitive load, and usability—factors that all contribute to learning outcomes. Rasch analysis, *t*-tests, and correlation analyses were used for evaluation, revealing that the developed system can generate digital organic chemistry tasks. Students performed equally well on simple digital and paper-pencil molecule-drawing tasks when they received an appropriate introduction to the digital tool. However, using the digital tool in two of three studies imposes a higher extraneous cognitive load than using paper and pencil. Nevertheless, the students rated the tool as sufficiently usable. A significant negative correlation between extraneous load and tool usability was found, suggesting room for improvement. We are currently concentrating on augmenting the functionality of the new e-learning tool to increase its potential for automatic feedback, even for complex tasks such as reaction mechanisms.

KEYWORDS

e-learning system, technology-enhanced assessment, higher education, cognitive load, organic chemistry, feedback

1 Introduction

In recent decades, the success and dropout rates in higher education have attracted particular interest (Aulck et al., 2016; Fischer et al., 2021; Fleischer et al., 2019; Heublein et al., 2020). For chemistry students, prior knowledge (in addition to satisfaction with the study content) is especially important for success (Fischer et al., 2021; Fleischer et al., 2019). It is therefore particularly important to promote the knowledge of students with little prior knowledge. This can be achieved by providing principle-based (feedback that provides learners with an explanation of why an answer is right or wrong, rather than just informing them whether an answer is right or wrong), just-in-time feedback (Eitemüller et al., 2023). In addition to knowledge of content, mastery of representations (e.g., routine use of skeletal formula) is essential for understanding organic chemistry. Representational competence mediates the relationship between prior knowledge and content knowledge in organic chemistry (Dickmann et al., 2019). Research indicates that students face challenges in producing representations at a very basic level such as drawing the correct molecule when given an IUPAC (International Union of Pure and Applied Chemistry) name (Bodé et al., 2016; Farhat et al., 2019) and connecting representations with relevant concepts (e.g., Anzovino and Lowery Bretz, 2016; Asmussen et al., 2023; Graulich and Bhattacharyya, 2017). Instructional support that promotes the integration of explanations and representations can enhance learning outcomes (Rodemer et al., 2021). Furthermore, the duration and level of previous chemistry education has been found to be a significant predictor for representational competence at university (Taskin et al., 2017).

Organic chemistry can be very complex, and hence, demands a high degree of the learners' working memory capacity to cope with it. To prevent cognitive overload, additional extraneous load, e.g., due to the processing of unfavorable instructions should be avoided to not impair learning (Paas and Sweller, 2014). Therefore, an e-learning system with sufficient usability that does not initiate extraneous processing through confusing operation is needed to enhance learning, by enabling learners to work independently on tasks at their own pace and receive personalized and timely error-related feedback. This type of learning system would present learners who have lower prior knowledge with an opportunity to keep up with those who have higher prior knowledge while providing all learners with more opportunities to practice the skills they have acquired. Additionally, implementing such an e-learning system could facilitate a shift from traditional courses that promote a one-size-fits-all approach to a flipped classroom, where students have more opportunities to engage in dialogue, receive feedback, and interact with educators (van Alten et al., 2019). However, an effective non-commercial and adequate e-learning and e-assessment tool that can be integrated into learning platforms such as Moodle or Ilias, allows teachers to create digitally those typical organic chemistry tasks they need for their courses, can evaluate student responses automatically and provide explanatory feedback; and whose handling does not induce extraneous cognitive load, that is, offers good usability, is still missing. This paper offers insights into the primary stages of the development and evaluation of an e-learning and e-assessment tool for organic chemistry. First, we summarize the findings of related studies. Next, we present an overview of the functionality of the developed learning tool. Then, we describe the design of the evaluation studies before presenting their results, and finally, discuss our findings.

1.1 Challenges in learning organic chemistry

Students' prior knowledge significantly predicts their success in chemistry studies (Fischer et al., 2021; Fleischer et al., 2019). Although students' knowledge of organic chemistry generally increases over the course of a semester, regardless of their prior knowledge (Averbeck, 2021), learners with low prior knowledge do not learn enough to catch up, so prior knowledge remains a significant predictor of course grade in basic organic chemistry courses. Thus, it is possible to identify students at a high risk of not passing the course through prior knowledge assessment (Hailikari and Nevgi, 2010). Research suggests that novices in organic chemistry tend to focus on surface structures of representations (such as geometric shapes) and therefore have difficulty establishing the relationships between structural features and concepts (Anzovino and Lowery Bretz, 2016). Furthermore, dependence on rote learning instead of meaningful learning (Grove and Lowery Bretz, 2012) and focus on the surface level of representations (Graulich and Bhattacharyya, 2017) appear to be problematic.

An analysis of students' difficulties in predicting products or comparing mechanisms using a modified version of Bloom's taxonomy provides an overview of the various types of difficulties that students encounter (Asmussen et al., 2023). While difficulties with factual knowledge are limited to the cognitive process of remembering, difficulties with conceptual knowledge arise in the cognitive processes of remembering, understanding, applying, analyzing, and evaluating. If students encounter difficulties in analyzing concepts, they may not be able to identify the relevant parts of a molecule or deduce relevant concepts from the representation. Consequently, even if information is depicted, students may not be able to use it because of their inability to comprehend it. Contrary to the other difficulty categories, this category leads to no or incorrect solutions for the task. Thus, it is essential to encourage students to use representations as useful tools, instead of viewing them as a burden. An analysis of student quizzes at the beginning of an advanced organic chemistry course revealed that approximately one-third of the students encountered challenges in drawing Lewis structures based on the name of the molecule. Of the errors detected, the prominent ones included (1) failure to draw a Lewis structure, instead of opting for some structural formula that often violated the octet rule, and (2) drawing molecules with incorrect carbon chains of varying lengths (Farhat et al., 2019). Further evidence suggests that students struggle with fundamental skills in organic chemistry, such as identifying functional groups, drawing molecules from their IUPAC names, and drawing meaningful structures that adhere to the octet rule (Bodé et al., 2016). Evidence shows that representational competence mediates the relationship between prior and content knowledge in organic chemistry (Dickmann et al., 2019). Research also suggests that instructional support, which facilitates the integration of explanations and representations, mediates learning outcomes (Rodemer et al., 2021). Furthermore, the duration and level of previous chemistry education has been found to be a significant predictor for representational competence at university (Taskin et al., 2017).

In summary, although representations are crucial in organic chemistry, evidence suggests that students face difficulties in using and comprehending basic representations. Consequently, students experience problems when attempting more advanced tasks such as reaction mechanisms, as these rely on the ability to use and comprehend representations. Previous studies have demonstrated the importance of prior knowledge for achieving success in the field of organic chemistry. It is thus crucial for students lacking prior knowledge to receive adequate instructional support to bridge their knowledge gaps and prevent them from dropping out.

This paper uses tasks that ask students to draw molecules based on their IUPAC names to explore the differences between paper-based and digital tasks as part of an initial evaluation of an e-learning and e-assessment tool that is currently being developed

1.2 Cognitive load and schema construction

Cognitive Load Theory concerns learning domains with high intrinsic cognitive load (Paas et al., 2005) and considers the limitation of working memory as a critical factor in creating learning materials (Sweller et al., 1998). Learners acquire cultural knowledge that has to be acquired with conscious effort, such as organic chemistry by actively constructing mental models incorporating previously attained knowledge and presented information (Paas and Sweller, 2014). Acquired knowledge can be stored in long-term memory without time or quantity limitations in the form of cognitive schemata. Schemata serve to organize not only pieces of information, but also the relations between them. Basic schemata can be combined to form more complex schemata (Paas et al., 2003a). Unlike long-term memory, working memory has a strictly limited capacity, meaning that only a few pieces of new information can be processed simultaneously in working memory (Paas and Sweller, 2014). Despite its complexity, a schema that transfers from long-term memory into working memory is considered a single element; therefore, it is not limited by working-memory constraints (Sweller et al., 1998). Hence, schemata not only organize and store knowledge, but also serve to relieve working memory. Within a given domain, experts and novices diverge in the progression and automation of domain-related schemata (Sweller, 1988). Experts use highly developed and automated schemata that enable them to perform complex tasks. Consequently, to acquire cultural knowledge such as organic chemistry, one should aim to develop coherent and highly automated cognitive schemata.

For the successful acquisition of cultural knowledge, a sufficient cognitive load during learning is crucial (Paas et al., 2005). The processing of new information in the working memory induces a cognitive load (Paas and Sweller, 2014). Cognitive load theory distinguishes between the intrinsic cognitive load that arises from learning content and the extraneous load that results from unfavorable processing with regard to learning (Paas and Sweller, 2014; Sweller, 2010; Sweller et al., 1998; Sweller et al., 2019). Intrinsic cognitive load is determined by the number of elements that must be processed simultaneously and their interrelatedness. In the domain of organic chemistry, remembering the structural formula of ethanol may overload the working memory capacity of someone who is completely unfamiliar with chemistry, chemical representations, and structural formulas. Without a cognitive schema that assigns chemistry-specific meaning to the representation, one would need to process every letter and line and their spatial arrangement in the working memory to draw the structural formula for ethanol when requested. An organic chemistry expert can rely on a number of fully automated chemistry-specific cognitive schemata that include knowledge about the alkyl group and hydroxy group, element symbols, the octet rule that informs about plausible and implausible bonding, limitations of the structural formula as a form of representation, and other unrepresented information such as the three-dimensional arrangement of molecular components, hybridization, and possibilities for rotation around sigma bonds. Therefore, learners' prior knowledge plays a crucial role in their individual intrinsic cognitive load. Higher prior knowledge allows learners to process more information simultaneously because they have already constructed fundamental schemata on which they can rely and expand, whereas learners with lower prior knowledge have fewer rudimentary or non-automated chemistry-specific schemata on which to rely (Paas et al., 2003b). For effective learning, it is essential that the learning material is not overly complex and aligns with the learner's prior knowledge to prevent an overwhelming intrinsic cognitive load. In the field of organic chemistry, learners need a good grasp of specific concepts such as the representation of atoms by element symbols, the octet rule, basic functional groups, hybridization, and molecular orbitals, which enable them to draw, for example, the Lewis structures of organic molecules without risking intrinsic cognitive overload by trying to remember meaningless letters and strings. These schemata are necessary to advance further skills such as understanding and predicting reaction mechanisms.

Extraneous cognitive load arises from the inappropriate processing of learning material, such as providing unnecessary and distracting information or misdirecting learners' attention. An example of distracting information is information on the effects of alcohol intake on humans when asked to recall the name and draw the Lewis structure of the corresponding molecules of the first 10 primary n-alcohols. An example of distracting attention at this stage and in the context of the task might be informing learners about alternative representations or about the inaccurate representations of the bond angle by the structural formula. As the extraneous cognitive load competes with the intrinsic cognitive load, reducing the extraneous cognitive load is crucial for successful learning in complex fields such as organic chemistry. Minimizing these distractions optimizes the load imposed on the working memory and releases the working memory capacity for schema construction and automation, which is essential for effective learning.

In summary, learning in complex domains, such as organic chemistry, induces a high intrinsic cognitive load. Learners require working memory, which is limited, to proceed with the intrinsic load. After the initial construction of the cognitive schemata, learners also need time to practice and automate their schemata. Additional extraneous load can impair learning (Paas and Sweller, 2014). The present study therefore investigates the cognitive load of students when working on tasks in which molecules are to be drawn on the basis of their IUPAC names.

1.3 Requirements for an effective e-learning system

As previously explained, learning environments must strive to minimize learners extraneous load. This is also true for e-learning tools, where such a load can result from non-intuitive handling, splitting attention as a consequence of multiple or overlapping windows, or other technical hindrances to the solution of the primary task. Therefore, an adequate learning environment must provide sufficient usability. Additionally, e-learning systems should present further opportunities for learners to practice newly acquired skills and thus promote schema automation. An e-learning system that automatically evaluates students' responses and provides feedback beyond that on knowledge of results can enhance learning by offering explanatory feedback that provides principle-based explanations (Johnson and Priest, 2014). Explanatory feedback supports learning by (1) informing learners about the correctness of their answers, (2) identifying gaps in their existing knowledge, and (3) creating opportunities to adjust mental models. One advantage of e-learning systems over traditional written homework is the opportunity to offer immediate feedback to learners rather than delayed feedback after the teacher has corrected the homework. Empirical studies suggest that students value immediate feedback (Malik et al., 2014). Another benefit of an e-learning system over presenting a sample solution in class is its ability to provide principle-based feedback for errors that students have actually made instead of discussing common errors or misunderstandings. Eitemüller et al. (2023) demonstrated this advantage in a general chemistry course, in which students who received individual error-specific feedback outperformed those who received only corrective feedback, after adjusting for prior knowledge. Moreover, an e-learning system enables students to study independently and take advantage of individually paced learning sessions and the option to pause if necessary. Several studies on the segmenting principle have revealed that in addition to presenting meaningful segments that aid in organizing information, pausing and providing extra time to process information enhances learning (Mayer and Pilegard, 2014; Spanjers et al., 2012).

In summary, an e-learning system with adequate usability and a low extraneous cognitive load can facilitate learning by enabling learners to work individually on tasks at their own pace and receive personalized real-time feedback on errors. This will provide learners who have low prior knowledge with a chance to catch up with those who have higher prior knowledge, and thereby promote schema automation by presenting opportunities to practice. Additionally, an e-learning platform can support the transformation of conventional courses that rely on uniformity into flipped classrooms, which encourages increased opportunities for questioning, feedback, and interaction between students and teachers (van Alten et al., 2019). This article presents the results of initial evaluation studies carried out as part of the development of an e-learning tool for organic chemistry.

1.4 Digital learning tools for organic chemistry

A digital learning environment for organic chemistry should offer opportunities to create typical tasks for the subject and allow an automatic evaluation of student responses. Presently, transferring manually drawn molecular structures to computerized systems remains impossible (Rajan et al., 2020). Hence, digital learning tools for organic chemistry are restricted to using images for multiple-choice items (Da Silva Júnior et al., 2018) or must offer the possibility of digitally drawing molecules. Efforts have been made to develop such a platform (Malik et al., 2014; Penn and Al-Shammari, 2008). For example, the University of Massachusetts and University of Kentucky, which collaborated with Pearson, developed such a tool (Chamala et al., 2006; Grossman and Finkel, 2023). Another example is orgchem101, an e-learning tool offered by the University of Ottawa, which consists of modules that can stand alone or be combined with courses (Flynn, 2023; Flynn et al., 2014). Meanwhile, the University of California has developed the Reaction Explorer, which is currently distributed by Wiley PLUS (Chen and Baldi, 2008, 2009; Chen et al., 2010; WileyPLUS and University of California, 2023). However, no adequate, non-commercial e-learning and e-assessment tools currently exists that can be integrated into learning platforms such as Moodle or Ilias. This tool should allow teachers to create digitally their own organic chemistry tasks necessary for their courses, automatically assess students' responses, provide explanatory feedback, and not induce an extraneous cognitive load through its use, thus providing good usability.

