education sciences

Special Issue Reprint

Engaged Student Learning and Inclusive Teaching Practices in Higher Education Chemistry

Edited by
Jack F. Eichler and Oluwatobi Odeleye

mdpi.com/journal/education

MDPI

Engaged Student Learning and Inclusive Teaching Practices in Higher Education Chemistry

Engaged Student Learning and Inclusive Teaching Practices in Higher Education Chemistry

Guest Editors

Jack F. Eichler
Oluwatobi Odeleye

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editors
Jack F. Eichler
Chemistry Department
University of California
Riverside, CA
USA

Oluwatobi Odeleye
Department of Chemistry
West Virginia University
Morgantown, VA
USA

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Education Sciences* (ISSN 2227-7102), freely accessible at: https://www.mdpi.com/journal/education/special_issues/XMLJQ058X1.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-4417-3 (Hbk)
ISBN 978-3-7258-4418-0 (PDF)
https://doi.org/10.3390/books978-3-7258-4418-0

Contents

About the Editors

Jack F. Eichler

Professor Eichler's scholarly activity focuses on developing, implementing, and assessing more engaging and inclusive learning environments for large-enrollment general chemistry courses. He was the principal investigator for the NSF Transforming Undergraduate Education in STEM (TUES) project, which developed, implemented, and disseminated a series of problem-based case studies from a two-year general chemistry/organic chemistry sequence. He also participated in the NSF Improving Undergraduate STEM Education (IUSE) project, in which he adapted these problem-based case studies to in-class activities for flipped classroom modules. His current work focuses on optimizing online learning environments for chemistry-related learning objectives, how the flipped classroom can be leveraged to facilitate a deeper conceptual understanding of core chemistry concepts, and using mastery outcome assessment structures to reduce historical equity gaps in large-gateway STEM courses.

Oluwatobi Odeleye

As a chemical education researcher, Dr. Odeleye's research interests revolve around different factors that influence students' attitudes towards chemistry and STEM fields in general. She believes that the classroom environment and teaching practices employed by instructors play essential roles in shaping these attitudes, which research has shown can affect student retention in STEM fields, especially regarding students from under-represented groups. She recently ran a workshop funded by the National Science Foundation (NSF-DUE #2232453) that brought math and chemistry educators together to improve their capacity to research methods for improving engagement in their introductory STEM classrooms. She strongly advocates for learner-centered classrooms and continues to investigate ways to aid instructors in making their STEM classroom environments more learner-oriented.

Preface

The retention of students in STEM degree pathways continues to be a problem in higher education, particularly for students from minority groups. Addressing this is of the utmost importance in recruiting and retaining high-quality individuals for the 21st-century STEM workforce and, perhaps more importantly, helping all students who are passionate about STEM achieve their career goals. It is in this context that this Special Issue of *Education Sciences* was conceived. "Engaged Student Learning and Inclusive Teaching in Higher Education Chemistry" presents 10 articles on how student engagement can be improved in higher education chemistry courses, potentially impacting retention and success in STEM pathways, and interventions to improve outcomes more directly for historically disadvantaged demographic groups.

Jack F. Eichler and Oluwatobi Odeleye
Guest Editors

education sciences

Editorial

Summary of the Special Issue from the Guest Editors

Jack F. Eichler [1,*] and Oluwatobi Odeleye [2]

[1] Department of Chemistry, University of California, Riverside, CA 92521, USA
[2] Department of Chemistry, West Virginia University, Morgantown, WV 26506, USA; oluwatobi.odeleye@mail.wvu.edu
* Correspondence: jack.eichler@ucr.edu

We are excited to bring you this Special Issue of *Education Sciences* titled "Engaged Student Learning and Inclusive Teaching in Higher Education Chemistry". The retention of students in STEM degree pathways continues to be a problem in higher education, particularly for students from minority groups. Addressing this problem is of the utmost importance in recruiting and retaining high-quality individuals for the 21st-century STEM workforce and, perhaps more importantly, helping all students passionate about STEM achieve their career goals. It is in this context that this Special Issue was conceived.

This Special Issue presents 10 articles describing several ways student engagement can be improved in higher education chemistry courses, potentially impacting retention and success in STEM pathways, or interventions that may lead more directly to improving outcomes for historically disadvantaged demographic groups. Articles more broadly related to ***improving classroom engagement and learning outcomes*** include: (1) a report on the use of pre-class activities that use multimedia simulations to foster conceptual understanding of particulate-level models (Herrington & Sweeder, 2024); (2) a study looking at the role of language and reading comprehension skills on learning chemistry (Buell & Pazicni, 2024); (3) a systematic review of specifications grading that finds emergent themes with respect to how practitioners use this assessment scaffold (Howitz, McKnelly & Link, 2025); (4) qualitative research that investigates the nature of peer-to-peer questions during in-class collaborative group work (Dahl, et al., 2025); and (5) a qualitative study that explores the student and faculty attitudes towards the learning and teaching of reaction mechanisms in organic chemistry (Odeleye & Tieu, 2025).

Articles that more directly ***investigate interventions that impact success and retention for historically underserved demographic*** include: (1) a study on the impact of supplemental instruction programs on student success and equity gaps (Pham & Ye, 2024); (2) a quantitative study that demonstrates how a mastery assessment structure can improve equity gaps in large enrollment general chemistry (Hartman & Eichler, 2024); (3) a qualitative study that describes how students with disabilities view their experiences classroom accommodations (DeKorver, 2025); (4) an essay on culturally relevant pedagogies and culturally responsive teaching, with a framework for re-envisioning chemistry classroom culture (Wang & Bussey, 2025); and (5) a meta-analysis of previous research related to improving academic success in introductory chemistry courses, with an emphasis on how effective learning environments can lead to equitable outcomes (Chestnut & Johnson, 2025).

We hope the articles published here encourage chemistry education researchers to continue thinking about ways to evaluate interventions being used in their classroom and engage with tools like supplemental instruction, mastery/specifications grading, and mental models in their research. For practitioners, this Special Issue aims to help chemistry instructors from all backgrounds make their classes more engaging and inclusive. Several articles in this issue highlight the importance of being student-focused as we seek to enhance the classroom environment and be more inclusive in our teaching. For example, one article highlights the importance of implementing teaching practices based on students' feedback to accommodate students from all backgrounds, and another discusses how peer-to-peer interactions influence students' experiences in general chemistry.

This Special Issue reflects the continued effort of researchers and instructors to reduce equity gaps and increase interest and retention in STEM pathways for all students. We would also like to emphasize that, though this Special Issue is devoted to chemistry education research and practice, the broader findings and lessons learned can and should be applied across the various STEM disciplines.

Author Contributions: Writing—original draft preparation, J.F.E. and O.O.; writing—review and editing, J.F.E. and O.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: JFE and OO each contributed articles to this Special Issues, however these manuscripts went through the normal peer review process and the *Education Sciences* editorial board approved these for publication. JFE and OO were not involved in the editorial decision-making process for these two articles.

List of Contributions

Buell, R. W., & Pazicni, S. (2024). Who learns from reading texts in general chemistry? *Education Sciences*, *14*, 1287. https://doi.org/10.3390/educsci14121287.

Chestnut, J., & Johnson, C. C. (2025). Factors influencing students' academic success in introductory chemistry: A systematic literature review. *Education Sciences*, *15*, 413. https://doi.org/10.3390/educsci15040413.

Dahl, T. M., Grieger, K., Miller, S., & Nyachwaya, J. (2025). Exploring the nature and role of students' peer-to-peer questions during an in-class collaborative activity. *Education Sciences*, *15*, 229. https://doi.org/10.3390/educsci15020229.

DeKorver, B. (2025). Putting inclusion into practice: Five commitments toward equity in teaching. *Education Sciences*, *15*, 84. https://doi.org/10.3390/educsci15010084.

Hartman, J. D., & Eichler, J. F. (2024). Implementing mastery grading in large enrollment general chemistry: Improving outcomes and reducing equity gaps. *Education Sciences*, *14*, 1224. https://doi.org/10.3390/educsci14111224.

Herrington, D. G., & Sweeder, R. D. (2024). Using simulations and screencasts in online preclass activities to support student building of mental models. *Education Sciences*, *14*, 115. https://doi.org/10.3390/educsci14020115.

Howitz, W. J., McKnelly, K. J., & Link, R. D. (2025). Delving into the design and implementation of specifications grading systems in higher education. *Education Sciences*, *15*, 83. https://doi.org/10.3390/educsci15010083.

Odeleye, O., & Tieu, N. (2025). Students' & faculty members' attitudes towards learning and teaching reaction mechanisms in organic chemistry. *Education Sciences*, *15*, 357. https://doi.org/10.3390/educsci15030357.

Pham, D., & Ye, L. (2024). Lifting the gate: Evaluation of supplemental instruction program in chemistry. *Education Sciences*, *14*, 1196. https://doi.org/10.3390/educsci14111196.

Wang, S., & Bussey, T. J. (2025). Re-envisioning classroom culture in an introductory general chemistry course: Description of a course redesign project. *Education Sciences*, *15*, 307. https://doi.org/10.3390/educsci15030307.

education sciences

MDPI

Article

Factors Influencing Students' Academic Success in Introductory Chemistry: A Systematic Literature Review

Jessica Chestnut and Carla C. Johnson *

Department of STEM Education, North Carolina State University, Raleigh, NC 27695, USA; jnchestn@ncsu.edu
* Correspondence: carlacjohnson@ncsu.edu

Abstract: Undergraduate introductory chemistry is a gatekeeping course preventing students from persisting in STEM degree programs. It is important to understand students' experiences of introductory chemistry and better support students as this course traditionally has high attrition and failure rates. This systematic literature review examines the factors of academic success for undergraduates in introductory chemistry courses and aims to understand how these factors differ for varying student groups. A meta-analysis of 35 articles uncovered three emergent themes for promoting students' academic success: course design, instructional tools and resources, and student learning and characteristics. Most notably, active learning environments, metacognitive assessments, and student affective variables such as identity and motivation emerged as significant predictors of students' academic success. Additionally, this review demonstrates how differences in student demographics, achievement levels, affective variables, and participation in chemistry affect the extent to which students succeed in this course. Student demographics were most frequently reported to cause disparities in course performance, with students from historically underrepresented populations exhibiting the most disadvantages in overall course performance. These findings signify the importance of creating effective learning environments in introductory chemistry for students from diverse backgrounds to achieve equitable outcomes and sustain STEM interest.

Keywords: chemistry; learning; teaching; retention; achievement; confidence

1. Introduction

The capacity to meet growing economic and workforce demands relies on the retention and completion rates within Science, Technology, Engineering, and Mathematics (STEM) undergraduate degree programs. A report by the President's Council of Advisors on Science and Technology (2012) projected a shortage of one million STEM graduates over the decade spanning 2012 to 2022. The demand for skilled professionals in STEM fields is on an upward trajectory, with the U.S. Bureau of Labor Statistics (2020) projecting continued faster growth for STEM employment compared to non-STEM employers. Another study by the U.S. Bureau of Statistics in (Fayer et al., 2017) reported roughly 8.6 million STEM jobs available in life science, mathematics, computer science, physical science, and engineering. This report also projected STEM employment will increase by 8.8% between 2018 and 2028 in the United States, with computer science occupations showing the largest growth (U.S. Bureau of Labor Statistics, 2020). There is a critical need for educational institutions to not only increase the number of graduates in STEM fields but also to ensure their readiness for the workforce.

Despite growing workforce demands for STEM professionals, approximately 40% of students who embark on a STEM major actually persist to their graduation (Chen, 2013; President's Council of Advisors on Science & Technology, 2012). STEM retention in 2013 was reported at a national average of 48%, with community colleges only reporting 30% retention for their STEM majors (Chen, 2013). Attrition rates for STEM majors have also been observed to be much higher than attrition rates of non-STEM majors. The perceived difficulty of STEM programs has been noted as a concern that diverts students from potential STEM degree programs and STEM careers. While many students will express interest in STEM fields as high school seniors, only 21% of high school graduates are academically underprepared for the coursework required in introductory STEM courses (ACT, 2018). Long-term retention and graduation of STEM undergraduates has been previously "predicted significantly by cumulative grade point average, financial need, aid (work-study, loan, and gift), gender, ethnicity, years living on campus, high school rank (HSR), ACT composite, out-of-state residence, and STEM status" (Whalen & Shelley, 2010, p. 45). When controlling for financial variables, Whalen and Shelley (2010) observed students from historically underrepresented populations to be significantly less likely to be retained in STEM fields or graduate within six years when compared to well-represented populations.

The disparities in retention and completion rates among different demographic groups adds an additional layer of complexity to persistence in STEM undergraduate degree programs. While retention rates for White, Asian, and male students in STEM are relatively higher, historically underrepresented gender and racial minorities exhibit markedly lower completion rates. Figueroa et al. (2017) found that only 25% of African American, Latino, and Native American students complete a STEM degree within six years, compared to 44.5% of their White and Asian counterparts. Another study found that degree completion rates are roughly half that of the national average for students from underrepresented racial and ethnic minorities. Disparities among gender are also observed in the retention and persistence of STEM graduates. Chen (2013) found that 43% of female community college students switched out of STEM majors while only 29% of their male peers left. Female STEM undergraduates are frequently underrepresented in STEM fields and achievement gaps have been observed to favor male students (Chen, 2013; Figueroa et al., 2017; Whalen & Shelley, 2010). Given the disparities in STEM that exist for students from traditionally underrepresented populations, research efforts should be focused on retaining a diverse group of students and reducing achievement gaps.

Significance of Introductory Chemistry

Several of the STEM degree programs with high rates of attrition see dropouts occurring within the first year of taking introductory coursework (Figueroa et al., 2017; Freeman et al., 2014; Seymour & Hewitt, 2000; Stone et al., 2018). In demand STEM programs such as engineering, computer science, and healthcare fields, all require introductory courses for undergraduates to pass before they can begin their major specific course content. Chemistry is often hailed as the "central science" given its foundational concepts on the structure of matter and interdisciplinary applications to many STEM degree programs (Tai et al., 2005). Introductory chemistry courses are a prerequisite for several STEM majors and are often considered a gateway, weed-out, or killer course for students (Bressoud, 2020; Lloyd & Eckhardt, 2010; Tai et al., 2005). Undergraduates cannot continue in their intended STEM major if they do not successfully pass their introductory coursework. Many students enrolled in introductory chemistry courses do not intend on majoring in chemistry or biochemistry fields (Gillespie, 1991), which poses an additional challenge to cater to a large audience of STEM undergraduates.