1.5 Development goals and research questions

The Universities of Bonn and Duisburg-Essen are currently developing and evaluating an e-learning and e-assessment tool for organic chemistry.

Hence, our goal (G) is to develop and evaluate an e-learning and e-assessment tool with the following characteristics:

> *G1*: Is non-commercial,
>
> *G2*: Can be integrated into learning platforms like Moodle or Ilias,
> *G3*: Allows teachers to create digitally organic chemistry tasks that are typical and needed for their courses,
>
> *G4*: Can evaluate student responses automatically, provide explanatory feedback, and offer students additional opportunities to study at their own pace, and
>
> *G5*: Does not induce an extraneous cognitive load due to its handling, thus providing good usability.

The development of the tool followed the instructional path that students take when learning organic chemistry. Therefore, we developed a digital tool that can perform simple drawing tasks for organic molecules. This study's objective was to evaluate the tool and establish guidelines for the implementation of complex tasks involving chemical reaction equations, mechanisms, and the transformation of molecule A to molecule B. It evaluated the (1) performance of students on simple molecule-drawing tasks using paper and pencil versus the digital tool, (2) perceived cognitive load of the students, and (3) usability of the tool. Goals 3 and 4 were therefore not the subject of this first evaluation study.

We examined the following research questions (RQ):

> *RQ1*: To what extent are students' abilities to solve digital molecule-drawing tasks comparable with their abilities to solve paper-pencil-based molecule-drawing tasks?
>
> *RQ2*: To what extent do cognitive load ratings for digital molecule-drawing and paper-pencil-based molecule-drawing tasks differ?
>
> *RQ3*: How do students rate the usability of the developed tool?

We assume that the digital format has several characteristics that can be decisive for learning. We therefore assume that a comparative media approach (Buchner and Kerres, 2023) can be helpful in gaining valuable insights into changing processes. We therefore explicitly do not use the comparative media approach based on a purely technology-centered understanding of teaching and learning. The first implication is that digital tasks necessitate the use of skeletal formulae rather than allow students to choose the formula they wish to use. The ability to use the skeletal formula without difficulty is a learning objective for a basic course in organic chemistry. The skeletal formula is a rather abstract representation in which a lot of essential information is not explicitly represented (e.g., carbon and hydrogen atoms or free electron pairs), compared to the valence structure formula or the Lewis structure, for example. This level of abstraction can pose a challenge to novice learners who struggle to read this representation. In addition, learners challenged by the task may be overwhelmed by the additional requirement of using the tool rather than drawing their ideas directly. Moreover, the use of a digital tool could lead to a tool-driven solution (e.g., avoiding repeated switching between required functions such as the option to draw single bonds or the option to insert heteroatoms). It is not necessary to mentally plan the drawing in paper format. Learners can solve tasks here in a content-driven way by developing their drawing step by step along the name of the molecule. Furthermore, the digital tool evaluated in this study lacks a notepad function for drawing or elaborating components before assembling them. Additionally, there is no option to cross off already-named components. Finally, the digital tool automatically adds or removes hydrogen atoms if the octet rule is violated, which can confuse students who are not aware of the violation and distract them from drawing. Hence, we considered that several aspects relevant to learning and performance differ between media and therefore assumed that the media comparison approach would be helpful in investigating whether learning or performance differs between media. Moreover, because we assumed that any differences in media could potentially affect cognitive load, we decided to collect data on cognitive load. Notably, any potential differences in performance or cognitive load are not a direct result of the medium used but rather of the changes in task processing presented above. Should this study yield evidence of disparities in students' abilities or cognitive load across different media, further investigation is necessary to establish which of the aforementioned factors is responsible for these differences (as the media itself is not regarded as a causal factor).

2 The e-learning and e-assessment tool called JACK

JACK (Striewe, 2016) is an e-learning and e-assessment tool developed at the University of Duisburg-Essen. Unlike general learning management systems (LMS) that are designed to perform general tasks related to running courses (e. g. group formation, dissemination of learning materials, provision of simple quizzes), JACK is specialized for conducting formative and summative assessments with complex assessment items and individual, instant feedback. Nevertheless, JACK can be integrated into any LMS that supports the Learning Tools Interoperability (LTI) Standard. To do so, teachers define learning activities within the LMS that actually serve as links to the JACK server. Students clicking on these links automatically get logged in to JACK and forwarded to the set of exercises or assignments defined by the teacher. Elaborate feedback is displayed directly in JACK, while result points can also be reported back to the LMS for further processing.

To develop an e-learning and e-assessment environment for organic chemistry, JACK was complemented with Kekule.js (Jiang et al., 2016), an open-access web-based molecular editor available under the MIT license. The editor is based on current web-technologies (i. e. HTML5) and compatible with all major browsers on desktop and mobile clients. This enhancement enables us to work on a new type of task (called: *molecule*) within JACK, which allows the drawing of organic molecules in skeletal formulas and automatic checking of whether a student's response is correct. Teachers can create molecule-based tasks using JACK. For instance, these tasks can require learners to draw a single molecule (e.g., "Draw the molecule described by the following IUPAC-name: *(S,S)*-pentane-2,4-diol," Figure 1, upper left) or multiple molecules (e.g., "Draw all isomers of 1,2-dibromocyclohexane, including stereoisomers"). To evaluate student responses automatically, teachers can provide a sample solution by drawing the expected molecules in a Kekule.js window in the feedback section of JACK. Feedback for correct (e.g., "Your answer is correct."; Figure 1, bottom right) and incorrect ("Unfortunately, your answer is incorrect. Please try again.") answers as students respond may be included. E-learning teachers can also record further optional feedback rules; for example, to check student responses for typical mistakes (e.g., whether a student has drawn *(R,S)*-pentane-2,4-diol, Figure 1, upper right, or just pentan-2,4-diol without providing further information regarding stereochemistry) and provide principle-based feedback to their students to correct typical mistakes (Johnson and Priest, 2014).

The recorded sample solution and student responses are transformed from graphical information (drawn molecules) into InChI codes (Heller et al., 2015; IUPAC, 2023), which are standardized strings. The InChI code (international chemical identifier) derived from the student response is compared with the recorded sample solution and with additional feedback rules; in the case of correspondence, the respective feedback is provided to the student. If there is no correspondence, feedback for an incorrect solution is provided to the student. The transformation from graphical information to InChI allows comparing the essential features of molecules such as stereochemistry, and is independent of extraneous features such as writing direction. The original drawing is also stored in JACK in a machine-readable format. Hence, students and teachers can always inspect the original drawings when reviewing and discussing responses. In addition, JACK can further analyze the drawing to search for features that cannot be expressed in InChI codes, such as the arrangement of molecules around the reaction arrows.

Hence, integrating Kekule.js into JACK enables the design of tasks that require students to draw organic molecules, thus automatically comparing their responses to correct sample solutions or well-known mistakes, and providing feedback. JACK offers additional types of tasks such as fill-in-the-blanks and multiple-choice tasks, as well as various course settings to create appropriate learning or assessment courses.

3 Methods

This section discusses the development of the aforementioned e-learning and e-assessment tool and presents the framework for the evaluation studies.

3.1 Framework for the evaluation studies

The initial version of the tool was available in the summer of 2022 and underwent an initial evaluation. Throughout the subsequent year,

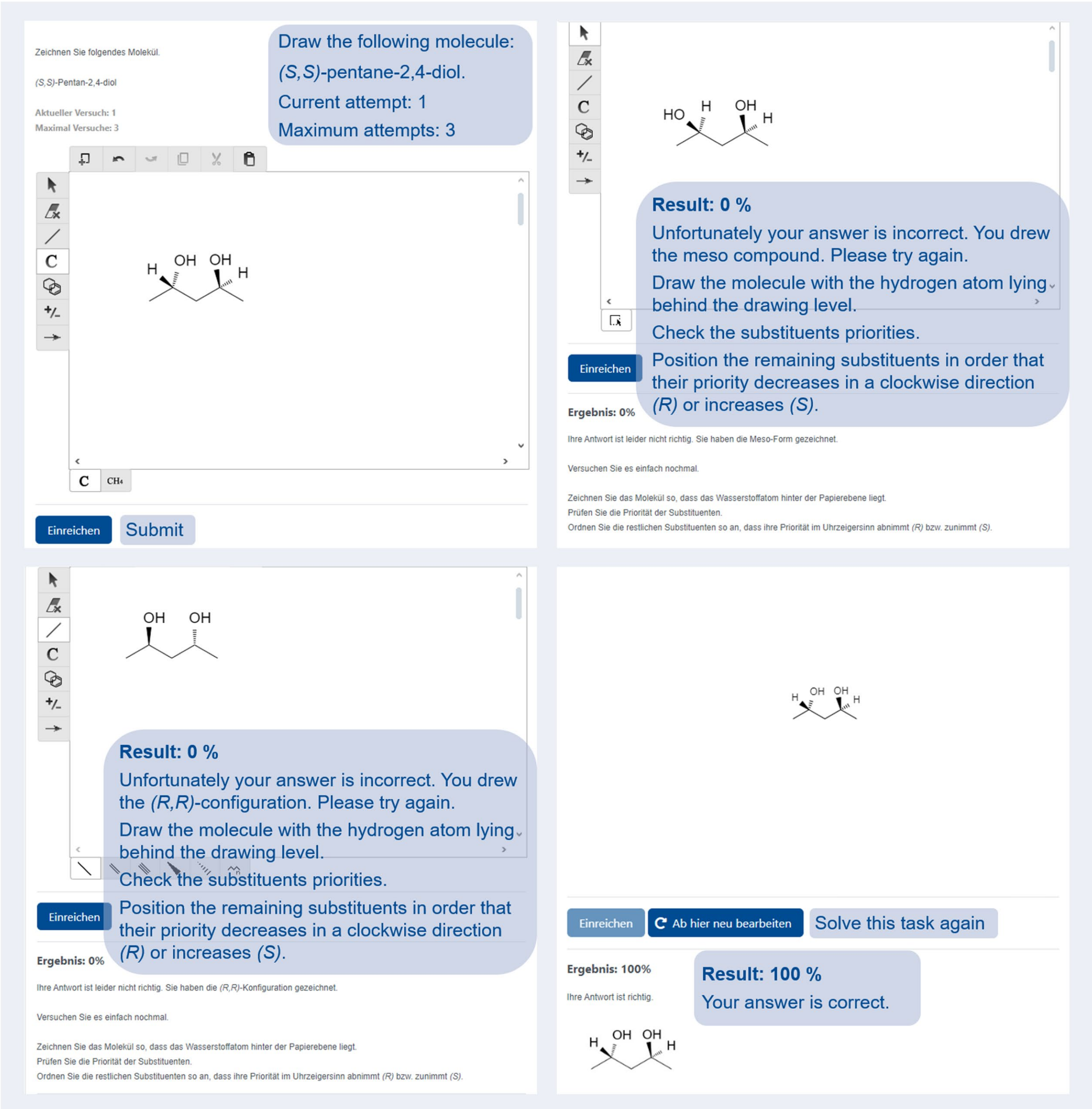


FIGURE 1
Screenshots from JACK. The figure shows a task (top left), two incorrect student answers with principle-based feedback (top right and bottom left), and a correct answer with a sample solution (bottom right). English translation of the originally German text is presented in blue.

the system became more sophisticated, following user feedback and feature requests from teachers. The e-learning mode of the tool (multiple attempts to solve the task, principle-based feedback, and knowledge of the results being available) was subsequently used for exercises in the beginner course on organic chemistry, and additional evaluations were conducted using the e-assessment mode (characterized by one attempt to solve the task, no principle-based feedback, and no knowledge of the results being available). All evaluation studies had the shared objective of examining (RQ1) to what extent students' abilities to solve digital molecule-drawing tasks compare to their abilities to solve paper-pencil-based molecule-drawing tasks (RQ2) to what extent cognitive load ratings for digital molecule-drawing and paper-pencil-based molecule-drawing tasks differ, and (RQ3) if their usability ratings were satisfactory. Our assumption for the study was that students who have the ability to draw molecules based on their IUPAC names should perform equally well in both formats. Cognitive load ratings were interpreted as indicators of perceived demand. In other words, comparable cognitive load ratings were interpreted in such a way that digital molecule-drawing tasks are perceived as being as demanding as paper-pencil molecule-drawing tasks and do not induce any further (extraneous) load (which might not change performance but might have a negative impact, for example, on motivation). Satisfactory usability ratings are those that are at least above the midpoint of the scale and are interpreted to mean that students do not dislike the digital molecule-drawing tasks.

For all studies, students were asked to participate in class. Students were informed that their participation was voluntary and did not affect their course grades. They were also informed about data protection and the use of their data for scientific research, and any questions were answered prior to data collection. At the end of each study, students were given the opportunity to ask questions again. Although IRB approval was not necessary at German universities, the guidelines concerning the ethical scientific practice of the Federal German Research Foundation (DFG) were applied.

To compare students' abilities to solve paper-pencil-based molecule-drawing tasks to digital molecule-drawing tasks, the following tasks in tandem with two very similar tasks were constructed. The digital task asks students to "Draw a 1,3-Dichloro-2-methylbutane molecule." The corresponding paper-pencil-based task asks students to "Draw a 1,2-Dichloro-2-methylpropane molecule." All tasks were originally constructed in German language. An overview of all IUPAC names used in study 4 is provided in Table 1. The students' answers were coded as correct (1) or incorrect (0). In the case of the paper-pencil format, this coding was performed by the first author of this study. The coding for the digital format was automatically done by the developed system.

Cognitive load ratings were collected using nine-point, labeled rating scales. For the first, second, and fourth studies, a German adaptation of the intrinsic and extraneous load items provided by Leppink et al. (2013) was used (Table 2) to measure delayed two-dimensional cognitive load (Schmeck et al., 2015; van Gog et al., 2012; Xie and Salvendy, 2000) after completing all molecule-drawing tasks in one format (paper-pencil-based or digital). Owing to the unsatisfactory reliability of the two-dimensional cognitive load rating scales in the third study, immediate cognitive load measures were subsequently collected after each molecule-drawing task using a unidimensional cognitive load instrument (Kalyuga et al., 2001).

TABLE 1 English translation of the originally German IUPAC names used for the item set of study 4.

Item number	Paper-pencil-based format	Digital format
1	Heptane	Octane
2	5-ethyl-4,4-dipropylnonane	5,5-dibutyl-4-propylnonane
3	Propanol	Ethanol
4	3-methylpentane-2,3-diol	2-methylpropane-1,3-diol
5	3-ethylhexanal	3-methylbutanal
6	2-hydroxy-3-methylbutanal	3-hydroxy-4-methylpentanal
7	Heptane-3,4,5-trione	Butane-2,3-dione
8	1,2-dichloro-2-methylpropane	1,3-dichloro-2-methylbutane
9	Butane-2-amine	Propane-2-amine
10	1-aminomethan-1-ol	2-aminoethan-1-ol
11	Hexanoic acid	Pentanoic acid
12	3-bromo-3,5-dichloro-1-nitro-4-propylheptane	3,5-dibromo-2-chloro-5-ethyl-1-nitrooctane
13	Ethanoic acid methyl ester	Methanoic acid ethyl ester
14	But-1-ene	But-2-yne

All items were constructed by adding the molecules' IUPAC-name to the following prompt: Draw a [fill in IUPAC-name]-molecule.

The usability ratings for the first and second studies were collected using nine-point, labeled rating scales and a German adaptation (Hauck et al., 2021) of the nine items originally provided by Brooke (1996). For the fourth study, a German version of the user experience questionnaire (UEQ, Laugwitz et al., 2008), which measures attractiveness (six items), perspicuity, efficiency, dependability, stimulation, and novelty (four items for each scale) using semantic differentials and seven-point rating scales, was administered.

The initial data analysis was performed using Winsteps (1-pl Rasch model, version 5.2.4.0; Boone et al., 2014). Data analysis was done separately for data from paper-pencil-based and digital material. Analyses in Winsteps were conducted in item-centered fashion. Rating scales were analyzed using a partial credit model. According to the Rasch model, there is a probabilistic relationship between the observed response behavior (item solution probability) and a latent characteristic (ability of the person) (Bond and Fox, 2007; Boone, 2016; Boone and Scantlebury, 2006). Within the Rasch model, the solution probabilities, ranging from 0 to 1, are transformed into logit values, having a value range from minus infinity to plus infinity (Boone and Staver, 2020; Boone et al., 2014). Typical measures range from +3 to −3. Easy items and people with lower personal abilities are indicated with negative values. Difficult items and individuals with higher person abilities are indicated with positive values (Boone et al., 2014; Linacre, 2023). In comparison to a raw sum score of correctly solved items, the person ability also considers whether the correctly solved items were easy or difficult items. An identical logit value for a person ability and an item difficulty corresponds to a 50% probability of solving an item by this person. A student who solves only very few and very easy items correctly will gain a low person ability. A student who solves the majority of items correctly and is even able to solve the difficult items will gain a high person ability. Hence, the person ability is a measure of how well a student performed. Person abilities were imported into IBM SPSS (26^{th} version), and further analyses were performed (paired *t*-tests and correlation analyses). The Rasch analysis provides two reliability measures: person and item reliabilities. Person reliability is used to classify the sample. A person reliability below 0.8 in an adequate sample means that the instrument is not sensitive enough to distinguish between high and low person ability. The person reliability depends on the sample ability variance, the length of the test, the number of categories per item and the sample-item targeting. A high person reliability is obtained with a wide ability range, many items, more categories per item and optimal targeting. If the item reliability is below 0.8 and the test length is adequate, this indicates that the sample size or composition is unsuitable for a stable ordering of items by difficulty (Linacre, 2023). Item reliability depends on item difficulty variance and person sample size. A wide difficulty range and a large sample result in higher item reliability. The item reliability is widely independent of test length.

Please note that Rasch values for person abilities for drawing molecules and cognitive load ratings are contrary to each other. Students who have drawn a large number of molecules correctly have high positive person abilities with regard to drawing molecules. Students who reported low cognitive load in the cognitive load ratings have high negative person abilities for the cognitive load ratings.

TABLE 2 German and English items of the adapted cognitive load questionnaire.

Subscale	Adapted version			Leppink et al. (2013)
	Item label	English translation	German version	
Intrinsic Load	[IL1p]	The topic covered in the tasks was very complex.	Der Fachinhalt, der den Aufgaben zugrunde liegt, war sehr komplex.	The topic/topics covered in the activity was/were very complex.
	[IL2p]	The tasks covered technical rules, that I perceived as very difficult.	Den Aufgaben liegen fachliche Regeln zugrunde, die ich als sehr schwierig empfunden habe.	The activity covered formulas that I perceived as very complex.
	[IL3p]	Working on the tasks requires a professional approach that I find challenging.	Die Bearbeitung der Aufgaben erfordert ein fachliches Vorgehen, das ich als herausfordernd empfinde.	The activity covered concepts and definitions that I perceived as very complex.
Extraneous Load	[EL1p]	The task was unclear to me.	Die Aufgabenstellung war mir unklar.	The instructions and/or explanations during the activity were very unclear.
	[EL2p]	The depiction of the tasks was not helpful for processing.	Die Darstellung der Aufgaben war für die Bearbeitung nicht hilfreich.	The instructions and/or explanations were, in terms of learning, very ineffective.
	[EL3p]	The tools (paper, pen) for working on the tasks were challenging.	Die Hilfsmittel (Papier, Stift) für die Aufgabenbearbeitung haben mich herausgefordert.	The instructions and/or explanations were full of unclear language.

These items refer to the paper-pencil version of the questionnaire. For the digital version of the questionnaire the examples in parenthesis [EL3] were replaced by: iPad, pen, drawing tool.

For all analyses, we report bias-corrected and accelerated 95% bootstrap confidence intervals (BCa 95% CI) from 1,000 bootstrapped samples in square brackets (Field, 2018). The effect size *d* for repeated measures of one group ($d_{RM,\ pooled}$) was calculated using the website *statistica* (Lenhard and Lenhard, 2023).

4 Results

4.1 Study 1

The first evaluation study was conducted at the end of the summer of 2022. University students (22 chemistry B.Sc. and water science B.Sc. majors taking the obligatory basic organic chemistry course, which only used paper-pencil-based exercises) were first asked to complete eight paper-pencil-based molecule-drawing tasks. Afterwards, they reported their perceived cognitive load using a multidimensional instrument adapted from Leppink et al. (2013). In the second part of the study, students were given an iPad and a brief introduction (10 min) to drawing molecules using JACK and Kekule.js, including a follow-along sample task. Once all the technical questions were answered, the students were asked to complete eight digital molecule-drawing tasks. Cognitive load and usability ratings were then collected.

The average mean for person ability regarding drawing molecules person ability ($N_{students} = 22$, $n_{pp\ items} = 8$ Person Reliability: 0.35, Item Reliability: 0.90, $n_{digital\ items} = 8$, Person Reliability: 0.63, Item Reliability: 0.85) in paper-pencil ($M = 0.69$, $SE = 0.34$) and digital ($M = -0.63$, $SE = 0.46$) tasks were compared using a *t*-test, which revealed a significant difference, $t(21) = 2.59$, $p = 0.017$, $d_{RM,\ pooled} = 0.438$, favoring the paper-pencil molecule-drawing tasks with higher person abilities, $\Delta M = 1.33$, BCa 95% CI [−2.268, −0.328]. Figure 2 provides two boxplot diagrams. Person abilities are plotted on the y-axis using the logit scale of the Rasch model. High person abilities are represented by high positive values. A mid-level person ability is represented by a value around zero. Low person abilities in drawing molecules are represented by negative values. The two formats (paper-pencil, digital) are plotted on the x-axis. The two boxplot diagrams show the distribution of person abilities for drawing molecules for both formats. The bold black lines represent the median for person abilities for each format (Field, 2018). The blues boxes represent the area into which the person abilities of half of the observed students fall (interquartile range). The whiskers represent the top and bottom 25% of person abilities. Differences regarding the position of the bold black line, the shape of the blue boxes and the positions of whiskers indicate differences between the formats. A lower position of the bold black line for the digital format in the diagram at the upper left (study 1, summer 2022) indicates that the students' person ability to solve digital tasks was lower than their ability to solve paper-pencil-based tasks.

Intrinsic cognitive load ratings were used to inspect the cognitive load which arises from the content, namely drawing molecules based on their IUPAC name. The more interacting elements a student has to process at the same time, such as the number of carbon atoms related to the name of the carbon chain, the functional group related to the name of a substituent, the number of alkyl groups and their positions, the higher the intrinsic cognitive load will be. Comparing the average mean for person ability regarding intrinsic load ratings ($N_{students} = 22$, $n_{il\ items\ pp} = 3$, Person Reliability: 0.71, Item Reliability: 0.87, $n_{il\ items\ digital} = 3$ Person Reliability: 0.70, Item Reliability: 0.96) for paper-pencil ($M = -0.78$, $SE = 0.32$) and digital ($M = -0.43$, $SE = 0.25$) tasks using a *t*-test revealed no significant differences, $t(21) = 1.06$, $p = 0.300$, $\Delta M = 0.36$, BCa 95% CI [−1.032, −0.238], which means students perceived equal intrinsic load while working on the paper-pencil and

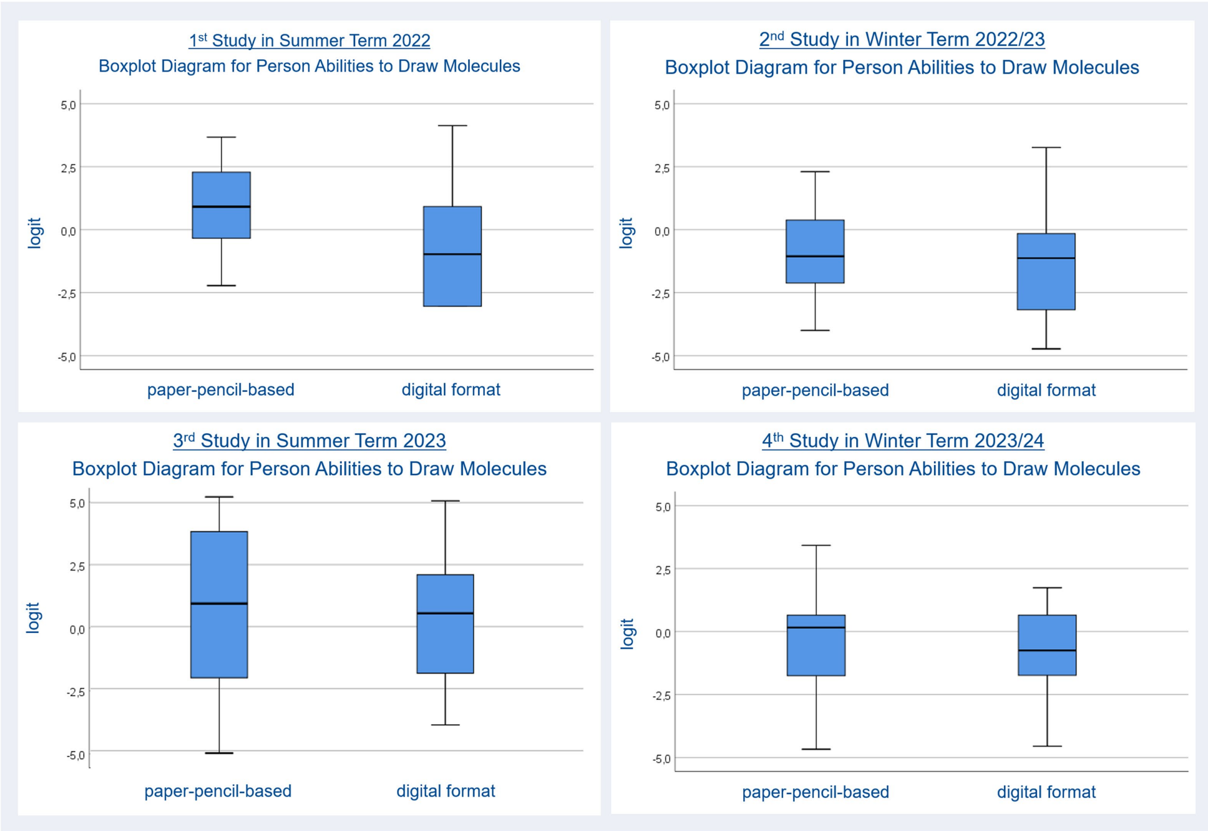


FIGURE 2
Person abilities in drawing molecules for paper-pencil-based and digital molecule-drawing tasks.

digital molecule-drawing tasks. Figure 3 (upper left) shows the comparison between formats regarding intrinsic cognitive load using boxplot diagrams. The comparable position of the median (bold black line) and the shape of the boxes and whiskers indicate no differences between formats. The single dot labeled with an 8 underneath one whisker represents an outlier (a student whose ratings are not fitting to the other students' ratings).

Extraneous cognitive load ratings were used to inspect the cognitive load which arises from processing not content related information. For example, a student maybe knows which functional group she has to draw to depict a molecule but she cannot remember where to find the option to add a heteroatom within the user interface of Kekule.js. Therefore, she has to try different options. In case she chooses a wrong user interface element on her first try, she has to remove the changes she made to her molecule, before she can come back to searching for an option to add the heteroatom. This testing by trial-and-error will require cognitive resources which are not available to solving the original task. The average mean for person ability regarding extraneous load ratings ($N_{students} = 22$, $n_{el\ items\ pp} = 3$, Person Reliability: 0.47, Item Reliability: 0.56, $n_{el\ items\ digital} = 3$, Person Reliability: 0.54, Item Reliability: 0.97) for paper-pencil ($M = -1.77$, $SE = 0.42$) and digital ($M = -0.38$, $SE = 0.17$) tasks were also compared using a t-test. The results revealed significantly lower mean person abilities for paper-pencil tasks, $t(21) = 3.21$, $p = 0.004$, $d_{RM,\ pooled} = 0.430$, $\Delta M = 1.49$, BCa 95% CI [−2.268, −0.648], meaning that students perceived lower extraneous load while working on such tasks. Again, boxplot diagrams were used to show differences between formats regarding extraneous cognitive load (Figure 4, upper left). Like in Figure 2 the differences regarding the position of the median and the distribution of person abilities regarding agreement with extraneous cognitive load ratings indicate differences between the two formats.

Person abilities for usability ratings for the digital tasks ($N_{students} = 22$, $N_{Items} = 9$, Person Reliability: 0.88, Item Reliability: 0.89) amounted to $M = -0.37$, $SE = 0.25$, slightly below the middle of the scale. Hence, the students' usability experience was slightly less than medium.

Person ability for cognitive load ratings and the usability of digital tasks were correlated. The intrinsic cognitive load was significantly correlated with usability, $r = -0.477$, $p = 0.025$, BCa 95% CI [−0.758, −0.082]. Hence, students who reported higher intrinsic load experienced lower usability. The correlation between extraneous load and usability was in line with this finding, $r = -0.455$, $p = 0.033$, BCa 95% CI [−0.762, −0.041]. Hence, students who reported higher extraneous load experienced lower usability.