A comprehensive meta-analysis by Freeman et al. (2014) reported an average failure rate of 33.4% for undergraduate STEM introductory courses. When looking specifically at introductory chemistry, institutions have reported failure rates exceeding 50% (Chambers & Blake, 2008). Academic success in introductory chemistry courses has lasting impacts on STEM persistence and student graduation rates. A study of community college chemistry students found that 32% of students received a D, F, or W in the course (Cohen & Kelly, 2019). This study further reports that 49% of the students failing chemistry changed their majors and 80% of those majors were changed to a non-STEM field. An analysis by Stone et al. (2018) revealed a direct correlation between grades in first-semester general chemistry and overall graduation rates. Students who achieved passing grades in introductory chemistry had a 73% graduation rate while those that failed the course had a 43% graduation rate. Only 14% of the students failing the chemistry course changed their majors to non-STEM fields and eventually graduated (Stone et al., 2018). Given these insights, it is evident that introductory chemistry courses are integral to the future of STEM education and workforce development. Prior studies seeking to understand academic success for students in introductory chemistry courses have revealed that student-centered learning, alternative assessments, student agency, and identity-based interventions help increase academic performance and persistence in STEM programs (Bressoud, 2020; Chen, 2013; Freeman et al., 2014; Stone et al., 2018; Ryoo & Winkelmann, 2021). It is important that students in post-secondary STEM coursework are engaged in active learning, as defined by Hartikanean et al. (2019) as "student-centered and activating instructional methods and instructor-led activities" as opposed to more "traditional, content-centered approaches, such as lecturing". Addressing the challenges within chemistry introductory courses through innovative teaching methods, supportive learning environments, and consideration of demographic influences is crucial for fostering success among a diverse student population.

This systematic literature review (SLR) examined the empirical research base regarding delivery of the introductory chemistry at higher education institutions and associated student academic success in this gateway course. The goal of this study was to examine research published over the most recent decade (2014–2023) and to generate a comprehensive list of factors which facilitate academic success for students from various ethnic/racial, gender, and socio-economic backgrounds in introductory chemistry coursework. York et al. (2015) has created a theoretical framework of academic success that comprises "academic achievement; acquisition of knowledge, skills, and competencies; and persistence and retention" (p. 2). For the purposes of this SLR, academic success will use traditional measures based on the students' academic performance and defined as passing the first semester general or introductory chemistry course with at least a 70% (C or higher) grade. As prior literature has shown, gateway introductory courses play a significant role in the academic journey of students, and introductory chemistry, in particular, has suffered from high rates of attrition (Chambers & Blake, 2008; Figueroa et al., 2017; Stone et al., 2018). The rationale for this research is grounded in the position that undergraduate introductory chemistry is required for many STEM-related degree programs, thus, it is critical to better understand how to support students well so that they complete the course and move forward in their programs.

This systematic literature review was focused on addressing the following questions: (1) What factors contribute to undergraduate students' academic success in introductory chemistry courses, and (2) How does success in undergraduate introductory chemistry differ based on student background?

2. Materials and Methods

2.1. Database Search and Article Selection

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021) guided to this SLR which focused on analyzing the factors of academic success for undergraduates in introductory chemistry. The purpose of this study was to determine how this field of inquiry had evolved over the most recent decade (2014–2023). The initial search for articles was conducted with multiple databases (ERIC, Academic Search Complete, Google Scholar, and Proquest); however, due to significant overlap in articles found, the authors decided to include the use of Academic Search Complete (ACS) and Educational Resources Information Center (ERIC). ACS was chosen for its multidisciplinary applications in higher education contexts and provides full-text literature for over 9000 journals (NC State University Libraries, 2024). ERIC was also selected as it serves as the primary database in education research contexts. The ERIC database covers a range of educational institutions and contains over 750 journals (NC State University Libraries, 2024). Both ACS and ERIC databases produced articles based on the following Boolean search term: introductory chemistry AND undergraduate AND (success OR achievement OR performance). Research studies were limited to full-text articles in scholarly, peer reviewed journals written only in English over the past nine full years (2014 to 2023). The ACS database produced 57 articles, and the ERIC database produced 77 articles. These results were exported as a CSV file and imported into Excel to undergo further screening of selection criteria.

2.2. Screening and Excluded Studies

PRISMA guidelines were utilized for identifying and screening research studies to be used in this SLR (Page et al., 2021). Figure 1 summarizes the process for identifying research articles from the ACS and ERIC databases, screening of the abstracts and full-text articles, and exclusion criteria to finalize the 35 studies to be included in this SLR. Of the articles produced from the initial ACS (n = 57) and ERIC (n = 77) database searches, seven duplicate articles were removed resulting in 127 articles subject for abstract screening. These abstracts were screened to ensure they were relevant to the guiding research questions of this SLR study. Articles were excluded from the study based on the following criteria: (1) articles not focused on undergraduate introductory chemistry, (2) articles which include discipline specific courses like organic chemistry, biochemistry, or upper-level courses, (3) articles focused on other STEM non-chemistry disciplines, (4) articles not focused on undergraduates in higher education, (5) articles that only discussed the chemistry laboratory and not classroom context, and (6) articles that were not empirical studies or published in peer reviewed scholarly journals. For example, during the abstract screening one research article titled "Using touch-screen technology, apps, and blogs to engage and sustain high school students' interest in chemistry topics" was rejected for its use in high school contexts and not higher education (Kim et al., 2014). Another research article titled "The roles of motivation and metacognition in producing self-regulated learners of college physical science: a review of empirical studies" was rejected for its use of interdisciplinary STEM fields such as physics (McDowell, 2019). The research article titled "Simple and inexpensive 3D printed filter fluorometer designs: User-friendly instrument models for laboratory learning and outreach activities" was rejected for its laboratory context and use in upper-level courses (Porter et al., 2017). A total of 86 abstracts were rejected during this initial screening process.

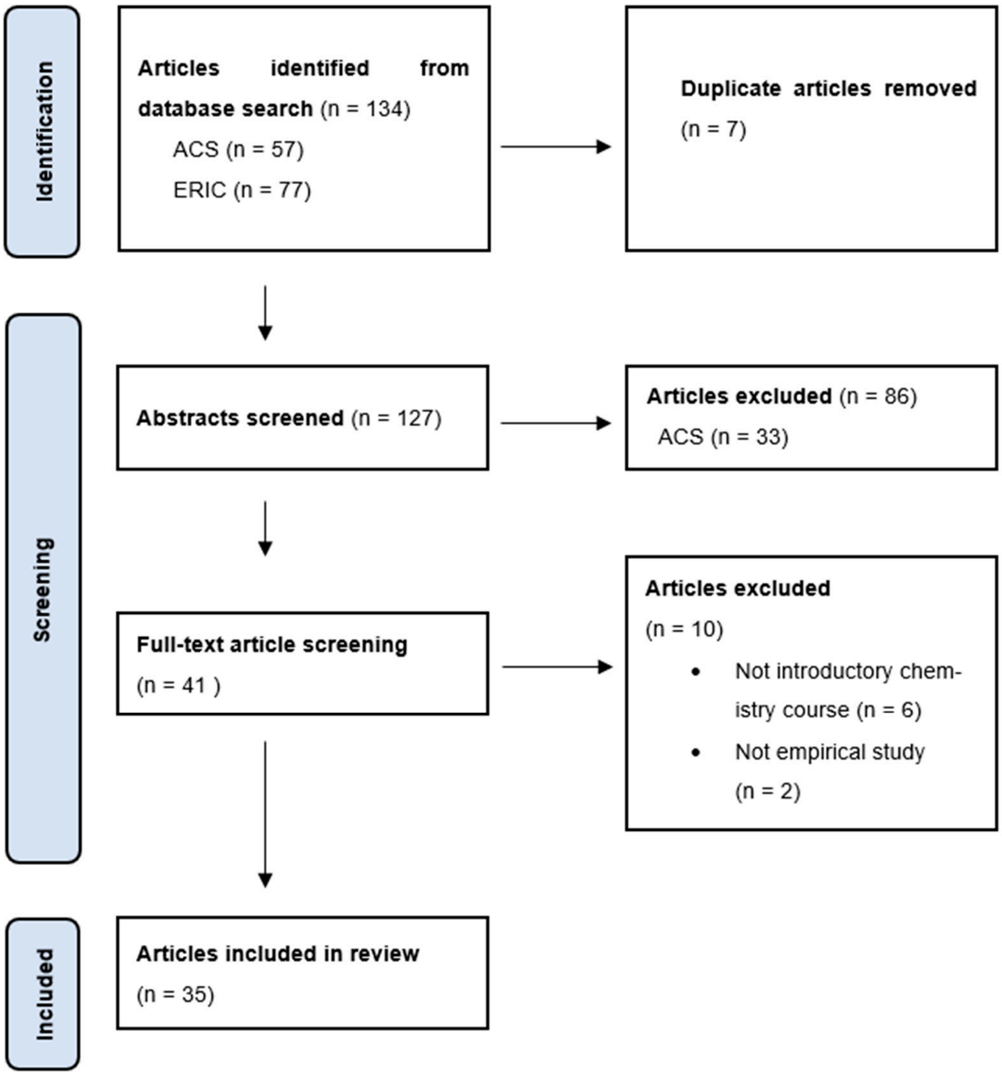


Figure 1. PRISMA Flow Diagram of Review Protocol. Note. Adapted from Page et al. (2021).

Once the abstract screening was complete, a total of 41 articles moved on to the next phase of full-text screening and were downloaded into the reference manager Zotero. Upon full-text reading, ten articles were rejected based on exclusion criteria. Six studies were rejected for focusing on upper-level chemistry courses or other science disciplines. For example, the research article titled "Benefits of a game-based review module in chemistry courses for nonmajors" was rejected due to its focus on a biochemistry course and not an introductory chemistry course (Stringfield & Kramer, 2014). Two studies were excluded for not being primary and empirical research studies. For instance, the research article titled "Radical awakenings—A new teaching paradigm using social media" was excluded as this was a review of a conference paper and not an empirical study (Sorensen-Unruh, 2017). Finally, the last two studies were excluded for not focusing on students in introductory chemistry courses and instead focusing on the validation of survey instruments. After the full-text screening process, a total of 35 research articles were included in the study.

2.3. Data Analysis

Rather than using a conceptual framework to structure the coding process, thematic synthesis was used to analyze the articles in this study where themes emerged from the primary studies (Thomas & Harden, 2008). The three stages of this approach include the following: (1) coding selected text, (2) development of descriptive/emergent themes, and (3) generation of analytical themes (Thomas & Harden, 2008). First, the 35 research articles included in this study were categorized based on the year of publication, journal, location of the study, and study type based on methodology. Descriptive statistics were also used to quantify these variables. The articles then underwent thematic analysis to understand the factors of academic success for introductory chemistry students and how these factors may differ based on various groups of students. Thematic analysis was guided by the research questions and took an iterative approach to refine inductive codes developed from research article analysis and produce emergent themes (Schreier, 2019). To answer the first research question, commonalities in initial codes were used to categorize factors of academic success. The articles were then analyzed thoroughly a second time to refine the categories and then finalized as analytical themes that were refined to answer the first research question regarding the factors of academic success of undergraduate introductory chemistry students. Thematic analysis was also used to answer the second research question. Studies were categorized to determine how they were investigating various group differences that affect academic success. For each article, findings, discussion, and implications were coded independently by the lead author. The second author on this manuscript independently coded a subset of articles which provided inter-rater reliability for the SLR analysis. Our team independently assessed intercoder reliability (ICR) for each code at the end of each "testing round". When 80% agreement on 95% of codes (Miles & Huberman, 1994) was achieved, the coding framework was finalized, and coding of remaining data will continue independently.

3. Results

The results of this study are presented through addressing the two research questions. The first research question examines the factors that are associated with student success in introductory chemistry coursework. The second question delves deeper to determine any nuanced differences by student subgroup. Details regarding the 35 articles included are provided in Table 1 such as journal of publication, location of study, and study type based on methodology. The year 2023 was the most common (n = 7, 20%) year of publication for articles in this study. Other years of publication included 9% (n = 3) of articles in 2014, 9% (n = 3) in 2015, 9% (n = 3) in 2016, 11% (n = 4) in 2017, 9% (n = 3) in 2018, 11% (n = 4) in 2019, 9% (n = 3) in 2020, 9% (n = 3) in 2021, and 6% (n = 2) in 2022. While the year 2024 was included in the search criteria, the database searches did not produce articles from 2024 since this search was conducted in February 2024.

The most common journal that 46% of studies (n = 16) were published in was the Journal of Chemical Education. Other journals that research articles were published in included Psychology of Women Quarterly (n = 1), Educational Technology Research and Development (n = 1), Journal of Experimental Education (n = 1), Educational Psychology (n = 1), Computers and Education (n = 1), Annals of the New York Academy of Sciences (n = 1), School Science and Mathematics (n = 1), Journal of Counseling Psychology (n = 1), Journal of the Scholarship of Teaching and Learning (n = 2), Chemistry Education Research and Practice (n = 4), EURASIA Journal of Mathematics, Science and Technology Education (n = 1), College Student Journal (n = 1), Journal of Science Education and Technology (n = 1), Electronic Journal of Science Education (n = 1), and Journal of Teaching and Learning with Technology (n = 1).

Table 1. Descriptive Overview of Selected Literature.