Overall, the first study showed that students' ability to solve paper-pencil molecule-drawing tasks was higher than their ability to solve comparable digital tasks, and that students reported higher extraneous cognitive load for digital molecule-drawing tasks and unsatisfactory usability. Hence, students who received very little instruction regarding the tool and previously learned to draw molecules in a paper-pencil format were hindered in their ability to

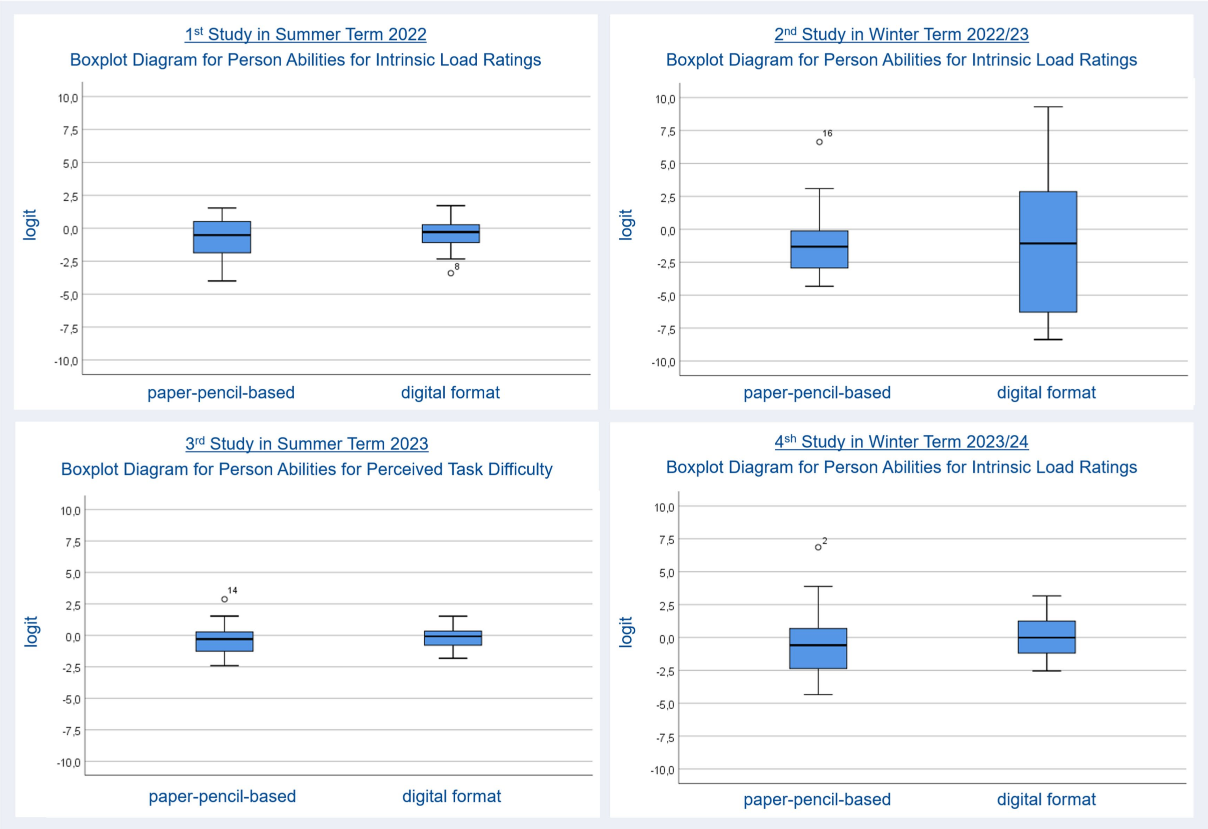


FIGURE 3
Person abilities for cognitive load ratings for paper-pencil-based and digital molecule-drawing tasks.

solve molecule-drawing tasks and perceived an extraneous load when trying to solve digital molecule-drawing tasks.

4.2 Study 2

To investigate whether the results of the first study would change if students were used to drawing molecules digitally, we extended the introduction to the digital tool. The first session of the organic chemistry course in the winter term was used to provide a longer introduction to the tool, with a training phase of eight tasks and personal support for problems and questions (approximately 30 min in total). Additionally, we asked students to bring their own devices and offered faculty iPads to students who did not bring their own devices. During the following weeks, the students were continuously asked to complete digital molecule-drawing tasks at home as part of the course exercises. After 1 month, data were collected from 21 bachelor's degree students preparing to become chemistry teachers. In Germany, these students study two subjects (in our case chemistry and another subject, e.g., mathematics) and educational science. Students first worked on paper-pencil molecule-drawing tasks and rated their perceived cognitive load before working on the digital tasks and reporting their perceived cognitive load and usability. In this study, we used 17 molecule-drawing tasks in tandem. Again, we collected cognitive load and usability ratings.

The average mean for person ability regarding performance ($N_{students} = 21$, $n_{pp\ items} = 17$, Person Reliability: 0.74, Item Reliability: 0.83, $n_{digital\ items} = 17$, Person Reliability: 0.78, Item Reliability: 0.86) in paper-pencil ($M = -1.03$, $SE = 0.39$) and digital ($M = -1.39$, $SE = 0.48$) tasks were compared using a t-test; this revealed no significant difference, $t(20) = 1.10$, $p = 0.285$. Boxplot diagrams (Figure 2, upper right) support this finding.

Item Reliability for intrinsic cognitive load ratings for the digital format is not acceptable. Low item reliability implies that the person sample is not large enough to confirm the item difficulty hierarchy (=construct validity) of the instrument; hence, our work excluded item analysis. A comparison of the average mean for person ability regarding intrinsic load ratings ($N_{students} = 21$, $n_{il\ items\ pp} = 3$, Person Reliability: 0.87, Item Reliability: 0.66, $n_{il\ items\ digital} = 3$ Person Reliability: 0.94, Item Reliability: 0.00) for paper-pencil ($M = -1.04$, $SE = 0.55$) and digital ($M = -1.05$, $SE = 1.25$) tasks using a t-test revealed no significant differences, $t(20) = 0.01$, $p = 0.990$, $\Delta M = 0.01$, BCa 95% CI [−1.851, −1.863], which means that students perceived equal content-related (intrinsic) load while working on the paper-pencil and digital molecule-drawing tasks. Boxplot diagrams (Figure 3, upper right) support this finding. Again, the single dot labeled with 16 above the left whisker represents an outlier.

The average mean for person ability regarding not content-related (extraneous) load ratings ($N_{students} = 21$, $n_{el\ items\ pp} = 3$, Person Reliability: 0.32, Item Reliability: 0.71, $n_{el\ items\ digital} = 3$, Person Reliability: 0.53, Item Reliability: 0.88) for paper-pencil ($M = -2.68$, $SE = 0.37$) and digital

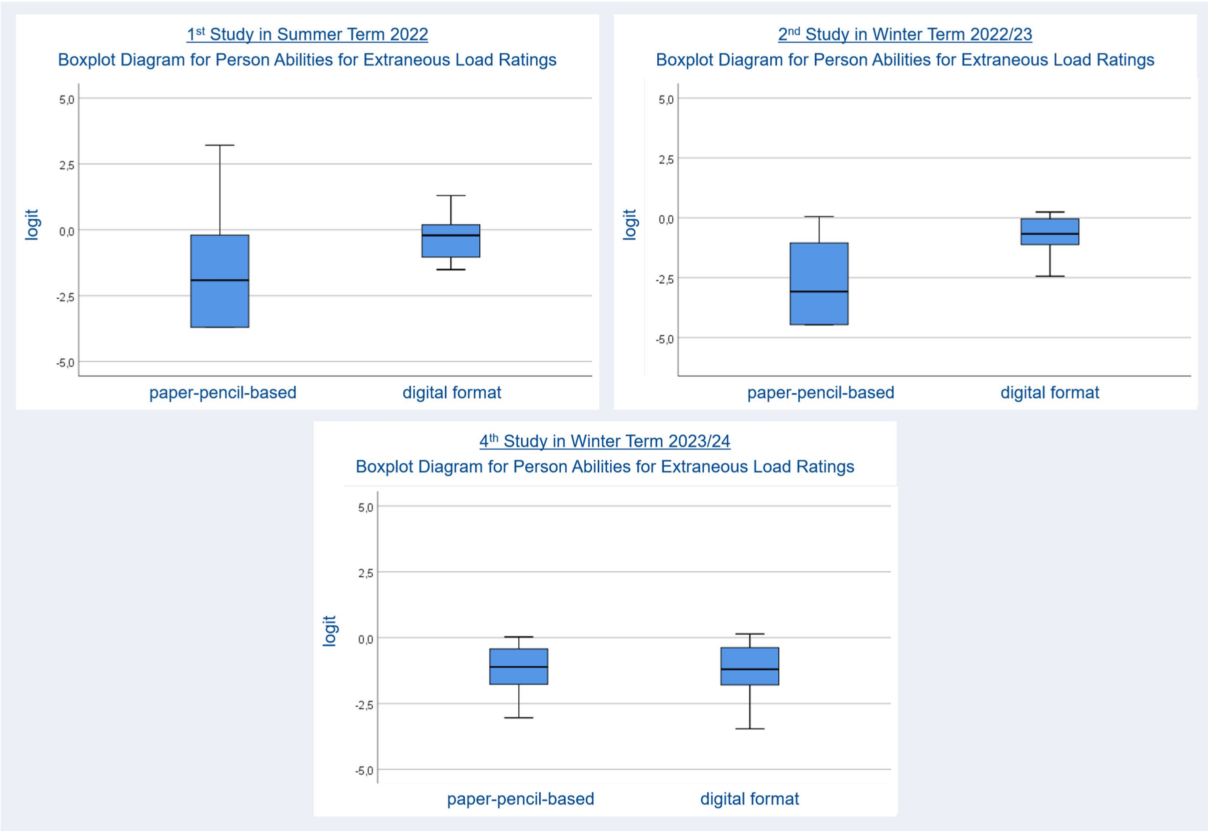


FIGURE 4
Person abilities for extraneous cognitive load ratings after working on paper-pencil-based and digital molecule-drawing tasks.

($M = -0.77$, $SE = 0.19$) tasks were also compared using a t-test. This revealed significantly lower mean ratings for person ability regarding extraneous load in paper-pencil tasks, $t(20) = 4.90$, $p \le 0.001$, $d_{RM, pooled} = 1.421$, $\Delta M = -1.91$, BCa 95% CI [−2.686, −1.106], meaning that students perceived lower extraneous load while working on the paper-pencil molecule-drawing tasks. Boxplot diagrams (Figure 4, upper right) support this finding.

Person abilities for usability ratings for the digital tasks ($N_{students} = 21$, $N_{items} = 9$, Person Reliability: 0.85, Item Reliability: 0.88) amounted to $M = 0.19$, $SE = 0.27$, slightly above the middle of the scale. Hence, the students' usability experience was slightly above medium level.

Person ability for cognitive load ratings and usability of the digital tasks were correlated. The correlation between extraneous load and usability is significant, $r = -0.738$, $p \le 0.001$, BCa 95% CI [−0.883, −0.479]. Hence, students who reported higher extraneous load experienced lower usability. The correlation between person ability for intrinsic cognitive load and usability was not significant. Hence, no significant relationship was found between intrinsic cognitive load and usability.

In conclusion, the results showed that the extended introduction, continuous use of digital molecule-drawing exercises during the course, and opportunity to work with a personal device helped reduce differences between the ability to draw molecules either paper-pencil-based or digitally. Equally perceived intrinsic cognitive load pointed to the same direction. However, students still perceived a higher extraneous load when working on the digital molecule-drawing tasks. Students' usability experience was a bit above medium level.

4.3 Study 3

Due to the further development of the tool, we have made various changes to the design of the organic chemistry exercise for the cohort that took part in the third study. However, these changes did not alter the tasks we used for the study.

In the summer of 2023, students were again introduced to the tool during the first session of the term; this included a training phase on their own devices and an opportunity to ask questions and receive personal support. Students were continuously asked to work on digital exercises as homework for the course during the following weeks. Three weeks later, data were collected from 26 undergraduate students (majoring in chemistry or water science). We used 14 molecule-drawing tasks in tandem this time. Students first worked on the digital drawing tasks before switching to the paper-pencil format. Considering the unsatisfactory reliability of the cognitive load rating scales, we used an immediate unidimensional single-item measure (Kalyuga et al., 2001). This item measures an overall cognitive load. Thus, it is not able to differentiate between extraneous or intrinsic load. In the third study, no usability ratings were collected (usability was investigated at a later point in the course; therefore, no usability ratings were reported for the third study).

The average mean for person ability regarding drawing molecules ($N_{students} = 26$, $n_{pp\ items} = 14$, Person Reliability: 0.86, Item Reliability: 0.85, $n_{digital\ items} = 14$, Person Reliability: 0.81, Item Reliability: 0.88) in paper-pencil ($M = 0.73$, $SE = 0.62$) and digital ($M = 0.23$, $SE = 0.50$) tasks were compared using a *t*-test; this revealed no significant difference: $t(25) = 1.21$, $p = 0.238$. The boxplot diagrams in Figure 2 (upper left) represent this finding.

The average mean for person ability regarding perceived task difficulty ratings ($N_{students} = 26$, $n_{pp\ item} = 1$ answered 14 times, Person Reliability: 0.94, Item Reliability: 0.95, $n_{item\ digital} = 1$ answered 14 times, Person Reliability: 0.72, Item Reliability: 0.94) of paper-pencil ($M = -0.28$, $SE = 0.23$) and digital ($M = -0.18$, $SE = 0.15$) tasks were also compared using a *t*-test. This revealed no significant difference, $t(25) = 1.05$, $p = 0.306$, $\Delta M = 0.10$, BCa 95% CI [−0.292, 0.101], which means that perceived task difficulty did not differ between digital and paper-pencil molecule-drawing tasks. The boxplot diagrams in Figure 3 (upper left) represent this finding. As these boxplots are based on perceived task difficulty ratings they are not directly comparable to the other boxplot diagrams in Figure 2 which are based on intrinsic cognitive load ratings. The dot labeled with 14 again represents an outlier.

In summary, the results of the third study support the findings of the second study and provide further evidence that students' performance in simple molecule-drawing tasks is independent of the format (paper-pencil or digital). Regarding extraneous cognitive load, it remains unclear whether the item used failed to detect differences or whether students did not perceive a higher level of extraneous cognitive load while working on digital molecule-drawing tasks. The findings require further investigation.

4.4 Study 4

Before the start of the winter term (2023/24), we attempted to improve the usability of the digital tool for an elementary organic chemistry course by deleting user interface elements that were irrelevant for working on the tasks (e.g., templates for heterocycles). At the beginning of the course, students were introduced to the tool during the first session of the term; this included a training phase on their own devices and the opportunity to ask questions and receive personal support. Afterwards, they had the opportunity to work on the first organic chemistry tasks in class, ask questions, or receive support. Students were continuously asked to work on digital exercises as homework for the course during the following weeks. Two weeks later, data were collected from 25 bachelor's degree students preparing to become chemistry teachers. Again, we used 14 molecule-drawing tasks in tandem, and the students first worked on the digital drawing tasks before switching to the paper-pencil format. We switched back to the multidimensional cognitive load measure we already used for the first and the second study and used the UEQ (Laugwitz et al., 2008) to measure usability.

The average mean for person ability regarding drawing molecules ($N_{students} = 25$, $n_{pp\ items} = 14$, Person Reliability: 0.78, Item Reliability: 0.86, $n_{digital\ items} = 14$, Person Reliability: 0.73, Item Reliability: 0.85) in paper-pencil ($M = 0.37$, $SE = 0.42$) and digital ($M = -0.86$, $SE = 0.39$) tasks were compared using a *t*-test, revealing no significant difference, $t(24) = 1.67$, $p = 0.107$. Again, boxplot diagrams were used to represent this finding (Figure 2, bottom right).

A comparison of the average mean for person ability regarding content-related (intrinsic) load ratings ($N_{students} = 25$, $n_{il\ items\ pp} = 3$, Person Reliability: 0.87, Item Reliability: 0.71, $n_{il\ items\ digital} = 3$ Person Reliability: 0.75, Item Reliability: 0.71) for paper-pencil ($M = -0.64$, $SE = 0.51$) and digital ($M = 0.07$, $SE = 0.32$) tasks using a *t*-test revealed no significant differences, $t(24) = 1.80$, $p = 0.085$, meaning that students perceived equal intrinsic load while working on the paper-pencil and digital molecule-drawing tasks (Figure 3, bottom right). Again, boxplot diagrams were used to represent this finding (Figure 3, bottom right). An outlier is shown by the dot labeled with 2.

We also compared the average mean for person ability regarding not content-related (extraneous) load ratings ($N_{students} = 25$, $n_{el\ items\ pp} = 3$, Person Reliability: 0.05, Item Reliability: 0.47, $n_{el\ items\ digital} = 3$, Person Reliability: 0.35, Item Reliability: 0.90) for paper-pencil ($M = -1.25$, $SE = 0.21$) and digital ($M = -1.50$, $SE = 0.25$) tasks using a *t*-test. This revealed no significant differences, $t(24) = 0.76$, $p = 0.458$. Hence, the students perceived equal extraneous load while working on the paper-pencil and digital molecule-drawing tasks. Again, boxplot diagrams were used to represent this finding (Figure 4, bottom right).

The mean person abilities for all subscales of the UEQ were above zero. Hence, students perceived usability as acceptable. The attractiveness of the tool was merely sufficient, but students' perceptions of efficiency and stimulation were very good as shown in Table 3.

Extraneous cognitive load was significantly negatively correlated with attractiveness: $r = -0.475$, $p = 0.016$, BCa 95% CI [−0.739, −0.114]; perspicuity: $r = -0.399$, $p = 0.048$, BCa 95% CI [−0.696, 0.169]; efficiency: $r = -0.552$, $p = 0.004$, BCa 95% CI [−0.796, −0.151]; and stimulation: $r = -0.585$, $p = 0.002$, BCa 95% CI [−0.802, −0.266], indicating that students who perceived high extraneous load rated the tool's attractiveness, perspicuity, efficiency, and stimulation to be lower and vice versa. We found no significant correlation between intrinsic load and tool usability.

In summary, the results of the fourth study support the findings of the second and third studies and provide further evidence that students' performance in simple molecule-drawing tasks is independent of the format (paper-pencil or digital). Regarding extraneous cognitive load, the fourth study provided evidence that after reducing the user interface, students perceived equal levels of extraneous cognitive load in both the paper-pencil and digital molecule-drawing tasks. Students considered the tool's usability sufficient, with values above zero for all usability subscales.

TABLE 3 Parameters for all subscales of the user experience questionnaire.

	N_{Items}	Person reliability	Item reliability	*M*	*SE*
Attractiveness	6	0.56	0.86	0.13	0.34
Perspicuity	4	0.68	0.88	0.82	0.55
Efficiency	4	0.71	0.84	1.15	0.60
Novelty	4	0.67	0.33	0.60	0.52
Dependability	4	0.67	0.78	0.67	0.56
Stimulation	4	0.87	0.90	2.18	0.84

$N_{Persons} = 25$.

5 Discussion

5.1 Discussion of the evaluation studies and their limitations

Figure 5 summarizes the results of the evaluation studies. The comparison between paper-pencil-based format and digital format regarding drawing molecules and cognitive load is represented individually. Undesirable differences between formats are marked with an X, missing statistical differences between formats are marked with a check mark. Additionally, results regarding usability are included. Figure 5 shows that we were able to improve the tool to reach satisfying results regarding drawing molecules, cognitive load, and usability over time.

The first study showed significant differences in person ability to solve paper-pencil-based and digital molecule-drawing tasks. This finding is consistent with the significantly higher extraneous cognitive load and equal intrinsic load observed for both formats. Please remember that intrinsic load arises from the content-related complexity of a task (e.g., drawing an ethane molecule is less complex than drawing a 3-ethylhexanal molecule because more structural elements have to be considered for drawing) whereas extraneous cognitive load arises from processing information which are not content-related (e.g., searching for an option to add a heteroatom required for depicting a functional group by trial-and-error). Thus, although the participants perceived the chemical content of the molecule-drawing task to be equally difficult (no significant difference regarding intrinsic cognitive load), their ability to solve such tasks was higher for the paper-pencil format. We used the Rasch model to calculate students' person ability to solve tasks asking them to draw molecules based on their IUPAC name. The person ability is based on students' performance. In comparison to a raw sum score of correctly solved items, the person ability also considers whether the correctly solved items were easy or difficult items. The lower ability to solve digital molecule-drawing tasks can be explained by the higher extraneous cognitive load of this format. Accordingly, the usability ratings were unsatisfactory, as they were below the scale center. That usability was significantly negatively correlated with intrinsic cognitive load meant that a higher intrinsic cognitive load was associated with reduced experienced usability and vice versa. Additionally, the extraneous cognitive load was significantly negatively correlated with usability. Specifically, students who perceived high extraneous load reported lower usability. These findings align with the cognitive load theory, suggesting that students who already perceive a high load from processing the relevant information to solve a task are at risk of perceiving cognitive overload due to unintuitive and unfamiliar handling of the tool, resulting in poorer usability. By contrast, participants who perceive a lower intrinsic load may have more working memory capacity to process the necessary information regarding tool handling, leading to better usability. Overall, the first study showed that students who had only minimal instruction in using the tool, previously learned to draw molecules in a traditional paper-pencil format, and used foreign devices experienced difficulty in solving molecule-drawing tasks and faced additional extraneous cognitive load when attempting to solve digital molecule-drawing tasks. The high extraneous load implies that the difficulty did not lie in the task itself but was induced by the unfamiliar medium or format.

Based on this assumption, we extended the participants' introduction to the digital tool in the next term and transitioned from using faculty devices to using students' personal devices, which were more familiar to them. Additionally, during the course of the term, we used the digital tool for knowledge acquisition. As there were no significant differences in person ability to complete the digital or paper-pencil-based molecule-drawing tasks, students were able to perform at similar levels when given their own devices, extended instructions, and regular training with the digital tool. Nevertheless, there was still a significantly higher extraneous cognitive load reported by the students when performing the digital drawing tasks. Again, no significant differences were found in intrinsic cognitive load. The usability rating slightly improved from the first to the second study and was found to be moderate. A significant correlation between intrinsic load and usability was no longer observed. Hence, when students were accustomed to using the tool, the perceived usability appeared unrelated to the perceived difficulty of the task. Similar to the first study, we found a negative correlation between the extraneous cognitive load and tool usability. This suggests that the tool's unintuitive handling places a load on the students' working memory. In summary, the results showed that the extended introduction, regular use of digital molecule-drawing exercises during the course,

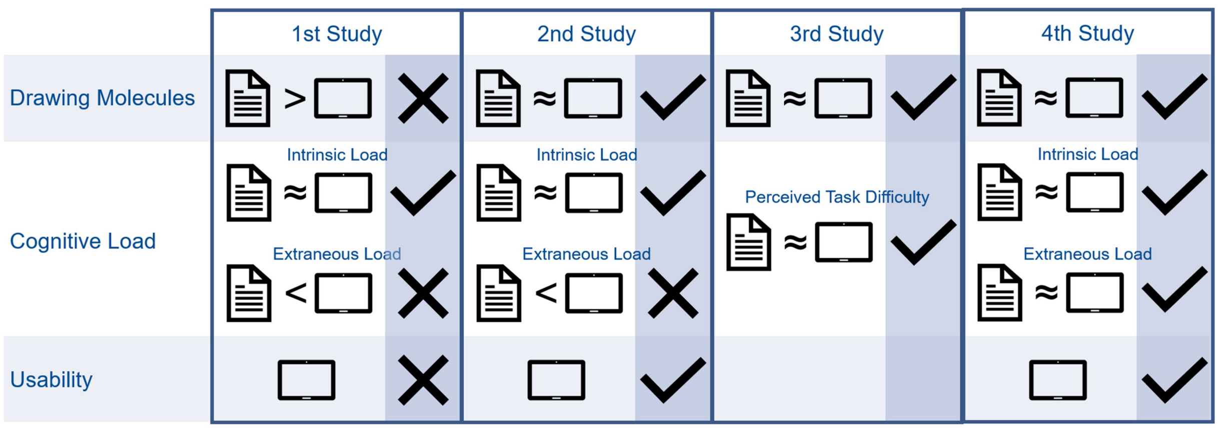


FIGURE 5
Overview of the evaluation studies' results.

and the opportunity to work with personal devices helped decrease the gap in students' ability to draw molecules using either paper and pencil or a digital tool. In conclusion, these findings indicate that the implemented measures effectively reduced the differences in performance. Nevertheless, the students still perceived a higher extraneous load when working on digital molecule-drawing tasks.

Students' feedback revealed that they found a combination of digital and paper-pencil homework exercises challenging to organize. However, the results from the second study showed no difference in person ability between the two formats when given an extended introduction, opportunities for regular use, and personal devices for students' use. The tool's ongoing improvements encouraged us to switch to entirely digital homework exercises for the basic organic chemistry course. The results of the third study confirmed the findings of the second study, which showed no differences in person abilities to solve molecule-drawing tasks using digital or paper-pencil-based formats. Unidimensional cognitive load measurements (which measure an overall load without being able to distinguish between intrinsic and extraneous load) did not detect differences in cognitive load between formats. However, it remained unclear whether the instrument was unable to detect differences or whether students, being used to drawing molecules digitally, did not perceive differences in the cognitive load. No usability ratings were collected in the third study.

To improve usability and reduce the extraneous load, we have customized the user interface by removing elements that are not necessary for processing the current tasks (e.g., templates for drawing heterocycles). Moreover, instead of working at home, we gave students the chance to work on the first organic chemistry tasks in class and provided support when needed. The fourth study showed equal abilities to solve molecule-drawing tasks, intrinsic load, and extraneous load for both formats, and ratings of sufficient to good for tool usability.

Overall, the evaluation studies demonstrated that the developed tool is functional. We provided evidence that students can accomplish digital drawing tasks using the tool as proficiently as using paper and pencil, provided that they are sufficiently familiar with the tool (RQ1). Nevertheless, the tool appeared to induce an extraneous cognitive load (RQ2) that should be reduced to relieve students' working memory. Reducing the extraneous cognitive load shows promise in improving unsatisfactory tool usability (RQ3), as these factors are closely related. An appropriate configuration of the Kekule.js widgets to accommodate task requirements involves hiding unnecessary controls and options, along with reducing extraneous cognitive load and improving usability (fourth study).

One limitation is that the chosen study design did not allow us to statistically determine the factors that cause a higher extraneous load for the digital drawing tool. Any adjustments we made were therefore based solely on assumptions derived from experience and discussions with students. One of these assumptions is that students who would normally not choose skeletal formula because of its abstract nature were forced by the digital tool to do so. Moreover, the use of a digital tool could lead to a tool-driven solution (e.g., avoiding repeated switching between required functions such as the option to draw single bonds or the option to insert heteroatoms). Unlike the paper format where students can develop their drawing step by step along the name of the molecule, a tool-driven solution requires a mental pre-structuring and planning of the drawing. We also deem the absence of a notepad function a deficiency, because this could help relieve working memory through note-taking. Although students were also allowed to take pen and paper to solve the digital tasks in the current format, this was only used spontaneously by a few individuals (so it could also be a strategy problem for the students). We can also imagine that the automatic correction of violations of the octet rule causes confusion among students. Occasionally, questions were asked in the tests that point in this direction (If I insert an oxygen atom here, suddenly a hydrogen atom also appears. How do I get rid of that?). Various designs for evaluation studies are required to understand the causes of this extraneous load. A comparison of the representation form used between paper-pencil and digital formats appears to be a productive method for obtaining information on students' preferred representation format. In situations in which students do not use the skeletal formula for the paper-pencil format, differences between media may occur because of the obligation to use an unfamiliar representation form for the digital format. Additional training for the use of skeletal formulas, the commonly used representation format for presenting chemical structures among professionals, offers the potential for improvement. Further investigation is necessary to better understand the causes of the extraneous cognitive load. Reliable instruments are required to measure both extraneous and intrinsic cognitive load. Aside from those used in the evaluation studies presented, other multidimensional cognitive load measurement instruments are available for learning scenarios with unknown potential (Klepsch et al., 2017; Krieglstein et al., 2023); however, comparable instruments for performance are still missing.

Our findings are limited by the unsatisfactory reliability of the multidimensional measurements of cognitive load used here as well as the potential inability of the unidimensional instrument to identify variances in extraneous cognitive load. Furthermore, the results are limited in that they rely on a small number of molecule-drawing tasks and the inclusion of students from only one university. In addition, our results are limited by the fact that the students study different subjects in the summer and winter semesters (Chemistry B. Sc. and Water Science B. Sc. in the summer semester, students preparing to become chemistry teachers in the winter semester). We recommend that future studies include students from other universities, which would require synchronization of basic organic chemistry courses and the use of the same exercises. Another limitation is the use of a single task format, specifically molecule-drawing tasks. Future research should investigate whether these results can be replicated for other tasks such as those dealing with chirality or reaction mechanisms. The results of our work are limited to assessments and do not consider learning, bearing in mind that the main objective of developing the digital tool was to enhance learning.

In summary, based on initial evaluations, it was feasible to implement a digital tool for organic chemistry courses. We developed a noncommercial (G1) e-learning and e-assessment tool that (G2) can be integrated into learning platforms such as Moodle or Ilias. Further efforts are needed to reduce the system-induced extraneous cognitive load and thereby enhance usability (G5). The next step is to assess whether integrating the tool's features—such as receiving individualized, explanatory feedback in real time, working at a self-paced rate, and having extra opportunities to practice—enhance students' learning outcomes and overall performance in organic chemistry (G4). A further step could be to look for different teachers who create their own typical organic chemistry tasks that they need for their courses (G3).

10.3389/feduc.2024.1355078

In addition to evaluating the use of a digital tool for an organic chemistry course, various other findings have emerged from discussions with students. For example, students expressed a dislike for the combination of digital and paper-pencil homework exercises, as they made the tasks challenging to manage. Students told us that working on digital and paper-pencil-based tasks makes it hard to remember whether they have solved all tasks as they had to check two sources. Additionally, they had problems with integrating paper-pencil-based and digital notes when preparing for exams. The participants also expressed their desire to export tasks, their responses, and the feedback they received to document their course progress and prepare for exams. Because offline copies of exams are required for exam documentation, the export-as-PDF feature was added to JACK, allowing students to save their results. Experiences from the summer of 2023 demonstrated that the implementation of a digital learning environment failed to improve student motivation to complete course exercises. Only a minority of the students completed the weekly exercises at home, while the majority appeared unprepared for the in-person sessions, expecting the teacher to provide them with a sample solution for review before the end-of-semester examination. As we decided to provide a worked-out example after three unsuccessful solution attempts, the format of the course's in-person sessions will change when the weekly homework exercises are done digitally. It is no longer necessary to use in-person sessions to provide sample solutions for all tasks. Hence, the use of digital homework exercises has the potential to enhance in-person sessions towards a more advanced involvement with organic chemistry, provided that students are willing and adequately prepared to attend classes. In short, the digital tool does facilitate the implementation of flipped classroom approaches (van Alten et al., 2019), with student motivation being the main challenge.