Article Reference	Journal	Study Location	Study Type
(An et al., 2022)	Journal of Chemical Education	U.S.	Quantitative
(Bergey et al., 2023)	Journal of Experimental Education	U.S.	Mixed Methods
(Bokosmaty et al., 2019)	Journal of Chemical Education	Australia	Mixed Methods
(Brown et al., 2015)	Journal of the Scholarship of Teaching and Learning	U.S.	Quantitative
(Bunce et al., 2017)	Journal of Chemical Education	U.S.	Mixed Methods
(Chan & Bauer, 2014)	Journal of Chemical Education	U.S.	Quantitative
(Carpenter et al., 2020)	Journal of Teaching and Learning with Technology	U.S.	Quantitative
(Clark, 2023)	Journal of Chemical Education	U.S.	Quantitative
(Cosio & Williamson, 2019)	Journal of Science Education and Technology	U.S.	Quantitative
(Cracolice & Busby, 2015)	Journal of Chemical Education	U.S.	Quantitative
(Edwards et al., 2023)	Journal of Chemical Education	U.S.	Quantitative
(Fink et al., 2020)	Chemistry Education Research	U.S.	Quantitative
(French et al., 2023)	Psychology of Women Quarterly	U.S.	Quantitative
(Gilewski et al., 2022)	Chemistry Education Research and Practice	U.S.	Mixed Methods
(Gulacar et al., 2019)	EURASIA Journal of Mathematics, Science and Technology Education	U.S.	Mixed Methods
(Hardin & Longhurst, 2016)	Journal of Counseling Psychology	U.S.	Quantitative
(Hawker et al., 2016)	Journal of Chemical Education	U.S.	Quantitative
(He et al., 2018)	Computers and Education	U.S.	Quantitative
(Kyne et al., 2023)	Chemistry Education Research and Practice	Australia	Mixed Methods
(Msonde & Van Aalst, 2017)	Educational Technology Research and Development	Tanzania	Mixed Methods
(Ott et al., 2018)	Journal of the Scholarship of Teaching and Learning	U.S.	Quantitative
(Perez et al., 2023)	Annals of the New York Academy of Sciences	U.S.	Quantitative
(Philipp et al., 2016)	Electronic Journal of Science Education	U.S.	Quantitative
(Revell, 2014)	Journal of Chemical Education	U.S.	Quantitative
(Smith et al., 2018)	Journal of Chemical Education	U.S.	Quantitative
(Talanquer, 2017)	Journal of Chemical Education	U.S.	Quantitative
(Talanquer & Pollard, 2017)	Journal of Chemical Education	U.S.	Quantitative
(Tang et al., 2014)	Journal of Chemical Education	U.S.	Quantitative
(Tashiro & Talanquer, 2021)	Journal of Chemical Education	U.S.	Quantitative
(Todd et al., 2021)	Chemistry Education Research and Practice	U.S.	Quantitative
(Van Duser et al., 2021)	College Student Journal	U.S.	Quantitative
(Wang et al., 2021)	Educational Psychology	U.S.	Mixed Methods
(Wong et al., 2023)	School Science and Mathematics	U.S.	Mixed Methods
(Ye et al., 2015)	Journal of Chemical Education	U.S.	Quantitative

Quantitative methodology was the most common method used by researchers as 71% of studies (n = 26) used quantitative analysis to understand factors of academic success for introductory chemistry undergraduates. The type of quantitative analysis varied and included methods such as surveys and descriptive statistics (Hardin & Longhurst, 2016; Revell, 2014; Ye et al., 2015), ANOVA or MANCOVA (French et al., 2023; Todd et al., 2021; Wang et al., 2021), linear or logistic regression analysis (Carpenter et al., 2020; Philipp et al., 2016; Van Duser et al., 2021), cluster analysis (Brown et al., 2015; Chan & Bauer, 2014), latent profile analysis (Perez et al., 2023), Wilcoxon sum-rank tests (An et al., 2022; Bancroft et al., 2020; Smith et al., 2018), chi squared tests (Cosio & Williamson, 2019; Talanquer & Pollard, 2017), and Mann–Whitney U tests (Hawker et al., 2016; Tang et al., 2014). The use of surveys to quantify factors or student characteristics was the most common method, with 58% (n = 18) of research studies using surveys and descriptive statistics.

Only nine studies (26%) used a mixed methods approach, and their quantitative methods were similar to the studies listed above. The qualitative methods used in the mixed methods studies included content analysis of a survey with open ended items (Bokosmaty et al., 2019), thematic analysis of focus groups (Gilewski et al., 2022), a word association test with open coding and network analysis (Gulacar et al., 2019), thematic analysis of student emails and evaluations of the course (Kyne et al., 2023), content analysis of concept maps (Wang et al., 2021; Wong et al., 2023), content analysis of student generated question logs (Bergey et al., 2023), student interviews and content analysis of discussion forums (Msonde & Van Aalst, 2017), and student interviews (Bunce et al., 2017). Only three studies had qualitative methods that involved direct interaction with students such as focus groups or interviews. No studies were identified that used only qualitative methodology to investigate factors of academic success for introductory chemistry students.

The vast majority (91%) of studies (n = 32) were conducted in the United States. Two studies were conducted in Australia and one study was conducted in Tanzania. It was also observed that the studies mostly investigated undergraduate students from four-year public universities. The only differences in institution types that were observed included one institution at a U.S. Naval Academy (Bunce et al., 2017) and one at a predominately Hispanic serving institution (An et al., 2022). Only one research article mentioned that their study was implemented at both a community college and public university (Gilewski et al., 2022).

3.1. Research Question 1: Factors of Student Success for Introductory Chemistry Undergraduates

Research question one explored the factors that influence academic success in undergraduates in introductory chemistry courses. Thematic analysis was used to identify the three emergent themes of course design, instructional tools and resources, and student learning and characteristics and the associated categories (Table 2).

Table 2. Summary of the Three Factors of Academic Success for Introductory Chemistry.

Theme	Category	Article Reference
Course design and learning environment (n = 18)	Instructional methods (n = 5)	Bancroft et al. (2020); Bokosmaty et al. (2019); Clark (2023); He et al. (2018); Tashiro and Talanquer (2021)
	Curricular redesign (n = 4)	Ott et al. (2018); Philipp et al. (2016); Smith et al. (2018): Talanquer and Pollard (2017)
	Assessment reform (n = 8)	Gulacar et al. (2019); Gilewski et al. (2022); Talanquer (2017); Tang et al. (2014); Todd et al. (2021); Wang et al. (2021); Wong et al. (2023); Ye et al. (2015)
	Online learning (n = 1)	Msonde and Van Aalst (2017)
Course resources and feedback (n = 5)	Instructional resource (n = 3)	Bunce et al. (2017); Cosio and Williamson (2019); Revell (2014)
	Instructor feedback (n = 2)	Carpenter et al. (2020); Kyne et al. (2023)
Student learning and characteristics (n = 12)	Student characteristics (n = 8)	Brown et al. (2015); Chan and Bauer (2014); Edwards et al. (2023); Fink et al. (2020); French et al. (2023); Hardin and Longhurst (2016); Perez et al. (2023); Van Duser et al. (2021)
	Student learning (n = 4)	An et al. (2022); Bergey et al. (2023); Cracolice and Busby (2015); Hawker et al. (2016)

Theme 1: Course Design and Learning Environment

The first emerging theme identified in this SLR was course design. A total of 18 studies (51%) discussed how the design of the chemistry course and learning environment contributed to overall student success in introductory chemistry courses. This theme consisted of two different types of course designs: active learning environments and an online learning environment.

Instructional Methods. A study by Tashiro and Talanquer (2021) compared student outcomes between a traditional, lecture-based chemistry course and a reformed course incorporating active learning strategies at a large public research university. Using hierarchical linear and logistical modeling, the study analyzes differences in student performance and assesses whether educational reforms help mitigate achievement gaps. Instructors either taught using active learning strategies with in-class clicker questions and collaborative learning activities for students or had a more traditional and lecture-based approach. The reformed course had higher course grade averages compared to the traditional course for all academic index ranges. It was observed that students with lower academic index ranges (lower-performing students) benefited from their participation in the reformed course; however, the final exams are different for the two courses (Tashiro & Talanquer, 2021). In relation to academic success in chemistry courses, the research suggests that structured learning environments, inclusive assessment practices, and balancing exam and coursework weights can significantly impact student achievement.

One of the research studies employed the strategy of a fully flipped instructional model to promote students' academic success in their introductory chemistry course (Bancroft et al., 2020). In the fully flipped model, lectures are prerecorded for students to watch outside of the classroom and in-person time traditionally reserved for lectures are instead used for students to work on problem solving in groups. Students in traditional lecture courses and the non-traditional flipped course had their course performance compared to understand the effects of a flipped instructional model. This study noted that withdrawals and D/F grades were decreased for all groups of students when using a flipped model (Bancroft et al., 2020).

Rather than creating a fully flipped instructional model, two studies only partially flipped their classrooms to investigate the effects on student performance in their chemistry courses (Bokosmaty et al., 2019; He et al., 2018). The partially flipped model still requires students to read materials or watch prerecorded lectures outside of the classroom but will still hold some time in class for lectures. Students are also asked to work on problem sets and engage in discussions in the partially flipped model. He et al. (2018) noted in their study that short-term academic achievement was not significantly affected for student final exam grades in the semester of taking the partially flipped introductory chemistry course, but their subsequent chemistry exam grades improved. This study demonstrated that a partially flipped instructional model had a long-term and not short-term benefit for academic achievement. Additionally, this study found that students in the partially flipped model exhibited higher levels of motivation and positive course perception compared to those in a traditional lecture course (He et al., 2018). An Australian university's implementation of a partially flipped instructional model also reported higher student satisfaction with the quality of teaching and learning resources available (Bokosmaty et al., 2019). As opposed to He et al.'s study, Bokosmaty et al. (2019) demonstrated significant short-term learning gains through improved academic performance in course grades. They also reported higher rates of student retention within the three introductory chemistry course sections the partially flipped model was implemented in.

Clark (2023) also studied reformed course utilizing active learning strategies as compared to traditional lecture-based courses. The researchers analyzed data from approximately 9000 students across multiple semesters, including both in-person and online courses. By using statistical regression analyses and controlling factors such as incoming preparation (measured by ACT scores), the study assessed the impact of teaching practices on achievement gaps. During the in-person courses, the instructor of one reformed course was involved in Modeling Instruction pedagogy and emphasized metacognition throughout the course to help students optimize their learning strategies. The instructor of the second reformed course was influenced by peer instruction pedagogy and uses a flipped classroom (Clark, 2023). Both courses used active learning and student-centered learning approaches, which significantly reduced the achievement gap for historically underrepresented populations compared to traditional lecture-based courses. In addition to active learning, it was observed that course structure and instructional strategies such as pre-class and post-class assignments, student engagement initiatives, and metacognitive learning interventions contributed to student success in the general chemistry course (Clark, 2023). The emergency switch to online instruction, however, had minimal to no student–student interaction or student-centered teaching and adapted a more didactic approach for the online instructional mode. The study concludes that while active learning is valuable, structured learning experiences outside of class time are equally important for reducing disparities in student performance.

Curricular Redesign. Curricular redesign was the focus of a study based out of the University of Arizona (Talanquer & Pollard, 2017). The chemistry department decided to reform their introductory chemistry course by creating a new curriculum focusing on student-centered approaches. Students engaged in an inquiry-based learning chemistry curriculum with some sections taking place in collaborative learning spaces to promote student interactions. Student performance with the reformed curriculum was seen to significantly increase while the failure rates of the standardized American Chemical Society (ACS) exam decreased from 38.5% to 29.2% (Talanquer & Pollard, 2017). Students that completed the reformed introductory chemistry course also showed improved subsequent organic chemistry course performance with more students receiving A's and a decrease in D and F grades compared to students not previously enrolled in the reformed introductory chemistry course sections. Similarly to the study from Bokosmaty et al. (2019), the student-centered approach from Talanquer and Pollard (2017) demonstrated long-term learning gains. Other active learning strategies included a POGIL model to enhance chemistry students' academic performance. One study using POGIL strategies with student teams working on inquiry problems reported an increase in student academic performance and a decrease in withdrawal rates for the course (Ott et al., 2018). Smith et al. (2018) also used an active learning environment with a POGIL approach in their introductory chemistry course. Since this course typically served students in a nursing major, the POGIL activities had a heavy emphasis on healthcare contexts. Data collected over ten semesters at two universities demonstrated a significant positive increase in students' chemistry self-concept. This result indicated that an active learning environment increased the students' belief in their ability to succeed in the introductory chemistry course due to the POGIL activities (Smith et al., 2018). The incorporation of health-related scenarios in the course increased student engagement, which also resulted in a significant increase in course grades.

Philipp et al. (2016) investigated the use of undergraduate teaching assistants (UTAs) in chemistry recitation sections. The UTAs had previously taken the same chemistry course and were employed due to their success in the course and recommendations from faculty. The recitation sections were student-centered and employed a peer mentoring approach. Other recitation sections were traditionally run by graduate teaching assistants (GTAs). The presence of the UTAs only significantly boosted the final exam scores for students that already had above average college GPA's (Philipp et al., 2016). The study found that UTA recitation sections increased the persistence of students in the next subsequent chemistry course regardless of the students' academic achievement. There were, however, no significant differences found in the final exam scores of students with the UTA-led recitation sections compared to the traditional recitations led by GTAs (Philipp et al., 2016). Overall, the studies using active learning approaches examined in this SLR suggest that active learning environments can improve students' academic performance and contribute to long-term learning gains.

Assessment Reform. Chemistry assessments were the most frequently reported tool revealed in this review of previous research. Eight studies reported that certain types of assessments can promote academic achievement for chemistry undergraduate students. Assessment types varied and included the prediction of wrong answers, measurement of linked concepts, concept maps, complexity of stoichiometry problems, word association tests, and retrieval quizzes. Talanquer (2017) created an assessment that required students to predict the wrong answers that another student may choose if they were only relying on their intuition to answer the question. This assessment intervention aimed to improve the analytical skills of students to help them work through chemical reasoning and become more aware of intuitive traps. Talanquer (2017) found that students did have an increase in their concept inventories and analytical skills, along with an increase in overall academic performance in the course.