5.2 Technical development and next steps

Overall, the integration of Kekule.js into JACK has resulted in a powerful tool that enables the digital implementation of a sufficient number of standard organic chemistry tasks and has potential for further enhancement, such as automating the evaluation of complete reaction mechanisms. Although further challenges must be addressed (e.g., automatic evaluation of mesomerism, transition states, arrows for electron transition, or the option to create and evaluate tasks that ask to mark parts of a molecule), the existing tool already offers additional learning opportunities and explanatory feedback that cannot be offered in the same amount during an in-person course. System development activities thus far have focused on the proper technical integration of Kekule.js into JACK, including the possibility

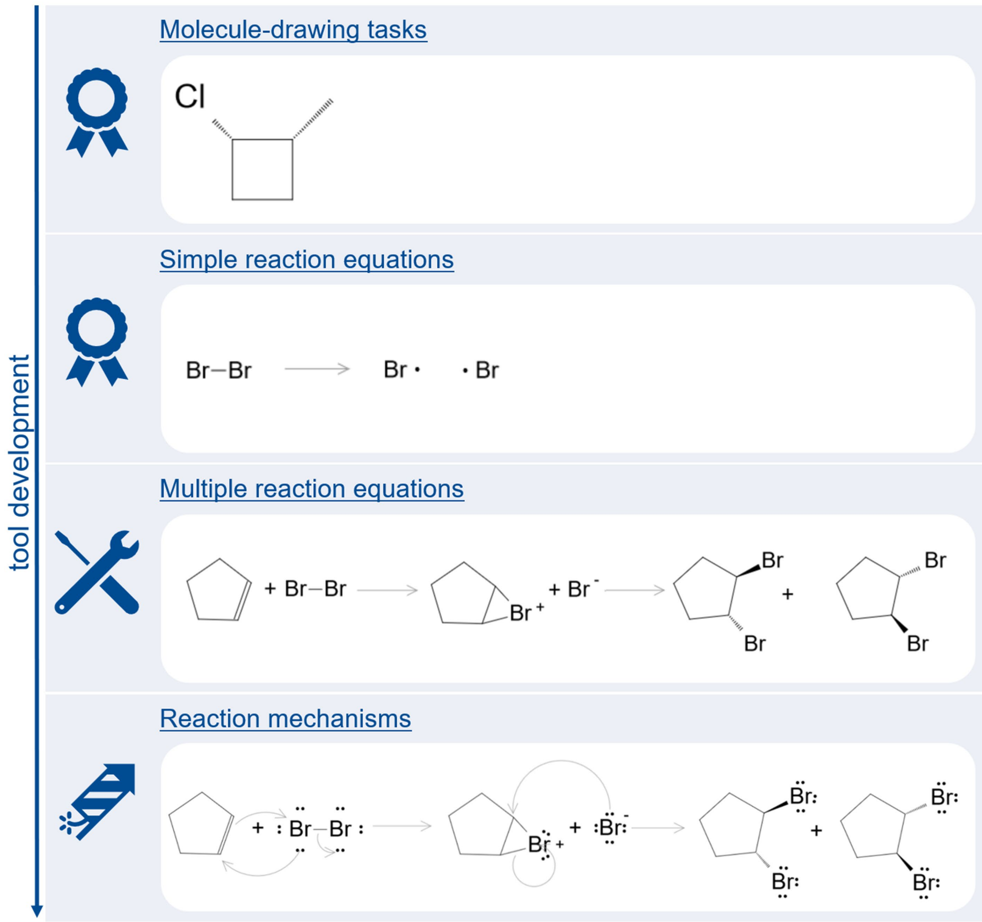


FIGURE 6
Tool development.

of configuring the editor interface and providing essential features for teachers to define rule-based feedback. Thus, the current state of development has not yet explored the potential additional benefits that digital drawing may provide with respect to automated analysis. The next development steps will extend the capabilities for automated answer analysis and rule-based feedback in two ways. First, JACK will not only analyze the InChI code generated from drawings, but also inspect the machine-readable representation of the drawing itself. This will not only allow the analysis of reactions with respect to the positions of molecules in relation to the reaction arrow, as already mentioned above, but also solve some other problems such as the representation; for example, it will also allow detecting additional annotations and color markings. Second, JACK will be linked to a Chemistry Development Kit (Steinbeck et al., 2006), allowing for the automatic detection of certain differences between molecules. This will save teachers the burden of listing all well-known errors in their feedback rules and instead allow them to write these rules on a conceptual level, defining feedback on types of errors instead of individual errors.

Automatic analysis of molecule-drawing tasks is based of comparing InChI-codes. During the 2 years in which evaluation studies were set, an additional function to automatically analyze reaction equations was implemented. This function is based on searching for reaction arrow and checking which molecules are placed in front (educts) and behind (products) or above (minor educts) or underneath (minor products) this reaction arrow. An extension of this function, which is currently under technical evaluation, allows to automatically evaluate reactions based on multiple reaction equations. The next step will be to inspect the machine-readable representation of the drawing itself, which will allow to inspect for lone electron pairs and electron pushing arrows. Hence, this function will enable us to automatically analyze reaction mechanisms. Figure 6 provides an overview over the tool's development with example tasks.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies involving humans because although IRB approval was not necessary at German universities, the guidelines concerning the ethical scientific practice of the Federal German Research Foundation (DFG) were applied. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KS: Conceptualization, Investigation, Methodology, Visualization, Writing – original draft. MS: Software, Writing – original draft. DP: Conceptualization, Writing – review & editing. AL: Funding acquisition, Project administration, Supervision, Writing – review & editing. MGo: Project administration, Software, Supervision, Writing – review & editing, Resources. MGi: Investigation, Project administration, Supervision, Writing – review & editing, Validation. MW: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This project was funded by Foundation Innovations in Higher Education. We thank the University of Bonn for funding the development of the e-learning tool within the Hochschulpakt and ZSL program of North Rhine-Westfalia State.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Anzovino, M. E., and Lowery Bretz, S. (2016). Organic chemistry students' fragmented ideas about the structure and function of nucleophiles and electrophiles: a concept map analysis. *Chem. Educ. Res. Pract.* 17, 1019–1029. doi: 10.1039/C6RP00111D

Asmussen, G., Rodemer, M., and Bernholt, S. (2023). Blooming student difficulties in dealing with organic reaction mechanisms – an attempt at systemization. *Chem. Educ. Res. Pract.* 24, 1035–1054. doi: 10.1039/d2rp00204c

Aulck, L., Velagapudi, N., Blumenstock, J., and West, J. (2016). Predicting student dropout in higher education. *2016 ICML Workshop on #Data4Good: Machine Learning in Social Good Applications*. doi: 10.48550/arXiv.1606.06364

Averbeck, D. (2021). "Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums: Der Einfluss kognitiver und affektiv-motivationaler Variablen [study success of chemistry freshmen: the influence of cognitive and affective variables]" in Studien zum Physik- und Chemielernen: Vol. 308. ed. A. Daniel (Berlin: Logos Berlin).

Bodé, N. E., Caron, J., and Flynn, A. B. (2016). Evaluating students' learning gains and experiences from using nomenclature101.com. *Chem. Educ. Res. Pract.* 17, 1156–1173. doi: 10.1039/c6rp00132g

Bond, T. G., and Fox, C. M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences. *2nd* Edn. New York: Lawrence Erlbaum Associates Publishers.

Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how? *CBE Life Sci. Educ.* 15:148. doi: 10.1187/cbe.16-04-0148

Boone, W. J., and Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science* 90, 253–269. doi: 10.1002/sce.20106

Boone, W. J., and Staver, J. R. (2020). Advances in Rasch analyses in the human sciences. Cham: Springer International Publishing.

Boone, W. J., Staver, J. R., and Yale, M. S. (2014). Rasch analysis in the human sciences. Heidelberg: Springer Netherlands.

Brooke, J. (1996). "SUS: a 'quick and dirty' usability scale" in Usability evaluation in industry. eds. P. W. Jordan, B. Thomas, B. A. Weerdmeester and I. L. McClelland. *First* ed (London, GB: Taylor and Francis: Bristol, PA).

Buchner, J., and Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Comput. Educ.* 195:104711. doi: 10.1016/j.compedu.2022.104711

Chamala, R. R., Ciochina, R., Grossman, R. B., Finkel, R. A., Kannan, S., and Ramachandran, P. (2006). EPOCH: an organic chemistry homework program that offers response-specific feedback to students. *J. Chem. Educ.* 83, 164–169. doi: 10.1021/ed083p164

Chen, J. H., and Baldi, P. (2008). Synthesis explorer: a chemical reaction tutorial system for organic synthesis design and mechanism prediction. *J. Chem. Educ.* 85, 1699–1703. doi: 10.1021/ed085p1699

Chen, J. H., and Baldi, P. (2009). No electron left behind: a rule-based expert system to predict chemical reactions and reaction mechanisms. *J. Chem. Inf. Model.* 49, 2034–2043. doi: 10.1021/ci900157k

Chen, J. H., Kayala, M. A., and Baldi, P. (Eds.). (2010). Reaction Explorer: Towards a Knowledge Map of Organic Chemistry To Support Dynamic Assessment and Personalized Instruction. In: *Enhancing Learning with Online Resources, Social Networking, and Digital Libraries*. eds. R. E. Belford, J. W. Moore, and H. E. Pence 1060, 191209. doi: 10.1021/bk-2010-1060.ch011

da Silva Júnior, J. N., Nobre, D. J., do Nascimento, R. S., Torres, G. S., Leite, A. J. M., Monteiro, A. J., et al. (2018). Interactive computer game that engages students in reviewing organic compound nomenclature. *J. Chem. Educ.* 95, 899–902. doi: 10.1021/acs.jchemed.7b00793

Dickmann, T., Opfermann, M., Dammann, E., Lang, M., and Rumann, S. (2019). What you see is what you learn? The role of visual model comprehension for academic success in chemistry. *Chem. Educ. Res. Pract.* 20, 804–820. doi: 10.1039/C9RP00016J

Eitemüller, C., Trauten, F., Striewe, M., and Walpuski, M. (2023). Digitalization of multistep chemistry exercises with automated formative feedback. *J. Sci. Educ. Technol.* 32, 453–467. doi: 10.1007/s10956-023-10043-2

Farhat, N. J., Stanford, C., and Ruder, S. M. (2019). Assessment of student performance on Core concepts in organic chemistry. *J. Chem. Educ.* 96, 865–872. doi: 10.1021/acs.jchemed.8b00913

Field, A. (2018). Discovering statistics using IBM SPSS statistics. *5th* Edn. Washington DC: SAGE.

Fischer, V., Walpuski, M., Lang, M., Letzner, M., Manzel, S., Motté, P., et al. (2021). Studienerfolg und Studienabbruch [what influences the decision to drop out? Longitudinal analyses of the interplay between study satisfaction, content knowledge and dropout intention in chemistry, engineering and social sciences]. *ZeHf* 4, 55–80. doi: 10.3224/zehf.v4i1.05

Fleischer, J., Leutner, D., Brand, M., Fischer, H., Lang, M., Schmiemann, P., et al. (2019). Vorhersage des Studienabbruchs in naturwissenschaftlich-technischen Studiengängen [prediction of student drop-out in STEM study programs]. *Z. Erzieh.* 22, 1077–1097. doi: 10.1007/s11618-019-00909-w

Flynn, A. B. (2023). Orgchem101 learning lab. Available at: https://www.orgchem101.com (Accessed July 20, 2023).

Flynn, A. B., Caron, J., Laroche, J., Daviau-Duguay, M., Marcoux, C., and Richard, G. (2014). Nomenclature101.com: a free, student-driven organic chemistry nomenclature learning tool. *J. Chem. Educ.* 91, 1855–1859. doi: 10.1021/ed500353a

Graulich, N., and Bhattacharyya, G. (2017). Investigating students' similarity judgments in organic chemistry. *Chem. Educ. Res. Pract.* 18, 774–784. doi: 10.1039/C7RP00055C

Grossman, R. B., and Finkel, R. (2023). ACE - achieving chemistry excellence organoc. Lexington: University of Kentucky.

Grove, N. P., and Lowery Bretz, S. (2012). A continuum of learning: from rote memorization to meaningful learning in organic chemistry. *Chem. Educ. Res. Pract.* 13, 201–208. doi: 10.1039/C1RP90069B

Hailikari, T. K., and Nevgi, A. (2010). How to diagnose at-risk students in chemistry: the case of prior knowledge assessment. *Int. J. Sci. Educ.* 32, 2079–2095. doi: 10.1080/09500690903369654

Hauck, D., Melle, I., and Steffen, A. (2021). Molecular orbital theory—teaching a difficult chemistry topic using a CSCL approach in a first-year university course. *Educ. Sci.* 11:485. doi: 10.3390/educsci11090485

Heller, S. R., McNaught, A., Pletnev, I., Stein, S., and Tchekhovskoi, D. (2015). Inchi, the IUPAC international chemical identifier. *J. Chem.* 7:23. doi: 10.1186/s13321-015-0068-4

Heublein, U., Richter, J., and Schmelzer, R. (2020). Die Entwicklung der Studienabbruchquoten in Deutschland [The development of student drop-out rates in Germany] (DZHW Brief 3|2020). Hannover, Germany: DZHW. doi: 10.34878/2020.03.dzhw_brief

IUPAC (Ed.) (2023). The IUPAC international chemical identifier (INCHI): International Union of Pure and Apllied Chemistry. Avilable at: https://iupac.org/who-we-are/divisions/division-details/inchi/

Jiang, C., Jin, X., Dong, Y., and Chen, M. (2016). Kekule.Js: an open source JavaScript Chemoinformatics toolkit. *J. Chem. Inf. Model.* 56, 1132–1138. doi: 10.1021/acs.jcim.6b00167

Johnson, C., and Priest, A. H. (2014). "The feedback principle in multimedia learning" in The Cambridge handbook of multimedia learning. ed. R. E. Mayer (Cambridge: Cambridge University Press).

Kalyuga, S., Chandler, P., Tuovinen, J., and Sweller, J. (2001). When problem solving is superior to studying worked examples. *J. Educ. Psychol.* 93, 579–588. doi: 10.1037/0022-0663.93.3.579

Klepsch, M., Schmitz, F., and Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* 8:1997. doi: 10.3389/fpsyg.2017.01997

Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., and Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educ. Psychol. Rev.* 35. doi: 10.1007/s10648-023-09738-0

Laugwitz, B., Held, T., and Schrepp, M. (2008). "Construction and evaluation of a user experience questionnaire" in Lecture notes in computer science. HCI and usability for education and work. ed. A. Holzinger (Berlin Heidelberg: Springer), 63–76.

Lenhard, W., and Lenhard, A. (2023). Psychometrica. Availabel at: https://www.psychometrica.de/effektstaerke.html (Accessed July 20, 2023).

Leppink, J., Paas, F., van der Vleuten, C. P. M., van Gog, T., and van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* 45, 1058–1072. doi: 10.3758/s13428-013-0334-1

Linacre, J. M. (2023). A User's guide to Winsteps® Ministep: Rasch-model computer programs. Available at: https://www.winsteps.com/winman/copyright.htm (Accessed July 23, 2024).

Malik, K., Martinez, N., Romero, J., Schubel, S., and Janowicz, P. A. (2014). Mixed-methods study of online and written organic chemistry homework. *J. Chem. Educ.* 91, 1804–1809. doi: 10.1021/ed400798t

Mayer, R. E., and Pilegard, C. (2014). "Principles for managing essential processing in multimedia learning: segmenting, pre-training, and modality principles" in The Cambridge handbook of multimedia learning. ed. R. E. Mayer (Cambridge: Cambridge University Press), 316–344.

Paas, F., Renkl, A., and Sweller, J. (2003a). Cognitive load theory and instructional design: recent developments. *Educ. Psychol.* 38, 1–4. doi: 10.1207/S15326985EP3801_1

Paas, F., and Sweller, J. (2014). "Implication of cognitive load theory for multimedia learning" in The Cambridge handbook of multimedia learning. ed. R. E. Mayer (Cambridge: Cambridge University Press).

Paas, F., Tuovinen, J. E., Tabbers, H., and van Gerven, P. W. M. (2003b). Cognitive load measurement as a means to advance cognitive load theory. *Educ. Psychol.* 38, 63–71. doi: 10.1207/S15326985EP3801_8

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., and Aubteen Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: optimizing learner involvement in instruction. *Educ. Technol. Res. Dev.* 53, 25–34. doi: 10.1007/BF02504795

Penn, J. H., and Al-Shammari, A. G. (2008). Teaching reaction mechanisms using the curved arrow neglect (CAN) method. *J. Chem. Educ.* 85:1291. doi: 10.1021/ed085p1291

Rajan, K., Brinkhaus, H. O., Zielesny, A., and Steinbeck, C. (2020). A review of optical chemical structure recognition tools. *J. Chem.* 12:60. doi: 10.1186/s13321-020-00465-0

Rodemer, M., Eckhard, J., Graulich, N., and Bernholt, S. (2021). Connecting explanations to representations: benefits of highlighting techniques in tutorial videos on students' learning in organic chemistry. *Int. J. Sci. Educ.* 43, 2707–2728. doi: 10.1080/09500693.2021.1985743

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., and Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instr. Sci.* 43, 93–114. doi: 10.1007/s11251-014-9328-3

Spanjers, I. A., van Gog, T., Wouters, P., and van Merriënboer, J. J. (2012). Explaining the segmentation effect in learning from animations: the role of pausing and temporal cueing. *Comput. Educ.* 59, 274–280. doi: 10.1016/j.compedu.2011.12.024

Steinbeck, C., Hoppe, C., Kuhn, S., Floris, M., Guha, R., and Willighagen, E. L. (2006). Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. *Curr. Pharm. Des.* 12, 2111–2120. doi: 10.2174/138161206777585274

Striewe, M. (2016). An architecture for modular grading and feedback generation for complex exercises. *Sci. Comput. Program.* 129, 35–47. doi: 10.1016/j.scico.2016.02.009

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cogn. Sci.* 12, 257–285. doi: 10.1207/s15516709cog1202_4

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educ. Psychol. Rev.* 22, 123–138. doi: 10.1007/s10648-010-9128-5

Sweller, J., van Merrienboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 261–292. doi: 10.1007/s10648-019-09465-5

Taskin, V., Bernholt, S., and Parchmann, I. (2017). Student teachers' knowledge about chemical representations. *Int. J. Sci. Math. Educ.* 15, 39–55. doi: 10.1007/s10763-015-9672-z van Alten, D. C., Phielix, C., Janssen, J., and Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: a meta-analysis. *Educ. Res. Rev.* 28, 100281–100218. doi: 10.1016/j.edurev.2019.05.003

van Gog, T., Kirschner, F., Kester, L., and Paas, F. (2012). Timing and frequency of mental effort measurement: evidence in favour of repeated measures. *Appl. Cogn. Psychol.* 26, 833–839. doi: 10.1002/acp.2883

WileyPLUS and University of California. (2023). *Reaction Explorer: Synthesis Explorer and Mechanism Explorer Reaction Tutorial System*. Available at: https://cdb.ics.uci.edu/cgibin/tutorial/ReactionTutorialSetupWeb.py (Accessed July 20, 2023).

Xie, B., and Salvendy, G. (2000). Prediction of mental workload in single and multiple tasks environments. *Int. J. Cogn. Ergon.* 4, 213–242. doi: 10.1207/S15327566IJCE0403_3

frontiers | Frontiers in Education

TYPE Perspective
PUBLISHED 11 November 2024
DOI 10.3389/feduc.2024.1376087

Check for updates

Extending access for all chemistry students with extended reality

Maria T. Gallardo-Williams[1]* and Cathi Dunnagan[2]

[1]Office for Faculty Excellence and Department of Chemistry, North Carolina State University, Raleigh, NC, United States, [2]Digital Education and Learning Technology Applications, North Carolina State University, Raleigh, NC, United States

OPEN ACCESS

EDITED BY
Gautam Bhattacharyya,
Missouri State University, United States

REVIEWED BY
Sarbjeet Kaur,
Curia Global, United States
Kevin Stewart,
Harding University, United States

*CORRESPONDENCE
Maria T. Gallardo-Williams
✉ mtgallar@ncsu.edu

RECEIVED 24 January 2024
ACCEPTED 17 October 2024
PUBLISHED 11 November 2024

CITATION
Gallardo-Williams MT and Dunnagan C (2024) Extending access for all chemistry students with extended reality.
Front. Educ. 9:1376087.
doi: 10.3389/feduc.2024.1376087

Equal access to instructor's time and attention in chemistry classes and laboratories can be a barrier experienced by students from historically excluded groups. An instructor's own biases will determine the nature of their interaction with students, and even well-meaning instructors can interact with students in slightly different ways, which might prevent certain students from having access to all the available instructional resources for the class. This is an additive problem, which may or may not be recognized in peer and student evaluations, and an issue that might escape self-reflection even in educators that are committed to diversity, inclusion, and justice. This issue conflates both actual and perceived biases, introducing a complex dynamic between instructor and student. Extended reality (XR) provides an avenue to generate materials that can be used to enhance or replace classroom instruction with a great degree of realism. In this paper we will discuss the implementation of a set of virtual reality (VR) organic chemistry labs. We will show that XR learning tools are by their very nature accessible and inclusive of a wide variety of students and will provide evidence from student reflections that shows that students from historically excluded groups find the XR content offered in our virtual reality labs more personal than in-person activities covering the same material.

KEYWORDS

virtual reality, augmented reality, extended reality, access, inclusion, organic chemistry

1 Introduction

Historically, chemistry has been a field dominated by white males. This situation was perpetuated for years due to disparities in access to quality education (Woolston, 2020; Van Dusen et al., 2022). This is evident in the underrepresentation of women in chemistry authorship (Cotton and Seiple, 2021) and gender differences in chemistry achievement and participation (National Science Foundation, National Center for Science and Engineering Statistics, 2017). Even with well-intentioned educators, inherent biases may shape interactions in ways that inadvertently limit access to course material. This subtle yet pervasive issue poses a complex dynamic between instructors and students, often manifesting through both actual and perceived differences in student-instructor interactions, including differential time and attention from instructors. Research consistently shows that students from historically marginalized groups face significant challenges in accessing equal opportunities in chemistry classes and laboratories (Kimble-Hill et al., 2020; Leopold and Smith, 2020; Neill et al., 2018). These challenges are often exacerbated by instructors' biases, which can limit students' access to instructional resources (Kimble-Hill et al., 2020). To address this, it is crucial for instructors to recognize their responsibility in supporting interactions and to implement activities such as reflective group work activities that benefit all students, especially from marginalized groups. (Leopold and Smith, 2020). Additionally, creating structured, inclusive classroom environments can help promote equitable participation (Neill et al., 2018).

Some interventions, such as growth mindset programs (Fink et al., 2018), and other supplemental programs to enhance at-risk students' academic skills and performance (Botch et al., 2007; Rath et al., 2012; Shields et al., 2012), have been shown to improve the performance of historically excluded minorities in chemistry. Small group activities, such as Peer-Lead Team Learning (PLTL) (Lewis and Lewis, 2005; Frey et al., 2018) and Process Oriented Guided Inquiry Learning (POGIL) are also known to support achievement in students from marginalized groups, particularly in large classes such as General Chemistry (Ruder and Hunnicutt, 2008; Vincent-Ruz et al., 2020).

These studies underscore the need for continued efforts to address the gender and racial disparities in the field of chemistry. However, a common feature of most supplemental chemistry achievement programs is their intensive nature (Lockie and Lanen, 1994). Providing one-on-one instruction, tutoring, or extended contact beyond what is offered in the classroom or laboratory setting imposes an extra burden on existing instructors or requires the hiring of additional instructional team members (Coletti et al., 2013). It also requires that students are motivated and available to engage with the intervention being offered at specific times/venues, and that instructors engage with all students in equitable ways. Although the intention is to be commended, many such programs do not persist due to staffing issues or limited participation, and it is our own personal experience that remedial interventions, even those with demonstrated benefits, are the first to go in lean budgetary conditions.

It is crucial to reconsider how we frame supplemental assistance in education. Labeling it as "remedial" or targeting specific groups, such as students with disabilities, can be detrimental. Adopting a positive framing approach and adhering to Universal Design for Learning (UDL) principles can enhance the accessibility and success of teaching and learning materials for all students (Brand et al., 2012; Tobin, 2021). UDL, a framework that emphasizes the design of instructional materials and activities to meet the needs of diverse learners, is particularly effective in this regard. It encourages the use of multiple means of representation, engagement, and expression, ensuring that all students can access and engage with the content. This approach is especially beneficial in online education, where it can increase student retention and engagement (Tobin, 2021). By incorporating UDL principles, educators can create a more inclusive and effective learning environment for all students. Extended reality applications can be informed by UDL principles (Parrish et al., 2021), and by their very nature make learning accessible in ways that were not previously available.

Recognizing the need for a transformative approach, many educators around the world have spent time and effort in the generation of materials that can be used to remedy such gaps. An alternative to the more traditional instructor-intensive inclusive resources is the use of pre-recorded video resources, which became prominent in many programs in the last 20 years, as the cost of creating and publishing videos decreased worldwide (Blonder et al., 2013; Box et al., 2017; Gillette et al., 2017). A logical next step was the creation of virtual and augmented reality materials to be used in chemistry courses (Dunnagan and Gallardo-Williams, 2020; Wright and Oliver-Hoyo, 2021). Such resources were widely repurposed as remote instructional materials, mostly to cover chemistry laboratory instruction during the recent COVID-19 pandemic disruption of instruction (Kelley, 2021), and have been relegated to a secondary plane in its aftermath (Link and Gallardo-Williams, 2022). This article delves into the realm of Extended Reality as a groundbreaking, and in some ways unexpected, solution to issues of unequal access. Extended reality (XR) is an umbrella term to refer to augmented reality, virtual reality, and mixed reality. XR emerges as a powerful tool to overcome the limitations of traditional classroom instruction, offering on-demand, universal immersive experiences that can enhance or even replace conventional teaching methods in certain settings. The degree of immersion in digital teaching tools within the XR realm can be highly individualized depending on the instructor and students' choices (Aguayo, 2021).

The integration of Extended Reality into education represents a paradigm shift in how we approach and deliver learning experiences. XR, encompassing both augmented and virtual reality, enables educators to create dynamic and interactive content that transcends the confines of traditional pedagogy. XR technologies have the potential to improve learning outcomes by enhancing interactivity and immersion (Logeswaran et al., 2021; Herur-Raman et al., 2021; Paye et al., 2021). In this context, we present a VR intervention: a suite of virtual reality organic chemistry labs (Dunnagan and Gallardo-Williams, 2020) to show that XR activities are by their very nature accessible and inclusive of a wide variety of students. We bring up this example and its student outcomes as part of a rapidly developing field. By addressing the crucial intersection of technology, diversity, and education, this work seeks to extend access and create a sense of belonging and community for chemistry students through the innovative use of XR tools.

2 Virtual reality: creating immersive learning environments

In 2019 we introduced a series of organic chemistry labs designed to be experienced in a fully immersive virtual environment (Dunnagan et al., 2019). These labs transcend the limitations of physical spaces, providing students with the opportunity to explore intricate reactions and phenomena in a risk-free, yet realistic setting. The immersive nature of VR not only enhances comprehension but also caters to diverse learning styles, ensuring that students from all backgrounds can actively participate and thrive in their educational journey.

When we designed the VR lab experiences, we were intentional in recruiting a diverse pool of teaching assistants, using inclusive language, and offering examples that were unbiased and appealing to all our students. Our strategy involved crafting VR laboratories that prioritized inclusivity and diversity. This encompassed curating content and leveraging the talents of several teaching assistants. Conscious efforts were made to reflect diversity in the virtual TA pool, aligning with the races, ethnicities, and gender orientations present in the Department of Chemistry at North Carolina State University. This intentional inclusivity received positive feedback from students, with 30% expressing favorable comments on the diverse virtual TA pool in course evaluations. This might appear at first sight to be a nod to diversity and inclusion initiatives; however, it goes beyond that, as it allows students to find a TA that looks like them among the pool of available instructors. While white students might have a variety of role models to choose from, the same is not true for students from marginalized backgrounds. The simple act of including diverse TAs can go a long way in creating community for these students.

To ensure a bias-free learning environment, the content of VR laboratories underwent a rigorous review process. Sourced from experienced TAs and reviewed by faculty, the scripts were scrutinized to eliminate any racially charged or gendered terms. The goal was to create a supportive and inclusive learning experience, reducing terms that may be perceived as biased. This multilayered review process contributed to the success of these realistic VR simulations in minimizing instructor bias and offering struggling students an opportunity to experience the best that the instructor has to offer.

However, when we tested the VR labs with the first group of students, we were surprised to discover in the data from historically marginalized groups in our sample population that their comments were not solely focused on the aspects of diversity and inclusion that we had included as part of the VR lab design. Instead, most of the students from historically marginalized groups also mentioned how much they appreciated the time and attention given to them by the virtual teaching assistant. These lifelike simulations appeared to mitigate instructor bias, as the material was generated independently of student presence, providing an opportunity for any student who may face challenges in a traditional setting to access the instructor's expertise under the best possible conditions.

The challenge of ensuring equitable access to support in higher education poses a significant hurdle, particularly for students from historically marginalized groups. Instructor-student interactions are greatly shaped by individual biases, creating varying levels of engagement and potentially limiting some students' ability to fully benefit from course content. This nuanced issue, frequently unnoticed in evaluations by peers and students, can persist even within faculty committed to fostering diversity and inclusion. It underscores the intricate relationship between instructor bias, student perception, and actual access to educational resources. In addressing this challenge, virtual reality (VR) emerges as a promising tool to create immersive materials that can either complement or substitute traditional classroom instruction with a high degree of realism and engagement.