Another assessment technique highlighted in this SLR was the measurement of linked concepts to understand how students' link chemistry content to their existing knowledge structures (Todd et al., 2021; Ye et al., 2015). These assessments were targeted at specific chemistry concepts and were incorporated in homework assignments and in-class exams to understand the conceptual links and misconceptions of chemistry students. Students would use the measurement of linked concept (MLC) assessment to respond true or false to items that would link big picture chemical concepts across a variety of topics. This assessment technique was able to identify the common misconceptions that students held about course content so that instructors could address these in class and improve the learning gains of the students (Ye et al., 2015). Gilewski et al. (2022) also used an assessment to measure linked concepts in introductory chemistry courses at a public university and community college. The MLC assessments were shown to significantly predict students' final exam scores in introductory chemistry. Additionally, the MLC assessment was paired with a metacognitive exercise where students "needed to look at the learning objectives they missed and write a plan for mastering the missed learning objectives" (Gilewski et al., 2022, p. 878). When paired with the metacognitive exercise, MLC scores significantly improved by 18%; however, this improvement did not translate to a statistically significant increase in final exam performance. Gilewski et al. (2022) also demonstrated that students had more engagement with course material because of the MLC assessment and metacognitive exercise. They reported that 87% of students reported revisiting learning objectives they missed and another 54% formulated plans to address these gaps such as reviewing lecture notes or engaging with more practice problems.

Another common type of assessment identified was the use of concept maps. One instance asked students to use concept maps in a chemistry course for the topic of enthalpy (Wang et al., 2021). In this study, students either had to write a paragraph about the concept map, fill in concept blanks on a partially completed concept map, or fill in labels on a partially completed concept map (Wang et al., 2021). It was found that students who were required to translate the concept map into complete sentences had significantly better performance on open-ended exam questions. However, no significant differences existed for multiple choice exam questions based on the different treatments of concept maps (Wang et al., 2021). It was also observed that students translating the concept map spent more time on this activity than the other concept map activities but spent less time answering posttest questions. This result suggests a more efficient retrieval and application of knowledge (Wang et al., 2021). Another study utilizing concept map assessments asked students to either fill in a blank concept map or correct an incorrect concept map based on the topic of electrochemistry (Wong et al., 2023). The concept map treatment reported that students filling in the concept map rather than correcting an incorrect map had better learning outcomes. Wong et al. (2023) further investigated the role of student interest and showed that higher student interest in the topic of electrochemistry also significantly impacted how well a student performed on the concept map activity and correlated with higher posttest scores as well.

One study used a different type of assessment and looked at complexity factors in stoichiometry problems on formative assessments and assessed what problem factors affected problem-solving success (Tang et al., 2014). The stoichiometry problems were randomly generated and differed in variables such as number format, units given, identity of an element, chemical equation, and substance. This study found that only the complexity for the three variables of number format, units given, and chemical equation significantly affected students' academic performance on problem solving (Tang et al., 2014). The stoichiometry problems were also assessed for their cognitive load based on the complexity of the problem and it was found that problems with higher cognitive loads for students resulted in lower student performance. Additionally, Tang et al. (2014) implemented the use of an eye tracking system which revealed that less successful students spent more time focusing on the problem statement and complex variables than students who performed well on the stoichiometry assessment.

Gulacar et al. (2019) created an assessment focusing on word associations to understand the role of students' knowledge structures when given a chemistry related stimulus word. The knowledge structures generated by students with a higher prior knowledge demonstrated more connections and cohesive structures. This study also noted that while mathematics knowledge was important for success in the chemistry course, mathematical background was not a significant influence on the students' chemistry knowledge structures (Gulacar et al., 2019). Students that had more interconnected knowledge structures with concepts such as energy or forces at the center were also noted to be students with high scores on the chemistry placement exam. The use of post-exam retrieval quizzes was another assessment technique used to understand their effect on student performance over the semester (Todd et al., 2021). This study hypothesized that "individuals who participate more in the retrieval practice quizzing will score higher on the cumulative final exam than individuals who elected not to participate" (Todd et al., 2021). The results from the retrieval quizzes revealed that students completing more than 50% of the quizzes performed significantly better on the cumulative final exam than students that completed less than 50% of the retrieval quizzes. There were no significant differences found between the during-term exam grades, suggesting that the use of retrieval quizzes was an effective assessment rather than reflecting general academic ability (Todd et al., 2021). Regardless of a students' achievement level, the retrieval quizzes provided evidence of a forward testing effect where students were able to retain information over time for the cumulative exam. The assessments observed in this study indicate that linking multiple concepts and using metacognitive strategies to reflect on problem-solving processes can increase student learning outcomes.

Online Learning Environment. In addition to courses designed around active learning environments, online learning was also observed to be an important course design factor that promoted students' academic success. There was only one study (Msonde & Van Aalst, 2017) that used an online learning environment as their research study context. This study investigated the differences between non-interactive learning, medium interactive learning, and high interactive learning in a virtual classroom taking place at a Tanzanian university. Main differences in the learning interactions were centered around the use of discussion boards and forums along with interactive assignments such as listening to scientific podcasts and reflecting on chemical connections through the online learning management system. The high interactive learning model had the most substantial gains in academic performance due to the combined use of discussion forums and podcasts. All designs were seen to improve with academic performance due to student engagement and interaction with the discussion forums (Msonde & Van Aalst, 2017). This study also found that the students using discussion boards and podcasts exhibited improved learning gains, most notably with higher order thinking skills related to analysis, synthesis, and evaluation of course content. Similarly to the active learning environments, improved learning outcomes were credited to the student-centered approaches such as discussion forums utilized in the online learning environment.

Theme 2: Course Resources and Feedback

The second emergent theme identified from the examination of research articles was instructional tools and resources. This theme included five studies (14%) that examined specific tools or resources that instructors implement in their introductory chemistry classrooms to boost student academic performance. Instructional tools and resources were further categorized into instructional resources and instructor feedback.

Instructional Resources. The second category of instructional resources was identified when studies discussed resources that were implemented into the classroom or made available outside of the classroom for students to engage with in their introductory chemistry course. Some of the resources included technology, study materials, and homework platforms. One such resource was the use of technologies such as a tablet PC, lecture capture software, and online homework in a chemistry course (Revell, 2014). The tablet PC was used during lecture presentations so the instructor could annotate slides in real-time, lecture capture software allowed students to rewatch the lectures at home, and the online homework platform of Sapling Learning was used as opposed to textbook problems with handwritten answers (Revell, 2014). This study did not find that the use of lecture replays was significantly correlated with higher grades in the course. Rather, it seemed that international, English as a Second Language (ESL), and students with already weak academic backgrounds used the lecture replays to help them to complete the course (Revell, 2014). The online homework through Sapling Learning was significantly correlated with academic performance, with students completing most homework assignments achieving higher exam and course grades. Additionally, Revell (2014) found positive student perceptions with the tablet PC being rated highest for enhancing student learning and instructor effectiveness. The online homework was also valued for its learning gains to students due to the instant feedback and multiple attempts allowed from the online platform. The use of all three instructional resources led to a significant improvement in student retention compared to previous semesters, with a 90% completion rate for the semester using the three technologies compared to an average of 71% in prior semesters (Revell, 2014).

Another instructional resource explored was the use of study materials such as lecture notes and prior assessments. Bunce et al. (2017) investigated the use of study resources in an introductory chemistry course at the U.S. Naval Academy, where students have many time constraints outside of the class due to academic and institutional obligations. Study resources utilized by students were observed to vary depending on the type of assessment. For example, students often used their lecture notes to study for instructor-written assessments while prior assessments or review guides were used to study for common exams (Bunce et al., 2017). This study demonstrated that study resources and the study behaviors of students are important to understand so that instructors can better support their learning processes.

The last instructional resource that was discussed included the study of homework completion versus student academic performance in introductory chemistry (Cosio & Williamson, 2019). In this study, students were assessed on their reasoning abilities through the Test of Logical Thinking (TOLT) and short-term learning gains through in-class clicker questions. In general, the students that completed their homework before the next lecture scored higher on exams compared to students that waited more than four days after lecture to complete their homework. Cosio and Williamson (2019) also found that students with low reasoning abilities based on their TOLT scores did not have a significant relationship to their overall exam score based on when they completed their homework. The completion of the homework also had a more significant correlation to long-term exam performance than the short-term measure of clicker questions (Cosio & Williamson, 2019). Research studies in this SLR demonstrate that a variety of instructional resources and how students engage with them affect the academic success of introductory chemistry students.

Instructor Feedback. Instructor feedback was found to help improve students' academic success in introductory chemistry is instructor feedback. One study analyzed instructor feedback through personalized emails that instructors sent to students which included evaluations on their course performance and advice on support systems and resources (Kyne et al., 2023). This study states that "affirmation from the feedback emails students received strengthened their belief in their own capabilities" (Kyne et al., 2023, p. 979). As a result, students that received personalized feedback emails had higher academic performance as compared to semesters when this instructor feedback was not provided. The mean course grades of students receiving the emails had a statistically significant improvement from 59.2% to 63.5% (Kyne et al., 2023). Additionally, this study showed that 85.3% of students were classified in a "good" grades category and 6.8% in a "poor" grades category with the addition of personalized feedback emails as compared to prior semesters without feedback at 78.5% and 17.9%, respectively (Kyne et al., 2023). Another form of instructor feedback was provided in Carpenter et al.'s (2020) study of exam wrapper feedback provided through an online learning management system. The online exam wrapper was designed to mimic one-on-one feedback sessions and asked reflective questions about students' exam preparation, study strategies, and areas of difficulty. Carpenter et al. (2020) noted that student completion rates of the optional exam wrappers were low, but for the students that did complete them, there was significant correlation between the use of the exam wrappers and course grades. A metacognitive awareness inventory was also distributed to students and findings show that these scores were positively correlated with students' performance in the course (Carpenter et al., 2020). Instructor feedback is an important instructional tool that can cause students to reflect on their performance in the course and take corrective actions towards improved academic success.

Theme 3: Student Learning and Characteristics

There were 12 studies (34%) for the theme student learning and characteristics. These studies examined how student-based factors such as their learning approaches and characteristics were important factors for undergraduate academic success in introductory chemistry. Measures of student affective characteristics such as their motivations, identity, and values along with the way students approach learning are significant indicators of how students perform in chemistry courses.

Student Characteristics. Eight studies investigated student characteristics to further understand how they affect academic performance in introductory chemistry courses. One study from French et al. (2023) explored the motivation of undergraduate chemistry students through the application of expectancy-value theory to understand links to their academic performance. Findings revealed that "students who dropped the course had significantly lower initial confidence about performance" compared to students who completed the course (French et al., 2023, p. 306). Final exam scores were also significantly and positively predicted by confidence in a students' performance, interest, utility values, and attainment values. Another study from Perez et al. (2023) utilized expectancy-value theory to understand patterns and outcomes of introductory chemistry undergraduates. Similarly to French et al.'s (2023) study, students with high confidence in their chemistry abilities and moderate utility and attainment values had higher exam scores. It was also stated that "students with the most adaptive profile of beliefs, according to theory, also had the most success" (Perez et al., 2023, p. 78). The majority of the students in the study held a strong belief in their ability to succeed in science, perceived only moderate costs to participate in science, and overall valued science. Thus, students with this motivation profile were not only seen to have higher exam scores, but also long-term persistence in STEM coursework (Perez et al., 2023).

The study by Edwards et al. (2023) examined the sense of belonging and persistence of students in the second sequence of general chemistry courses. The research examined two dimensions of social belonging where the first dimension referred to students' feelings of connection to peers, instructors, and the course environment. The other dimension referred to belonging uncertainty, which reflects the students' doubts about whether they truly belong in the course. It is worth noting that this study took place in the COVID-19 pandemic, so both sequences of general chemistry were taught online. The researchers used survey data and performance metrics to analyze the relationship between students' social belonging in the first semester sequence of general chemistry and how it affects their decision to persist to the second semester of general chemistry.

The findings reveal that course performance alone did not entirely explain students' persistence to the second semester of general chemistry. While many students with strong grades continued, a notable portion of high-achieving students (including some who received A's) did not progress to the second semester course (Edwards et al., 2023). Furthermore, the research found that first-semester performance did not predict students' early second-semester course sense of belonging, suggesting that external factors, beyond academic success, influence students' perceptions of belonging in chemistry courses. The findings emphasize that inclusive teaching practices, such as fostering social connection, and providing affirmation of student capabilities should be implemented throughout both semesters of general chemistry to improve persistence.

The research article by Fink et al. (2020) investigates the role of students' sense of belonging in predicting academic success and retention in a two-semester general chemistry sequence. The study was conducted at a private, research-intensive university and involved first-year students enrolled in a two-semester sequence of general chemistry courses. The researchers collected data on students' demographic backgrounds, academic preparation (including math scores, chemistry content knowledge, and AP coursework), participation in Peer-Led Team Learning (PLTL), and measures of perceived belonging and belonging uncertainty, which were surveyed at the beginning and end of the semester. The study's methods involved using statistical analyses like ANCOVA, regression, and logistic regression to examine relationships between belonging, demographics, academic preparation, performance, and attrition.

The study found that academic preparation, including prior chemistry knowledge and AP coursework, positively predicted belonging, while students with weaker academic backgrounds felt lower belonging and higher uncertainty (Fink et al., 2020). Additionally, students' early-semester belonging significantly predicted exam performance in both semesters, even after controlling for preparation, demographics, and PLTL participation. Higher belonging was associated with better exam scores, while higher belonging uncertainty was linked to lower performance in the second semester. Moreover, late-semester belonging in the first semester of general chemistry was a significant predictor of attrition, with students who felt lower belonging at the end of the semester more likely to leave the chemistry sequence before the second semester. The study concluded that a strong sense of belonging is an important factor for academic success and persistence in general chemistry.

Other student affective characteristics such as intellectual accessibility, emotional satisfaction, math self-concept, chemistry self-concept, self-efficacy, and test anxiety have been used to identify at-risk students in introductory chemistry courses (Chan & Bauer, 2014). Students in this study were required to take validated instruments such as the Chemistry Self-Concept Inventory (CSCI), Attitude toward the Subject of Chemistry Inventory (ASCI), and Motivated Strategies for Learning Questionnaire (MSLQ) to understand their cognitive and affective characteristics. Students with high scores on the survey instruments exhibited stronger beliefs in their performance of chemistry, more interest in science, and a better self-concept. In turn, students with higher scores on the survey instruments were significantly correlated to higher exam grades in their introductory chemistry course (Chan & Bauer, 2014). Students considered at risk had low beliefs of self-efficacy and self-concept and lower exam grades in the course. Social cognitive changes and student affective characteristics such as STEM interest and self-efficacy were also studied by Hardin and Longhurst (2016). Their research demonstrated that "lower self-efficacy, outcome expectations, and/or supports, as well as higher barriers predict lower interest and persistence in STEM" (Hardin & Longhurst, 2016, p. 234). Students with higher course grades demonstrated a stronger belief in their ability to perform in their introductory chemistry course. Student attitude was another characteristic investigated with student academic achievement in chemistry courses (Brown et al., 2015). The ASCI was given to students to quantify their attitudes and analyze performance on student assessments such as practical, tutorial, online web-based learning, and final exam. Overall, weak positive correlations were found between students' attitudes toward chemistry and their final exam performance (Brown et al., 2015). The final study aiming to understand student characteristics and academic performance in undergraduate chemistry was performed by Van Duser et al. (2021). In this study, they looked at background characteristics of students rather than affective characteristics. Background characteristics of interest included a students' prior ACT scores, SAT scores, and high school GPA. Both high school GPA and ACT math scores were found to be significant predictors of a student's performance in introductory chemistry. This study also noted that there was a difference in the four instructors teaching chemistry, where students with instructor #1 were 2.6 times more likely to pass the course than students taking instructor #4 (Van Duser et al., 2021). Overall, students' cognitive, affective, and background characteristics all play an important role in their academic achievement. Affective characteristics such as identity, self-efficacy, and motivation are more significant factors than a student's attitude towards chemistry.

Student Learning. The way students approach learning in their introductory chemistry course was also revealed to be an important factor for academic success. Four studies investigated factors of student learning such as question logs, metacognitive processes, and learning styles. Bergey et al. (2023) examined student generated questions during chemistry lectures to understand their relationship to exam performance. Students wrote their questions in a log, and it was found that a higher number of questions generated during lectures correlated with lower levels of perceived comprehension. Students that found their questions were resolved during the lecture period reported higher levels of perceived comprehension. On average, students generated one to two questions per lecture, with a median of nine questions every lecture (Bergey et al., 2023). The type of questions that students produced also correlated to exam performance. Questions that sought verification or clarification were classified as closed syntax questions and resulted in better exam performance, but other question types and exam performance were not statistically significant in correlation (Bergey et al., 2023). Additionally, the study found that students improved in their metacognitive accuracy with the use of question logs for the first exam, but this did not continue for the remainder of the semester.

Another study of student learning sought to understand how students monitor and predict their exam performance over time. Hawker et al. (2016) asked students to report the grade they believed they would receive as a question at the end of their exams. This study saw a distinct difference in exam evaluations where students that performed highly on their exam had more accurate evaluations of their exam performance than those who scored poorly. A large majority of students (89%) were observed to estimate their exams at a higher grade than they performed (Hawker et al., 2016). Students were also observed to improve their metacognitive monitoring through exam evaluations immediately after the first exam but did not continue to improve significantly over time. Cracolice and Busby (2015) also examined student learning and its relation to academic success through factors such as alternate conceptions, intelligence, scientific reasoning ability, and attitude towards chemistry. Prior knowledge and the way that students conceptualized chemistry was found to be a significant predictor for performance on ACS exams. Students' scientific reasoning abilities were also significant predictors of exam performance (Cracolice & Busby, 2015). This study revealed that "both alternate conceptions about topics typically covered in first-semester general chemistry and scientific reasoning ability… influence general chemistry content knowledge after a semester of instruction" (Cracolice & Busby, 2015, p. 1793). Alternate conceptions can be resistant to course content and negatively affect exam scores if students do not correct misconceptions.

Lastly, student's learning approaches at a Hispanic serving institution were analyzed to understand how they are related to chemistry course achievement. Final and ACS exam scores were used along with the Revised Approaches to Studying Inventory (RASI) to measure learning approaches as surface, strategic or deep (An et al., 2022). This study indicated that students increased in surface learning approaches and decreased in strategic and deep learning approaches over the semester, which suggests a shift towards rote memorization. Students with high strategic and deep learning approaches were observed to have the highest average ACS exam scores and course grades while those with high surface learning approaches had the lowest scores (An et al., 2022). This study demonstrated that the learning styles of students are directly linked to their academic performance in introductory chemistry. Studies included in this SLR highlighted that factors such as the way students approach learning and reflect on their learning processes play a significant role in their overall understanding of chemistry content.

3.2. Research Question 2: Student Success and Demographic Backgrounds

Research question two aimed to understand any differences in academic success based on demographic backgrounds of students in undergraduate introductory chemistry courses. There were 11 research studies that reported observed differences in academic success based on student demographics such as gender, race, and ethnicity.

Eight studies highlighted that academic outcomes were significantly different for students from historically underrepresented racial and ethnic populations. Bancroft et al. (2020) investigated student demographic differences based on a flipped instructional model intervention. They found that this model significantly improved course grades for Black and Latinx students as they were more likely to achieve higher grades in the flipped instructional model than in traditional lectures. This helped to close the achievement gap between Black and Latinx students compared to White or Asian students receiving the same flipped instruction. Interestingly, the flipped model introduced another performance gap between students from different socioeconomic backgrounds. Students from a low SES background showed much less improvement than students from mid to high SES backgrounds (Bancroft et al., 2020). Another study investigating active learning environments also reported differences for students in historically underrepresented populations. Ott et al. (2018) sparingly discussed student demographic differences based on the implementation of a POGIL model in the introductory chemistry classroom. They found that historically underrepresented students in active learning environments using POGIL strategies were positively impacted, and overall achievement gaps were reduced (Ott et al., 2018). Specifics about the students from historically underrepresented groups were not provided in the study. Additionally, Harri et al. (2020) found that after controlling for academic experiences, students from underserved groups we more likely to persist in the course than their peers if they received a “C” or better grade, which was coined the “hyperpersistent zone”.

Revell (2014) also spoke briefly about the differences in student demographics based on the use of instructional technologies. It was found that students from international or ESL backgrounds used lecture replays more often, but the use of lecture replays was not significantly correlated to final exam performance in the course. Another study taking place at a Hispanic serving institution looked at the differences between Hispanic and non-Hispanic students with regard to their learning approaches and course performance (An et al., 2022). This study found that there were no differences in the learning approaches between Hispanic and non-Hispanic students. However, An et al. (2022) did find a small but statistically significant difference in the ACS exam scores of students that use strategic and deep learning approaches that favored non-Hispanic students when compared to Hispanic students.

Student affective characteristics based on expectancy-value theory was also noted to differ for students from different demographics. Perez et al. (2023) found that women, first-generation college students, and traditionally underrepresented racial and ethnic groups were overrepresented in profiles that had mixed values and costs and moderate to high confidence. This study observed that students from underrepresented populations tend to have higher costs, lower confidence, and less value associated with science. Women and traditionally underrepresented students also showed lower exam performance but no significant difference in long-term science persistence (Perez et al., 2023). Student demographic differences such as race and ethnicity were also investigated by French et al. (2023) in their study of understanding student motivations on their academic performance. This study concluded that there was “no evidence that students from underrepresented racial/ethnic groups reported lower initial chemistry motivation relative to students from well-represented racial/ethnic groups” (French et al., 2023, p. 306). A small and statistically significant finding demonstrated that students from underrepresented racial and ethnic groups had a lower final exam score than students from more well-represented groups. French et al. (2023) also found that attainment value was a stronger and more positive predictor of final exam scores for underrepresented racial and ethnic minorities compared to White and Asian American students. However, there was no evidence of confidence about performance, interest, utility value or attainment value differences for underrepresented and well-represented groups at the beginning of the semester.

The findings by Clark (2023) indicate that reformed courses significantly reduced achievement gaps for historically underrepresented students, even when in-class active learning was removed due to emergency remote teaching during the COVID-19 pandemic. However, the study found that gender-based performance gaps persisted, with male students consistently outperforming female students on exams, while female students performed better in laboratory assessments. The achievement gap based on gender does reverse in lab where females are performing better than males. Achievement gaps based on demographics remain fairly constant throughout the semester, but the gap is smaller and not always significant for gender differences on the final exam (Clark, 2023). The persistence of gender gaps suggests that additional interventions are needed to support female students in exam-based assessments.

The study by Tashiro and Talanquer (2021) examines the disparities in student performance based on sex and underrepresented populations in general chemistry courses. The study finds that reformed courses provide benefits for students with lower academic preparation, particularly in off-sequence courses (spring semester enrollment). Notably, female students outperformed male students in off-sequence courses, whereas in the on-sequence (fall) traditional course, male students had higher grades. Both the traditional and reformed courses showed that there were achievement gaps that favored males over females, but this gap was significantly reduced in the reformed course (Tashiro & Talanquer, 2021). The on-sequence reformed course demonstrated more equitable outcomes between sexes, suggesting that course structure and grading weight distribution influence academic disparities. Despite these improvements, racial and ethnic inequities persisted. The study observed that traditional courses placed heavier emphasis on exams, where male and white or Asian students performed better, while the reformed course incorporated more coursework-based assessments, benefiting female and historically underrepresented students.

In addition to differences in race and ethnicity, nine studies observed different academic outcomes based on gender. French et al. (2023) investigated gender differences in motivation and academic performance for students in introductory chemistry. Gender differences revealed that men reported greater confidence in their chemistry performance compared to women, but women reported greater utility and attainment values than men. The study also found that men performed slightly better than women on final exams (French et al., 2023). Another study by Talanquer and Pollard (2017) helped to narrow performance gaps for students from different gender groups based on their implementation of a reformed curriculum. Female students were shown to have a significant increase in their ACS exam performance, thus reducing the achievement gap for gender (Talanquer & Pollard, 2017). They also noted that students from underrepresented racial and ethnic groups benefitted from a significant improvement in their grade averages and reduced failure rates.

Gulacar et al. (2019) observed gender differences in their study of chemistry students' knowledge structures. Findings revealed that female students had more densely connected and cohesive knowledge structures than those of male students. Gender differences in the chemistry knowledge structures suggest that females and males have a different way of organizing and conceptualizing chemistry knowledge. Another study by Hardin and Longhurst (2016) sought to understand the gender differences in students' self-efficacy and interest and how this ultimately affected course outcomes in undergraduate chemistry. Their study found that "women demonstrated significantly lower STEM and coping self-efficacy and less STEM interest than did men" and men also "experienced a significant increase, on average, in perceived support for obtaining a STEM degree" (Hardin & Longhurst, 2016, p. 237). The gender gap in academic performance did not narrow over the semester in the chemistry courses studied by Hardin and Longhurst (2016). The final research study investigating gender differences reported variances in student exam evaluations. It was observed that female students had higher rates of accurate evaluations of their exam performance compared to male students (Hawker et al., 2016). Females were found to have statistically significant and higher accuracy averages on three out of five semester exams compared to males. However, there was no significant difference in males and females with regard to their overall exam grades (Hawker et al., 2016).

The study by Edwards et al. (2023) examined the sense of belonging and persistence of students in the second sequence of general chemistry courses along with disparities based on gender differences. The study found that belonging uncertainty was a significant predictor of persistence for some female students, meaning that even when their grades were comparable to male students, higher uncertainty about their belonging led some women to discontinue chemistry studies. Additionally, at the beginning of the second semester course, female students reported higher belonging uncertainty than their male peers, despite no significant gender differences in final grades from the first semester course (Edwards et al., 2023). The study concludes that fostering a sense of belonging and reducing stereotype threat is needed to help retain female students in general chemistry. Similarly to work by Edwards et al. (2023), another study by Fink et al. (2020) found that students' early sense of belonging in general chemistry, particularly in the first semester, varied by gender and race. Female students, especially those from historically underrepresented groups, reported lower belonging and greater belonging uncertainty compared to male students. Belonging is especially crucial for students from underrepresented backgrounds and women, who reported lower belonging and higher uncertainty, potentially contributing to disparities in performance and retention.

Only one study examined differences in student demographics related to financial need. In their study about student background characteristics and chemistry academic performance, Van Duser et al. (2021) looked at the receipt of Pell Grants to determine its relationship to chemistry achievement. This study found that the receipt of Pell Grants was not a statistically significant predictor of chemistry academic performance. Other studies lacked information related to students' socioeconomic status and their academic success.

4. Discussion

This SLR aimed to synthesize existing literature of undergraduate introductory chemistry courses to identify factors of student academic success (York et al., 2015) and any differences across varying groups of students. Our findings revealed that a large majority of studies used quantitative methodology. Of the mixed methods studies, only a few utilized focus groups or interviews to have direct interaction with chemistry undergraduate students. None of the research studies only used qualitative methodology to understand academic success in introductory chemistry undergraduates. Qualitative methods such as interviews and focus groups will utilize students' voices to fully understand their lived experiences and can provide richer details and context than quantitative methodology (Libarkin & Kurdziel, 2002). This SLR found a need for more qualitative research and case studies to understand how students make sense of their experiences in introductory chemistry courses.

Some of the research studies examined student performance in subsequent courses or credits taken in upperclassmen years. Many of the articles did not use a longitudinal study to understand academic performance and student outcomes over the course of their entire academic career. Studies presented in this SLR largely discussed academic success over the course of one semester, yet further insights can be gained from examining students over the course of several years. Institutions were largely U.S. based public universities with only international universities from Tanzania ($n = 1$) and Australia ($n = 2$). There is a need to investigate introductory chemistry courses at more international universities to gain insights into how institutions abroad promote academic success for their students. Additionally, many of the universities were public four-year institutions, with only one noted to be a minority serving institution (An et al., 2022) and another a U.S. Naval Academy (Bunce et al., 2017). This SLR shows a deficit in examination of institution types on student academic success in undergraduate chemistry. Historically Black Colleges and Universities (HBCU) serve an important demographic of historically underrepresented students and should be included in investigations of undergraduate chemistry courses. Moreover, none of the studies included private institutions such as Ivy League universities, which could arguably have different expectations and factors of academic success than the studies included in this study.

Community colleges are also important academic institutions that are often under-recognized. Community colleges have been a growing entry point for students wanting to major in STEM (Snyder & Cudney, 2017), yet attrition in STEM is often highest at community colleges (Chen, 2013). Only one study (Gilewski et al., 2022) in this SLR used both a community college and public four-year university to investigate academic success for introductory chemistry undergraduates, but differences between the two institutions were not addressed. Community colleges cater to a diverse body of students, including a significant proportion of first-generation college students, low-income students, and students from underrepresented backgrounds (Cohen & Kelly, 2019). Despite their role in the STEM education pipeline, community colleges often operate with fewer resources than their four-year counterparts, yet they are tasked with the responsibility of preparing nearly half of the students who graduate with bachelor's and master's degrees in STEM fields (Hagedorn & Purnamasari, 2012). Considering that many community college students who come from low SES backgrounds are first-generation college students and maintain jobs while studying, it is unclear how academic interventions such as flipped instructional models may affect community college students. Future research efforts should be aimed at students in introductory chemistry courses at community colleges to understand how academic success may be different when compared to students at public four-year universities.

4.1. Factors of Academic Success in Undergraduate Introductory Chemistry

This SLR found three emergent themes for student academic success (York et al., 2015) in introductory chemistry courses: course design, instructional tools and resources, and student learning and characteristics. The theme of course design and learning environments saw many active learning environments utilized to promote students' academic success. Students in active learning environments have previously demonstrated the ability to retain more information and perform better on assessments than students in traditional learning environments (Bressoud, 2020; Clark, 2023; Freeman et al., 2014; Tashiro & Talanquer, 2021). Flipped instruction emerged as one of the active learning environments that helps promote student academic success in introductory chemistry courses. The flipped or partially flipped approach helps to reduce the cognitive load and working memory for students by allowing for them to take notes on prerecorded lectures outside of the classroom. Working memory is "affected by the inherent nature of material and by the manner in which the material is presented" (Kirschner, 2002, p. 4). The nature of chemistry requires students to understand submicroscopic concepts and connect them to the macroscopic world students are familiar with. Chemistry content comes with a plethora of new vocabulary, symbols, reactions, equations, and relationships for students to grasp. As a result, chemistry can cause substantial stresses to working memory as students try to organize and conceptualize information (Schuttlefield et al., 2012). Cognitive load theory assumes that students have a limited working memory and can only process a certain amount of information before being overloaded (Kirschner, 2002). This SLR highlighted instructional models focused on student-centered approaches such as flipped classrooms and POGIL models to help reduce strain for cognitive loads and working memory. Understanding the stressors that chemistry imparts on working memory and cognitive loads can lead chemistry instructors to design their courses in ways that facilitate more effective learning.

Only one of the studies (Msonde & Van Aalst, 2017) discussed online learning environments. Online learning has become popular in recent years with higher enrollment of online introductory courses, yet these courses have high rates of attrition compared to their in-person counterparts (Laing & Laing, 2015; Lederman, 2021). The area of research concerning online learning environments for introductory chemistry is underrepresented in this SLR to fully understand how best to promote academic success for students in virtual introductory chemistry courses. In recent years, and particularly during the COVID-19 pandemic, online instruction has become a popular mode of instruction for students. Online learning can limit student interactions and isolate students from on-campus support systems (Laing & Laing, 2015; Lederman, 2021). Msonde and Van Aalst (2017) used student-centered approaches to help increase engagement and interactions among students in a virtual environment. Further research should look at effective online learning in chemistry courses to understand how these virtual environments can promote better educational outcomes for chemistry students given the traditionally higher attrition rates of online courses.

Many studies within the second emergent theme of course resources and feedback focused on assessments to improve student academic outcomes. One study (Tang et al., 2014) demonstrated how the complexity of stoichiometry problems are connected to a student's performance on the assessment. This study is also rooted in cognitive load theory as working memory is influenced by the manner in which chemical content is presented to students (Kirschner, 2002). Other studies highlighted the importance of student reflection and opportunities to engage with metacognitive processes through assessments. Metacognition, or thinking about thinking, can help students regulate their learning and develop a more profound understanding of chemical concepts (Lavi et al., 2019). This study revealed that assessments utilizing metacognitive prompts increased academic success for chemistry undergraduates. Additionally, it is important for students to link chemical concepts across topics to gain a deeper understanding of chemistry as a whole. Instructor feedback was observed to not only improve student grades but also promote student self-efficacy by affirming their abilities. Incorporating personalized instructor feedback on student assessments that utilize reflective, metacognitive, or crosslinking prompts can encourage students to engage more deeply with the curriculum rather than rely on rote memorization.

Other instructional tools and resources that were observed to promote academic success for students included the effective use of homework and quizzes. The more practice and engagement students have with course materials, the better their academic outcomes. This SLR did not uncover the use of innovative or alternative approaches such as technological tools and written assignments. Traditionally, chemistry courses do not provide students the opportunity to explore their own interests and engage deeply with course content (Bressoud, 2020; Bokosmaty et al., 2019). Alternative assessments, such as essays or papers, can be utilized to add more agency to student learning and allow students to investigate topics they are interested in. Student essays have been previously used in chemistry courses for students to write about their own passions and interests as it relates to the chemistry course (Asher et al., 2023). These student essays were seen to increase the persistence of students in STEM degree programs. Further research into instructional tools and resources that can foster agency among students is needed to understand any benefits to academic success and persistence in chemistry.

The final emergent theme revealed that student learning and characteristics were an important factor for academic success in introductory chemistry. Science identities and motivations as understood by expectancy-value theory were most often investigated to comprehend how student affective characteristics affect their course performance. Students that have high confidence in their abilities to succeed, value chemistry and science, and have low costs to participate in science, are those most likely to academically succeed (Brown et al., 2015; French et al., 2023; Perez et al., 2023). Specific academic interventions to help foster students' confidence or improve their value perception were not presented. This

SLR uncovered a need for further investigation of student identities using other theoretical frameworks aside from expectancy-value theory. Additional insights of students' identity and their affective characteristics can be gleaned from the use of theories such as identity-based motivation (Oyserman et al., 2017), self-determination theory (Black & Deci, 2000), or Bandura's social cognitive theory (Gryka et al., 2017).

One study (Smith et al., 2018) discussed the use of health-related scenarios in a general chemistry course for predominately nursing majors but did not look at student affective characteristics as a result of the more identity-congruent instruction. None of the studies presented in this SLR discussed student agency and the importance of students exerting choice and control over their learning. Students often take a passive role in STEM courses and do not have many opportunities for agency. It has been previously shown that adding more agency to student learning can help foster a sense of engagement and belonging in STEM through the use of innovative teaching methods (Ryoo & Winkelmann, 2021). Linking course content back to student's interests or future goals is especially needed considering the majority of students in introductory chemistry are not chemistry or biochemistry majors. Any efforts to connect course content to students' knowledge, experiences, and future goals could have profound impacts. Future research efforts can look at promoting student choice and design interventions to understand how they affect student learning and characteristics along with promoting academic performance.

It was also noted that studies did not specifically examine professors and how their professional development or pedagogical perspectives can help promote academic success of students. One study (Van Duser et al., 2021) observed differences in student academic performance and passing rates based on the instructor teaching the course, but did not provide details why this was the case. Van Duser et al. (2021) only sought to understand differences in student background characteristics and later uncovered the instructor differences in their analysis. Future research efforts should be aimed at understanding how professors come to develop and enact their personal pedagogies. Class observations were not a method identified in the 35 research articles in this SLR and would be beneficial for exploring professors' pedagogical practices in chemistry courses. There is an expansive literature base on educators' pedagogical content knowledge (PCK) and how it affects students' performance in primary and secondary education but seldom reciprocated in STEM higher education contexts (Fraser, 2016; Mahler et al., 2017; Park & Oliver, 2008). Chemistry instructors' PCK and student academic performance or persistence in STEM degree programs is an understudied area in STEM higher education.

4.2. Differing Factors for Varying Student Groups

Many studies analyzing the differences in student groups and academic performance focused on differences in student demographics such as gender, race, and ethnicity. Results from these studies indicated students from historically underrepresented populations such as Black, Hispanic, or female students, were at a higher disadvantage when compared to well-represented populations in STEM such as White or Asian males. Introductory chemistry courses often expose students to many American or European White males as many of the elements, compounds, structures, theories, equations, and chemical reactions are named after them. Students are aware that many of the individuals deemed important or significant in STEM do not come from a historically underrepresented background (Dancy et al., 2020). Current occupations in STEM also lack diversity, especially in engineering and computer science fields (National Center for Science & Engineering Statistics, 2023). When chemistry instructors are presenting important historical scientists in their course, they can also take the time to present scientists from underrepresented racial, ethnic, and gender groups. Research has shown that students personally identifying with a scientist or role model in STEM are more likely to have interest in STEM and have a greater sense of belonging (Edwards et al., 2023; Fink et al., 2020; Gladstone & Cimpian, 2021). It is important for chemistry courses to teach about scientists that are reflective of the students taking the course in order to promote more identity congruence and a stronger sense of belonging for students. Many studies in this SLR only reported differences in student demographics and academic performance and did not discuss specific interventions that could be used to decrease achievement gaps and provide more equitable outcomes. There is a need for future studies to examine how students from historically underrepresented populations can be supported in introductory chemistry courses to overcome barriers of entry into STEM degree programs.

The Bancroft et al. (2020) study demonstrated that a flipped instructional model does not improve the academic success of all students in introductory chemistry courses. While the achievement gap closed for students from traditionally underserved groups such as Black and Latinx students, another performance gap was introduced for students from varying SES backgrounds. Flipped instruction requires students to use technology on their own time outside of the classroom to watch and take notes on prerecorded lectures. Students from low SES backgrounds may have more time restrictions than other students due to their own work responsibilities or family obligations based on their financial status. Studies in this SLR seldom discussed first-generation or low SES students. These backgrounds should be further investigated to understand student experiences of introductory chemistry courses with the goal of creating targeted supports. Academic resources should be leveraged to help students most at risk of dropping or failing their first semester in undergraduate gateway courses like chemistry.

Students enrolled in introductory chemistry courses are frequently majoring in other degree programs than chemistry- or biochemistry-related fields. None of the studies analyzed in this SLR looked at differences in academic performance based on the intended major of students. Examining student outcomes based on their intended major could provide insights into the degree programs that are most at-risk when taking chemistry courses and develop program-based interventions. Additionally, learning outcome differences were not analyzed between different institutions. Future studies can focus efforts at understanding any disparities that exist among different institution types such as public four-year universities, private universities, HBCU's, small liberal arts colleges, or community colleges. While differences in learning outcomes were observed based on achievement level, none of the studies targeted students that dropped out of the course to understand their experiences. Recruiting participants that have dropped, withdrew, or failed an introductory chemistry course would provide a deeper understanding of why they left and what they need to be academically successful. Studies are also needed to understand how students utilize academic resources such as textbook use, tutoring, and student-formed study groups to perform well in chemistry.

Lastly, student differences in academic performance were observed based on student affective characteristics, usually through the lens of expectancy-value theory. Students that had interest in STEM and stronger attitudes of STEM persisted through their chemistry course more often. Additionally, students that had high confidence in their STEM abilities, valued STEM, and exhibited low costs for participating in STEM were also more academically successful (Chan & Bauer, 2014; Perez et al., 2023). Even though student affective characteristics are strong predictors of their performance in introductory chemistry courses, only two studies examined differences in these variables on student learning outcomes. Future studies implementing targeted interventions to understand their effects on student affective characteristics are needed. Changing a students' perception of their identity, self-efficacy, interest, and motivation in STEM can have profound impacts for STEM persistence and workforce development.

5. Limitations

A limitation of this study could be the inclusion of only two databases for this SLR. Our initial search included two additional databases (Google Scholar and Proquest). However, due to significant overlap in articles produced, we decided to focus on only two selected search engines. Potentially expanding this SLR to include more databases may have resulted in more studies, which may provide additional insights into the factors of academic success and student differences observed in undergraduate introductory chemistry courses.

6. Conclusions

Introductory chemistry serves as a gatekeeper course for undergraduate STEM degree programs and plays an important role in the retention and persistence of STEM graduates. This SLR revealed three key factors that contribute to introductory chemistry students' academic success including course design and learning environment, course resources and feedback, and student learning and characteristics. First, it is critical for an introductory chemistry course to move from traditional formats to a more student-centered, active-learning format. This could include using a flipped classroom model where students spend the majority of time in class working in collaborative groups and engaging actively with the content. The instructor should utilize formative assessment practices to gauge student learning during class meetings and provide real-time feedback during group work time. Second, creating access for students to chemistry content through strategies that build identity and motivation—seeing themselves as a successful person in chemistry—is key to engaging all students in the course including first-generation students, as well as those from historically underserved and underrepresented groups in chemistry and STEM overall. Creating access and building community within the introductory chemistry is critical for success—as the literature in this review indicated student demographics such as gender, race, and ethnicity are the most reported factors linked to disparities in academic outcomes in introductory chemistry.

The findings of this SLR highlight the need for targeted interventions to support equitable learning environments for a diverse group of students. Chemistry instructors should consider the use of active learning strategies and curriculum that can connect to students' experiences and future goals. Instructors should also consider the use of varied assessments and opportunities for reflection on metacognitive processes paired with constructive feedback for students. Additionally, chemistry interventions to promote students' lived experiences and science identity should be further developed to promote better educational outcomes for students.

There is a need for future research efforts to be focused on initiatives that address the barriers various groups of students encounter in introductory chemistry coursework. This can include the examination of mentorship programs, academic support services, and interventions to create a supportive learning environment. Through implementing the factors that influence academic success and working to eliminate disparities among groups of students, introductory chemistry courses can provide better educational outcomes for students along with increasing their persistence in STEM.

Author Contributions: Conceptualization, J.C. and C.C.J.; Methodology, J.C.; Validation, J.C.; Formal analysis, J.C. and C.C.J.; Data curation, J.C.; Writing—original draft, J.C. and C.C.J.; Writing—review & editing, J.C. and C.C.J.; Supervision, C.C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: An inventory of articles included in the systematic literature review are available along with the code sheet upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

ACT. (2018). *The condition of STEM 2017*. ACT. Available online: https://www.act.org/content/dam/act/unsecured/documents/cccr2017/CCCR_National_2017.pdf (accessed on 16 February 2024).

An, J., Guzman-Joyce, G., Brooks, A., To, K., Vu, L., & Luxford, C. J. (2022). Cluster analysis of learning approaches and course achievement of general chemistry students at a Hispanic serving institution. *Journal of Chemical Education*, *99*(2), 669–677. [CrossRef]

Asher, M. W., Harackiewicz, J. M., Beymer, P. N., Hecht, C. A., Lamont, L. B., Else-Quest, N. M., Priniski, S. J., Thoman, D. B., Hyde, J. S., & Smith, J. L. (2023). Utility-value intervention promotes persistence and diversity in STEM. *Proceedings of the National Academy of Sciences*, *120*(19), e2300463120. [CrossRef] [PubMed]

Bancroft, S. F., Fowler, S. R., Jalaeian, M., & Patterson, K. (2020). Leveling the field: Flipped instruction as a tool for promoting equity in general chemistry. *Journal of Chemical Education*, *97*(1), 36–47. [CrossRef]

Bergey, B. W., Cromley, J. G., Kaplan, A., & Bloxton, J. D. (2023). Do students' questions during chemistry lectures predict perceived comprehension and exam performance? *The Journal of Experimental Education*, *91*(3), 411–430. [CrossRef]

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, *84*, 740–756. Available online: https://psycnet.apa.org/doi/10.1002/1098-237X(200011)84:6%3C740::AID-SCE4%3E3.0.CO;2-3 (accessed on 16 February 2024). [CrossRef]

Bokosmaty, R., Bridgeman, A., & Muir, M. (2019). Using a partially flipped learning model to teach first year undergraduate chemistry. *Journal of Chemical Education*, *96*(4), 629–639. [CrossRef]

Bressoud, D. (2020). Talking about leaving revisited: Persistence, relocation, and loss in undergraduate STEM education. *Notices of the American Mathematical Society*, *67*(09), 1. [CrossRef]

Brown, S. J., White, S., Sharma, B., Wakeling, L., Naiker, M., Chandra, S., Gopalan, R., & Bilimoria, V. (2015). Attitude to the study of chemistry and its relationship with achievement in an introductory undergraduate course. *Journal of the Scholarship of Teaching and Learning*, *15*, 33–41. [CrossRef]

Bunce, D. M., Komperda, R., Dillner, D. K., Lin, S., Schroeder, M. J., & Hartman, J. R. (2017). Choice of study resources in general chemistry by students who have little time to study. *Journal of Chemical Education*, *94*(1), 11–18. [CrossRef]

Carpenter, T. S., Beall, L. C., & Hodges, L. C. (2020). Using the LMS for exam wrapper feedback to prompt metacognitive awareness in large courses. *Journal of Teaching and Learning with Technology*, *9*(1), 79–91. [CrossRef]

Chambers, K. A., & Blake, B. (2008). The effect of LearnStar on student performance in introductory chemistry. *Journal of Chemical Education*, *85*, 1395−1399.

Chan, J. Y. K., & Bauer, C. F. (2014). Identifying at-risk students in general chemistry via cluster analysis of affective characteristics. *Journal of Chemical Education*, *91*(9), 1417–1425. [CrossRef]

Chen, X. (2013). *STEM attrition: College students' paths into and out of STEM fields* (Statistical Analysis Report. NCES 2014-001). National Center for Education Statistics.

Clark, T. M. (2023). Narrowing achievement gaps in general chemistry courses with and without in-class active learning. *Journal of Chemical Education*, *100*(4), 1494–1504. [CrossRef]

Cohen, R., & Kelly, A. M. (2019). Community college chemistry coursetaking and STEM academic persistence. *Journal of Chemical Education*, *96*(1), 3–11. [CrossRef]

Cosio, M. N., & Williamson, V. M. (2019). Timing of homework completion vs. performance in general chemistry. *Journal of Science Education and Technology*, *28*(5), 523–531. [CrossRef]

Cracolice, M. S., & Busby, B. D. (2015). Preparation for college general chemistry: More than just a matter of content knowledge acquisition. *Journal of Chemical Education*, *92*(11), 1790–1797. [CrossRef]

Dancy, M., Rainey, K., Stearns, E., Mickelson, R., & Moller, S. (2020). Undergraduates' awareness of white and male privilege in STEM. *International Journal of STEM Education*, *7*(1), 52. [CrossRef]

Edwards, J. D., Torres, H. L., & Frey, R. F. (2023). The effect of social belonging on persistence to general chemistry 2. *Journal of Chemical Education*, *100*(11), 4190–4199. [CrossRef]

Fayer, S., Lacey, A., & Watson, A. (2017). *Spotlight on STEM*; US Bureau of Labor Statistics. Available online: https://www.bls.gov/spotlight/2017/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future/ (accessed on 16 February 2024).

Figueroa, T., Cobian, K., Hurtado, S., & Eagan, K. (2017, March 4). *Trends and pathways for STEM major aspirants: A look at national data*. 9th Conference on Understanding Interventions That Broaden Participation in Science Careers Conference, San Antonio, TX, USA.

Fink, A., Frey, R. F., & Solomon, E. D. (2020). Belonging in general chemistry predicts first-year undergraduates' performance and attrition. *Chemistry Education Research and Practice*, *21*(4), 1042–1062. [CrossRef]

Fraser, S. P. (2016). Pedagogical content knowledge (PCK): Exploring its usefulness for science lecturers in higher education. *Research in Science Education*, *46*, 141–161. [CrossRef]

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. [CrossRef]

French, A. M., Else-Quest, N. M., Asher, M., Thoman, D. B., Smith, J. L., Hyde, J. S., & Harackiewicz, J. M. (2023). An intersectional application of expectancy-value theory in an undergraduate chemistry course. *Psychology of Women Quarterly*, *47*(3), 299–319. [CrossRef]

Gilewski, A., Litvak, M., & Ye, L. (2022). Promoting metacognition through measures of linked concepts with learning objectives in introductory chemistry. *Chemistry Education Research and Practice*, *23*(4), 876–884. [CrossRef]

Gillespie, R. J. (1991). What is wrong with the general chemistry course? *Journal of Chemical Education*, *68*, 192−194.

Gladstone, J. R., & Cimpian, A. (2021). Which role models are effective for which students? A systematic review and four recommendations for maximizing the effectiveness of role models in STEM. *International Journal of STEM Education*, *8*, 59. [CrossRef] [PubMed]

Gryka, R., Kiersma, M. E., Frame, T. R., Cailor, S. M., & Chen, A. M. H. (2017). Comparison of student confidence and perceptions of biochemistry concepts using a team-based learning versus traditional lecture-based format. *Pharmacy Teaching and Learning*, *9*(2), 302–310. [CrossRef]

Gulacar, O., Milkey, A., & McLane, S. (2019). Exploring the effect of prior knowledge and gender on undergraduate students' knowledge structures in chemistry. *EURASIA Journal of Mathematics, Science and Technology Education*, *15*(8), em1726. [CrossRef]

Hagedorn, L. S., & Purnamasari, A. V. (2012). A realistic look at STEM and the role of community colleges. *Community College Review*, *40*(2), 145–164.

Hardin, E. E., & Longhurst, M. O. (2016). Understanding the gender gap: Social cognitive changes during an introductory stem course. *Journal of Counseling Psychology*, *63*(2), 233–239. [CrossRef]

Harri, R. B., Mack, M. R., Bryant, J., Theobald, E. J., & Freeman, S. (2020). Reducing achievement gaps in undergraduate general chemistry could lift underrepresented students into a "hyperpersistent zone". *Scientific Advances*, *6*(24), eaaz5687. [CrossRef]

Hartikanean, S., Rintala, H., Pylvas, L., & Nokelanien, P. (2019). The concept of active learning and the measurement of learning outcomes: A review of research in engineering higher education. *Education Sciences*, *9*(4), 276. [CrossRef]

Hawker, M. J., Dysleski, L., & Rickey, D. (2016). Investigating general chemistry students' metacognitive monitoring of their exam performance by measuring postdiction accuracies over time. *Journal of Chemical Education*, *93*(5), 832–840. [CrossRef]

He, W., Holton, A. J., & Farkas, G. (2018). Impact of partially flipped instruction on immediate and subsequent course performance in a large undergraduate chemistry course. *Computers & Education*, *125*, 120–131. [CrossRef]

Kim, H., Chacko, P., Zhao, J., & Montclare, J. K. (2014). Using touch-screen technology, apps, and blogs To engage and sustain high school students' interest in chemistry topics. *Journal of Chemical Education*, *91*(11), 1818–1822. [CrossRef]

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*, 1–10.

Kyne, S. H., Lee, M. M. H., & Reyes, C. T. (2023). Enhancing academic performance and student success through learning analytics-based personalised feedback emails in first-year chemistry. *Chemistry Education Research and Practice*, *24*(3), 971–983. [CrossRef]

Laing, C. L., & Laing, G. K. (2015). A conceptual framework for evaluating attrition in online courses. *e-Journal of Business Education & Scholarship of Teaching*, *9*(2), 39–55.

Lavi, R., Shwartz, G., & Dori, Y. J. (2019). Metacognition in chemistry education: A literature review. *Israel Journal of Chemistry*, *59*(6–7), 583–597. [CrossRef]

Lederman, D. (2021). *Detailing last fall's online enrollment surge*. Inside Higher Ed. Available online: https://www.insidehighered.com/news/2021/09/16/new-data-offer-sense-how-covid-expanded-online-learning (accessed on 16 February 2024).

Libarkin, J. C., & Kurdziel, J. P. (2002). Research methodologies in science education: The qualitative-quantitative debate. *Journal of Geoscience Education*, *50*(1), 78–86. [CrossRef]

Lloyd, P. M., & Eckhardt, R. A. (2010). Strategies for improving retention of community college students in the sciences. *Science Educator*, *19*(1), 33–41.

Mahler, D., Großschedl, J., & Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students' performance. *International Journal of Science Education*, *39*(2), 213–237. [CrossRef]

McDowell, L. D. (2019). The roles of motivation and metacognition in producing self-regulated learners of college physical science: A review of empirical studies. *International Journal of Science Education*, *41*(17), 2524–2541. [CrossRef]

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage Publications, Inc.

Msonde, S. E., & Van Aalst, J. (2017). Designing for interaction, thinking and academic achievement in a Tanzanian undergraduate chemistry course. *Educational Technology Research and Development*, *65*(5), 1389–1413. [CrossRef]

National Center for Science and Engineering Statistics (NCSES). (2023). *Diversity and STEM: Women, minorities, and persons with disabilities 2023* (Special Report NSF 23-315). Available online: https://www.nsf.gov/reports/statistics/diversity-stem-women-minorities-persons-disabilities-2023 (accessed on 16 February 2024).

NC State University Libraries. (2024). *Adult and higher education key databases*. Available online: https://www.lib.ncsu.edu/databases/eric (accessed on 16 February 2024).

Ott, L. E., Carpenter, T. S., Hamilton, D. S., & LaCourse, W. R. (2018). Discovery learning: Development of a unique active learning environment for introductory chemistry. *Journal of the Scholarship of Teaching and Learning*, *18*(4), 161–180. [CrossRef]

Oyserman, D., Lewis, N. A., Yan, V. X., Fisher, O., O'Donnell, S. C., & Horowitz, E. (2017). An identity-based motivation framework for self-regulation. *Psychological Inquiry*, *28*(2–3), 139–147. [CrossRef]

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., & Mulrow, C. D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*(71), n71. [CrossRef]

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, *38*, 261–284. [CrossRef]

Perez, T., Robinson, K. A., Priniski, S. J., Lee, Y., Totonchi., D. A., & Linnenbrink-Garcia, L. (2023). Patterns, predictors, and outcomes of situated expectancy-value profiles in an introductory chemistry course. *Annals of the New York Academy of Sciences*, *1526*, 73–83. [CrossRef]

Philipp, S. B., Tretter, T. R., & Rich, C. V. (2016). Undergraduate teaching assistant impact on student academic achievement. *Journal of Science Education*, *20*(2), 1–13.

Porter, L. A., Chapman, C. A., & Alainiz, J. A. (2017). Simple and inexpensive 3D printed filter fluorometer designs: User-friendly instrument models for laboratory learning and outreach activities. *Journal of Chemical Education*, *94*(1), 105–111. [CrossRef]

President's Council of Advisors on Science and Technology. (2012). *Report to the president, engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*; Executive Office of the President, President's Council of Advisors on Science and Technology. Available online: http://files.eric.ed.gov/fulltext/ED541511.pdf (accessed on 16 February 2024).

Revell, K. D. (2014). A comparison of the usage of tablet PC, lecture capture, and online homework in an introductory chemistry course. *Journal of Chemical Education*, *91*(1), 48–51. [CrossRef]

Ryoo, J., & Winkelmann, K. (Eds.). (2021). *Innovative learning environments in STEM higher education: Opportunities, challenges, and looking forward*. Springer International Publishing. [CrossRef]

Schreier, M. (2019). *Qualitative content analysis* (P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams, Eds.). SAGE Research Methods Foundations. [CrossRef]

Schuttlefield, J. D., Kirk, J., Pienta, N. J., & Tang, H. (2012). Investigating the effect of complexity factors in gas law problems. *Journal of Chemical Education*, *89*(5), 585–591. [CrossRef]

Seymour, E., & Hewitt, N. (2000). *Talking about leaving: Why undergraduates leave the sciences* (pp. 1–448). Westview Press.

Smith, A. L., Paddock, J. R., Vaughan, J. M., & Parkin, D. W. (2018). Promoting nursing students' chemistry success in a collegiate active learning environment: "If I have hope, I will try harder". *Journal of Chemical Education*, *95*(11), 1929–1938. [CrossRef]

Snyder, J., & Cudney, E. (2017). Retention models for STEM majors and alignment to community colleges: A review of the literature. *Journal of STEM Education*, *18*(3), 48–57.

Sorensen-Unruh, C. (2017). ConfChem conference on select 2016 BCCE presentations: Radical awakenings—A new teaching paradigm using social media. *Journal of Chemical Education*, *94*(12), 2002–2004. [CrossRef]

Stone, K., Shaner, S., & Fendrick, C. (2018). Improving the success of first term general chemistry students at a liberal arts institution. *Education Sciences*, *8*(1), 5. [CrossRef]

Stringfield, T. W., & Kramer, E. F. (2014). Benefits of a game-based review module in chemistry courses for nonmajors. *Journal of Chemical Education*, *91*(1), 56–58. [CrossRef]

Tai, R. H., Sadler, P. M., & Loehr, J. F. (2005). Factors influencing success in introductory college chemistry. *Journal of Research in Science Teaching*, *42*, 987–1012.

Talanquer, V. (2017). Concept inventories: Predicting the wrong answer may boost performance. *Journal of Chemical Education*, *94*(12), 1805–1810. [CrossRef]

Talanquer, V., & Pollard, J. (2017). Reforming a large foundational course: Successes and challenges. *Journal of Chemical Education*, *94*(12), 1844–1851. [CrossRef]

Tang, H., Kirk, J., & Pienta, N. J. (2014). Investigating the effect of complexity factors in stoichiometry problems using logistic regression and eye tracking. *Journal of Chemical Education*, *91*(7), 969–975. [CrossRef]

Tashiro, J., & Talanquer, V. (2021). Exploring inequities in a traditional and a reformed general chemistry course. *Journal of Chemical Education*, *98*(12), 3680–3692. [CrossRef]

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *45*, 8. [CrossRef]

Todd, K., Therriault, D. J., & Angerhofer, A. (2021). Improving students' summative knowledge of introductory chemistry through the forward testing effect: Examining the role of retrieval practice quizzing. *Chemistry Education Research and Practice*, *22*(1), 175–181. [CrossRef]

U.S. Bureau of Labor Statistics. (2020). *Employment in STEM occupations*. Available online: https://www.bls.gov/emp/tables/stem-employment.htm (accessed on 16 February 2024).

Van Duser, K. E., Yan, X., Lucas, C. M., & Cohen, S. K. (2021). Predicting and supporting student performance in a high fail and high incompletion course: An exploratory study of introduction to general chemistry. *College Student Journal*, *55*(2), 135–144.

Wang, Z., Adesope, O., Sundararajan, N., & Buckley, P. (2021). Effects of different concept map activities on chemistry learning. *Educational Psychology*, *41*(2), 245–260. [CrossRef]

Whalen, D. F., & Shelley, M. C. (2010). Academic success for STEM and non-STEM majors. *Journal of STEM Education*, *11*(1), 45–60.

Wong, R. M., Alpizar, D., Adesope, O. O., & Nishida, K. R. A. (2023). Role of concept map format and student interest on introductory electrochemistry learning. *School Science and Mathematics*, *124*(1), 18–31. [CrossRef]

Ye, L., Oueini, R., & Lewis, S. E. (2015). Developing and implementing an assessment technique to measure linked concepts. *Journal of Chemical Education*, *92*(11), 1807–1812. [CrossRef]

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research & Evaluation*, *20*(1), 5.

education sciences

Article

Students' & Faculty Members' Attitudes Towards Learning and Teaching Reaction Mechanisms in Organic Chemistry

Oluwatobi Odeleye * and Nghiem Tieu *

Department of Chemistry, West Virginia University, Morgantown, WV 26506, USA
* Correspondence: oluwatobi.odeleye@mail.wvu.edu (O.O.); nut00001@mix.wvu.edu (N.T.)

Abstract: Organic Chemistry has typically been identified as a difficult course for many undergraduate students and has a notoriously high failure rate. The part of the subject dealing with reaction mechanisms is considered the most challenging area, and several papers have been published on how to facilitate students' understanding of mechanisms. During Fall 2022 and Spring 2023, we surveyed 127 students, and interviewed 3 students and 5 faculty members about their opinions towards teaching and learning Organic Chemistry, especially reaction mechanisms. The students' attitudes were surveyed through the Attitude towards the Subject of Chemistry Inventory (ASCIv2), and its relationship with grades was also investigated. The results show that most students have negative attitudes towards Organic Chemistry; however, those with more positive attitudes performed better in the course. Students mostly viewed Organic Chemistry as a course required for their major/degree or professional exams without knowing the actual applications of the subject in their respective fields. Professors were able to relate organic chemistry to other fields besides chemistry (Health Science) but found it difficult to give examples of where else reaction mechanisms would be used outside of Organic Chemistry. A suggestion for a change of Organic Chemistry course is discussed in the conclusion of this study.

Keywords: organic chemistry; reaction mechanisms; perspective

1. Introduction

Organic Chemistry has typically been identified as a difficult course for many undergraduate students and has a notoriously high failure rate (Barr et al., 2010; Dwyer & Childs, 2017; Paulson, 1999; Szu et al., 2011). Many STEM fields, besides Chemistry, require Organic Chemistry as a prerequisite, including Health Sciences such as Medicine, Pharmacy, Dentistry, and Nursing. Since this is a difficult course, it has been regarded as one of the barriers preventing students from accomplishing their goals, as many fail the course and some even lose interest in pursuing what they originally planned (Gupta & Hartwell, 2019; Zhang et al., 2020). The question of exploring why this course is so hard for most students has long been a research interest for Chemical Education community researchers. As studies found, multiple factors affect students' perceptions of the course, such as the content of the course, the approach students take to tackle the course's content, instructional approaches, etc. Collini et al. (2024) explored students' attitudes before and after the course and analyzed the themes that shaped students' attitudes (Collini et al., 2024). They found that students tend to link Organic Chemistry with difficulties, memorization, and the large volume of material students had to work on. "Educators" also appears consistently as a theme across their data, affecting the students both positively and negatively. Vilia et al. studied the relationship between attitudes and reasoning abilities and found a positive correlation between the two (Vilia et al., 2017). Dwyer and Childs (2017) investigated perspectives through evaluations of content difficulties from students and teachers. There were agreements about what topics were easy or hard, but the reasons given by the teachers and students are vastly different (Dwyer & Childs, 2017). Other studies found that students don't approach Organic Chemistry as an abstract subject but utilize a more rote memorization approach (Anzovino & Bretz, 2015; Henderleiter et al., 2001; Hermanns & Kunold, 2022). These articles also pointed out that students who studied Organic Chemistry by rote memorization of the details tended to perform worse than those who understood the materials and applied them as abstract concepts. To battle these problems, instructors and educators have taken different approaches to enhance students' performance in Organic Chemistry. Many of these approaches resulted in better student perceptions, better interest in the course, and higher grades. Some examples include utilizing virtual reality for immersive learning (Dunnagan et al., 2020; Edwards et al., 2019), active learning (Crimmins & Midkiff, 2017), metacognition enhancement (Blackford et al., 2023; Blackie et al., 2023), etc. A study in 2022 found that a little over half of Pre-Medicine students do not view Organic Chemistry as an important prerequisite for their intended career path (Dixson et al., 2022). Summarizing all the ideas together, these above-mentioned studies do agree with the premise that students' perceptions in Organic Chemistry play a vital role in their outcomes for the course.

It is important to note that among students taking Organic Chemistry, there are a lot of students who are not Chemistry majors. There is also a widely accepted notion that teaching the subject to non-Chemistry majors is harder compared to Chemistry majors (Henary et al., 2015; Zotos et al., 2021; Yearty & Morrison, 2019). Nevertheless, studies have been performed to facilitate students' learning and performance in Organic Chemistry, whether for Chemistry or non-Chemistry majors. In studies where the focus of improvement is on the subjects' content, reaction mechanisms, a major content area of the course, is considered the most challenging (Anzovino & Bretz, 2015; Hermanns & Kunold, 2022; Zotos et al., 2021; Schweiker, 2020). Recent studies mainly focused on trying to help students understand "reaction mechanisms" better (Anzovino & Bretz, 2015; Hermanns & Kunold, 2022; Zotos et al., 2021), but very few have investigated the importance of these "reaction mechanisms" topics to students outside of the Chemistry major. A commentary on Premedical curriculum suggests that students taking second-semester Organic Chemistry can choose between two options: One focuses on bioorganic (and is encouraged for medical students) and the other focuses more on mechanisms and physical organic chemistry (Shulman, 2013). This approach has been taken by Oberlin College for more than 30 years, and it promptly suggests that an advanced understanding of reaction mechanisms is not necessary for medical students. Another study of Organic Chemistry in Medical Education found that the course helps improve students' critical thinking and provides basic understanding for future courses in the medical field, but the article did not touch on reaction mechanisms (Dixson et al., 2022). It is undeniable that Organic Chemistry helps a lot for medical students, as research has proven (Barr et al., 2010; Higgins & Reed, 2007; Shulman, 2013), but there is no study on the same question for reaction mechanisms. Hence, there exists a need to explore this gap in teaching Organic Chemistry into the question of whether non-Chemistry students, especially premedical students, should study every bit of reaction mechanisms, their perceptions of Organic Chemistry, and how it affects their performance.

Students' attitudes and perceptions in Organic Chemistry and in General Chemistry courses had been evaluated through different instruments, mostly Likert-scale surveys. Attitude towards the Subject of Chemistry Inventory (ASCI) is a typical Likert-scale instrument that measures different aspects of how students feel about Chemistry, including difficulty (easy–hard), complexity (simple–complicated), challenging nature, clarity, etc.

This instrument has been validated and used in many studies in the literature (Chang & Menke, 2022; Xu et al., 2015; Xu & Lewis, 2011). There are currently three versions of this instrument. The first version (ASCIv1) includes 20 items with 5 factors, while the second version (ASCIv2) includes only 8 items corresponding to 2 factors. Both versions have been validated, tested for reliability, and are supposed to convey similar information. Thus, ASCIv2 is an easier scale to administer as it has fewer questions, takes little time, and is more convenient than the first version (Xu & Lewis, 2011). The third version (ASCIv3), however, is not a fully reworked version but a version with a different item order from ASCIv2. In ASCIv2, Item 2 is Complicated–Simple, and Item 8 is Chaotic–Organized. In ASCIv3, Items 2 and 8 switch places. This change was made to avoid potentially inflated measurement errors when three items that belong to a single factor are shown in a row (Mooring et al., 2016; Rocabado et al., 2019). However, such errors have not been investigated thoroughly in these articles. A few other instruments to measure students' attitudes towards Chemistry have been developed, such as the Attitudes Toward Chemistry Lessons Scale (ATCLS) (Cheung, 2009) and Meaningful Learning in the Laboratory (MLLI) (Galloway & Bretz, 2015). The ATLCS is a modification of the previous attitude-measuring scale. It consists of 12 items of a 7-point Likert scale, with 4 dimensions for interest in chemistry: theory lessons, lab work, beliefs about school Chemistry, and behavioral tendencies to learn Chemistry. However, the scale was created for high school students and was only used and validated in Hong Kong and China instead of the United States of America. The MLLI scale measures students' cognitive and affective perceptions about their learning experiences in the undergraduate chemistry laboratory with 31-question pre- and post-surveys. This instrument focuses more on the meaningful learning outcomes from students instead of their attitudes or perceptions during laboratory sessions. However, these scales do not appear to be used as frequently in studies as the ASCIv2 scale. Other researchers prefer asking students to comment about their attitudes or perceptions without sending out Likert-scale surveys (Dunnagan et al., 2020; Dwyer & Childs, 2017; Edwards et al., 2019). In various studies, attitudes towards Chemistry (measured by ASCIv2) are often linked to other aspects of learning. Nennig et al. measured students' attitudes in a lecture-only Inorganic Chemistry course and compared them between online and in-person courses. The research found that no matter how the course was offered, online or in person, students have comparable intellectual accessibility and emotional satisfaction towards Chemistry (Nennig et al., 2020). Another study by Kahveci, conducted on high school students, formulated that attitudes and performances in Chemistry are positively related (Kahveci, 2015). Among studies using the ASCIv2, there is one study that proved that students' attitudes and achievement are not correlated (Damo & Prudente, 2019) or are weakly correlated (Brown et al., 2015). Nonetheless, this study prefers working with the ASCIv2 because it brings about the information we need, because it has been validated, and because of its common use in recent literature.

While the existing research provides valuable insights into the perspectives and experiences of organic chemistry students, it has been pointed out that the content of reaction mechanisms for non-Chemistry majors and the students' perceptions of this area haven't been looked at in the literature. This study aims to investigate students' and instructors' attitude towards Organic Chemistry (particularly reaction mechanisms), and qualitatively assess the necessity of such content for non-Chemistry students. In detail, we seek to answer the following research questions:

(1) What are students' attitudes about studying Organic Chemistry reaction mechanisms and the subject's importance for non-Chemistry major students?
(2) What are students' perceptions of the role that learning Organic Chemistry plays in their major and future career?

(3) What are instructors' perceptions towards teaching and learning Organic Chemistry mechanisms?

The completion of this project will provide a view from both learners and teachers about Organic Chemistry courses and how "reaction mechanisms" relate to the students not majoring in Chemistry. From that base, depending on the results, teachers or curriculum organizers may change their work or methods accordingly.

2. Materials and Methods

To understand students' attitudes, surveys were sent to students in West Virginia University (WVU) throughout 2 semesters, Fall 2022 and Spring 2023. The questions on the survey included demographic information, students' attitudes using ASCIv2 instrument (Xu & Lewis, 2011), Likert-scale questions on their attitude towards reaction mechanisms, some short-answer questions, and a question about whether the student would be interested in a follow-up interview. In Fall 2022, the survey was incorporated as an email sent from the CHEM233 (Organic Chemistry 1) course instructors to the students, with no form of compensation. In Spring 2023, the survey was distributed through the MIX Survey Tuesday channel of WVU (a channel that supports survey research by automatically sending surveys requested by researchers to all WVU students every Tuesday), also with no form of compensation. Table 1 shows the