The analysis of data collected from the user study evaluating our VR materials for organic chemistry laboratories also sheds light on how students interact with VR instructors. Feedback provided by students from historically excluded groups highlights the perceived impartiality of the virtual instructor, the ability to engage with course material independently, and the advantage of remote access as key desirable aspects of the VR learning experience (Dunnagan et al., 2019). During the evaluation, 23% of students from historically excluded groups expressed satisfaction with the direct attention received from the virtual Teaching Assistant (TA). The prerecorded TA interactions, designed to appear accessible and supportive, proved to be a significant factor for students, highlighting the unanticipated impact on their learning experience. One of the student reflections included the following statement:

> "I have never had a TA look me in the eye for so long and take such care to explain a concept to me. This felt very personal."

The traditional classroom setting, prone to implicit biases, can hinder the educational experience for students, especially in close-quarters interactions such as laboratory settings. Virtual reality (VR) laboratories address this challenge by providing constant, supportive availability of TAs, minimizing the likelihood of biased interactions. Moreover, the virtual TA dynamic eliminates the potential for intimidation, a common concern in in-person labs (Dunn et al., 2023), leading to positive feedback from students who have not perceived the virtual instructor as intimidating or impatient.

Conscious efforts were made to reflect diversity in the virtual TA pool, aligning with the races, ethnicities, and gender orientations present in the Department of Chemistry at North Carolina State University. This intentional inclusivity received positive feedback from students, with 30% expressing favorable comments on the diverse virtual TA pool in course evaluations. This might appear at first sight to be a nod to diversity and inclusion initiatives, however it goes beyond that, allowing students to find a TA that looks like them among the pool of available instructors. While white students might have a variety of role models to choose from, the same is not true for students from marginalized backgrounds. The simple act of including diverse TAs can go a long way in creating community for these students.

In addition to the perceived impartiality of the instructor and the diversity of available instructors, students appreciated the ability to engage independently with the material and the convenience of remote access. Student comments, often shared during open microphone portions of interviews, were candid and personal, providing valuable insights into the overall positive impact of the VR learning experience. We had ample opportunity to test the usefulness of the VR lab experiences during the COVID-19 pandemic disruption, and found the online labs to be a suitable substitute for the in-person experience (Dunnagan and Gallardo-Williams, 2020).

In the summer of 2020, we used the Meaningful Learning in the Laboratory Instrument (MLLI) (Galloway and Bretz, 2015) to gauge students' cognitive and affective expectations before the virtual lab course and their experiences with virtual reality upon course completion. Students who participated in virtual reality laboratories reported more positive affective experiences than anticipated, expressing minimal frustration or confusion during the laboratory sessions (Williams et al., 2022).

Our exploration not only showcases the technical aspects of the VR organic chemistry lab experiences but also emphasizes their inherent accessibility and inclusivity. The crux of our discussion revolves around best practices for engaging students in introductory organic chemistry lab courses, with a particular emphasis on reaching those from historically marginalized groups. Through compelling evidence drawn from student reflections, we make the point that XR content fosters a more personalized and inclusive learning environment compared to traditional in-person activities covering the same material for some students.

In addressing the diverse needs of students, a hybrid approach combining both in-person and virtual experiences emerges as a favorable compromise, a trend that is currently seen in the health science education field (Pottle, 2019). This strategy accommodates the preference for in-person laboratory sessions while embracing the inclusive advantages not only for students from historically excluded minorities, but for any student experiencing attendance constraints like pregnancy or military deployment, or even students with unique safety considerations. Since chemistry is a laboratory-based discipline, and laboratory instruction is central to its mission (Seery, 2020), enhancing laboratory offerings with XR options might open the field to students that traditionally would not be able to participate.

3 Conclusion

In conclusion, this article highlights the current advances and future directions of an XR project in chemistry education. The presented VR intervention demonstrates the potential to overcome barriers related to unequal access and biased interactions. By fostering inclusivity and accessibility, XR can contribute to creating a more equitable and engaging learning environment for all chemistry students. The most impactful part of this work relates to an unintended outcome: Providing all students with equal access to the instructor's attention and all the resources available in an online environment. As we continue to explore and refine XR applications, the goal is to pave the way for a transformative shift in how chemistry education is delivered and experienced.

Looking forward, the integration of XR in chemistry education holds significant promise. The success of our VR intervention suggests that these technologies can address longstanding issues of unequal access and biased interactions in traditional classroom settings. Future research should explore the scalability of XR interventions and their long-term impact on student success, particularly for historically excluded minorities. Additionally, efforts should be made to integrate XR seamlessly into curricula, ensuring sustained benefits beyond temporary disruptions like the COVID-19 pandemic.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving humans were approved by the North Carolina State University Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MG-W: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. CD: Conceptualization, Investigation, Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding for both projects described in this article was provided by North Carolina State's Digital Education and Learning Technology Applications Unit, in the form of grants and technical support for product development.

Acknowledgments

The authors are thankful to their respective units for supporting this work, and especially to Michael Cuales, Arthur Earnest, David Tredwell, John Gordon, and Michael Castro for their technical expertise and creative vision.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aguayo, C. (2021). Mixed reality (XR) research and practice: exploring a new paradigm in education. *Pacific J. Technol. Enhanced Learn.* 3, 41–42. doi: 10.24135/pjtel.v3i1.104

Blonder, R., Jonatan, M., Bar-Dov, Z., Benny, N., Rap, S., and Sakhnini, S. (2013). Can you tube it? Providing chemistry teachers with technological tools and enhancing their self-efficacy beliefs Chem. *Educ. Res. Pract.* 14, 269–285. doi: 10.1039/C3RP00001J

Botch, B., Day, R., Vining, W., Stewart, B., Hart, D., Rat, K., et al. (2007). Effects on student achievement in general chemistry following participation in an online preparatory course. Chemprep, a voluntary, self-paced, online introduction to chemistry. *J. Chem. Educ.* 84, 547–553. doi: 10.1021/ed084p547

Box, M. C., Dunnagan, C. L., Hirsh, L. A. S., Cherry, C. R., Christianson, K. A., Gibson, R. J., et al. (2017). Qualitative and quantitative evaluation of three types of student-generated videos as instructional support in organic chemistry laboratories. *J. Chem. Educ.* 94, 164–170. doi: 10.1021/acs.jchemed.6b00451

Brand, S. T., Favazza, A. E., and Dalton, E. M. (2012). Universal design for learning: a blueprint for success for all learners. *Kappa Delta Pi Rec.* 48, 134–139. doi: 10.1080/00228958.2012.707506

Coletti, K. B., Covert, M., Dimilla, P. A., Gianino, L., Reisberg, R., and Wisniewski, E. O. (2013). Understanding the factors influencing student participation in supplemental instruction in freshman chemistry. Atlanta, GA: Paper presented at the 120th ASEE annual conference and exposition.

Cotton, A. D., and Seiple, I. B. (2021). Examining gender imbalance in chemistry authorship. *ACS Chem. Bio* 16, 2042–2046. doi: 10.1021/acschembio.1c00142

Dunn, A. L., Decker, D. M., Cartaya-Marin, C. P., Cooley, J., Finster, D. C., Hunter, K. P., et al. (2023). Reducing risk: strategies to advance laboratory safety through diversity, equity, inclusion, and respect. *J. Am. Chem. Soc.* 145, 11468–11471. doi: 10.1021/jacs.3c03627

Dunnagan, C. L., Dannenberg, D. A., Cuales, M. P., Earnest, A. D., Gurnsey, R. M., and Gallardo-Williams, M. T. (2019). Production and evaluation of a realistic immersive virtual reality organic chemistry laboratory experience: infrared spectroscopy. *J. Chem. Educ.* 97, 258–262. doi: 10.1021/acs.jchemed.9b00705

Dunnagan, C. L., and Gallardo-Williams, M. T. (2020). Overcoming physical separation during COVID-19 using virtual reality in organic chemistry laboratories. *J. Chem. Educ.* 97, 3060–3063. doi: 10.1021/acs.jchemed.0c00548

Fink, A., Cahill, M. J., McDaniel, M. A., Hoffman, A., and Frey, R. F. (2018). Improving general chemistry performance through a growth mindset intervention: selective effects on underrepresented minorities. *Chem. Educ. Res. Pract.* 19, 783–806. doi: 10.1039/C7RP00244K

Frey, R. F., Fink, A., Cahill, M. J., McDaniel, M. A., and Solomon, E. D. (2018). Peer-led team learning in general chemistry I: interactions with identity, academic preparation, and a course-based intervention. *J. Chem. Educ.* 95, 2103–2113. doi: 10.1021/acs.jchemed.8b00375

Galloway, K. R., and Bretz, S. L. (2015). Development of an assessment tool to measure students' meaningful learning in the undergraduate chemistry laboratory. *J. Chem. Educ.* 92, 1149–1158. doi: 10.1021/ed500881y

Gillette, A. A., Winterrowd, S. T., and Gallardo-Williams, M. T. (2017). Training students to use 3-D model sets via peer-generated videos facilitates learning of difficult concepts in an introductory organic chemistry course. *J. Chem. Educ.* 94, 960–963. doi: 10.1021/acs.jchemed.7b00155

Herur-Raman, A., Almeida, N. D., Greenleaf, W., Williams, D., Karshenas, A., and Sherman, J. H. (2021). Next-generation simulation—integrating extended reality technology into medical education. *Front. Virtual Real.* 2:693399. doi: 10.3389/frvir.2021.693399

Kelley, E. W. (2021). LAB Theory, HLAB Pedagogy, and Review of Laboratory Learning in Chemistry during the COVID-19 Pandemic. *J. Chem. Educ.* 98, 2496–2517. doi: 10.1021/acs.jchemed.1c00457

Kimble-Hill, A. C., Rivera-Figueroa, A., Chan, B. C., Lawal, W. A., Gonzalez, S., Adams, M. R., et al. (2020). Insights gained into marginalized students access challenges during the covid-19 academic response. *J. Chem. Educ.* 97, 3391–3395. doi: 10.1021/acs.jchemed.0c00774

Leopold, H., and Smith, A. (2020). Implementing reflective group work activities in a large chemistry lab to support collaborative learning. *Educ. Sci.* 10:7. doi: 10.3390/educsci10010007

Lewis, S. E., and Lewis, J. E. (2005). Departing from lectures: an evaluation of a peer-led guided inquiry alternative. *J. Chem. Educ.* 82, 135–139. doi: 10.1021/ed082p135

Link, R. D., and Gallardo-Williams, M. (2022). We should keep developing digital laboratory resources in the postpandemic era. *J. Chem. Educ.* 99, 519–520. doi: 10.1021/acs.jchemed.1c01197

Lockie, N. M., and Lanen, R. J. (1994). Supplemental instruction for college chemistry courses. *New Directions for Teaching and Learning*, 63–74.

Logeswaran, A., Munsch, C., Chong, J. Y., Ralph, N., and McCrossnan, J. (2021). The role of extended reality technology in healthcare education: towards a learner-centered approach. *Future Healthc. J.* 8, e79–e84. doi: 10.7861/fhj.2020-0112

National Science Foundation, National Center for Science and Engineering Statistics (2017). Women, minorities, and persons with disabilities in science and engineering: 2017, special report NSF 17–310. Arlington, VA: National Center for Science and Engineering Statistics, Directorate for Social, Behavioral and Economic Sciences, National Science Foundation.

Neill, C., Cotner, S., Driessen, M., and Ballen, C. J. (2018). Structured learning environments are required to promote equitable participation. *Chem. Educ. Res. Pract.* 20, 197–203. doi: 10.1039/C8RP00169C

Parrish, A. H., Kouo, J. L., Carey, L. B., and Swanson, C. (2021). "Implementing universal design for learning in the virtual learning environment" in Handbook of research on transforming teachers' online pedagogical reasoning for engaging K-12 students in virtual learning. eds. M. Niess and H. Gillow-Wiles (Philadelphia, PA: IGI Global), 42–66.

Paye, C. L., Dunnagan, C. L., Tredwell, D., and Gallardo-Williams, M. T. (2021). Connecting the dots: Lewis structure builder web app as a review tool for organic chemistry. *J. Chem. Educ.* 98, 2704–2708. doi: 10.1021/acs.jchemed.1c00213

Pottle, J. (2019). Virtual reality and the transformation of medical education. *Future Healthc. J.* 6, 181–185. doi: 10.7861/fhj.2019-0036

Rath, K. A., Peterfreund, A., Bayliss, F., Runquist, E., and Simonis, U. (2012). Impact of supplemental instruction in entry-level chemistry courses at a midsized public university. *J. Chem. Educ.* 89, 449–455. doi: 10.1021/ed100337a

Ruder, S. M., and Hunnicutt, S. S. (2008). "POGIL in chemistry courses at a large urban university: a case study" in Process-oriented guided inquiry learning. eds. R. S. Moog and J. N. Spencer (Washington, DC: American Chemical Society), 133–147.

Seery, M. K. (2020). Establishing the laboratory as the place to learn how to do chemistry. *J. Chem. Educ.* 97, 1511–1514. doi: 10.1021/acs.jchemed.9b00764

Shields, S. P., Hogrebe, M. C., Spees, W. M., Handlin, L. B., Noelken, G. P., Riley, J. M., et al. (2012). A transition program for underprepared students in general chemistry: diagnosis, implementation, and evaluation. *J. Chem. Educ.* 89, 995–1000. doi: 10.1021/ed100410j

Tobin, T. J. (2021). Reaching all learners through their phones and universal design for learning. *J. Adult Learn. Knowl. Innov.* 4, 9–19. doi: 10.1556/2059.03.2019.01

Van Dusen, B., Nissen, J., Talbot, R. M., Huvard, H., and Shultz, M. (2022). A QuantCrit investigation of society's educational debts due to racism and sexism in chemistry student learning. *J. Chem. Educ.* 99, 25–34. doi: 10.1021/acs.jchemed.1c00352

Vincent-Ruz, P., Meyer, T., Roe, S. G., and Schunn, C. D. (2020). Short-term and long-term effects of POGIL in a large-enrollment general chemistry course. *J. Chem. Educ.* 97, 1228–1238. doi: 10.1021/acs.jchemed.9b01052

Williams, N. D., Gallardo-Williams, M. T., Griffith, E. H., and Bretz, S. L. (2022). Investigating meaningful learning in virtual reality organic chemistry laboratories. *J. Chem. Educ.* 99, 1100–1105. doi: 10.1021/acs.jchemed.1c00476

Woolston, C. (2020). White men still dominate in UK academic science. *Nature* 579:622. doi: 10.1038/d41586-020-00759-1

Wright, L., and Oliver-Hoyo, M. (2021). Development and evaluation of the H NMR MoleculAR application. *J. Chem. Educ.* 98, 478–488. doi: 10.1021/acs.jchemed.0c01068

Frontiers in Education

Explores education and its importance for individuals and society

A multidisciplinary journal that explores research-based approaches to education for human development. It focuses on the global challenges and opportunities education faces, ultimately aiming to improve educational outcomes.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact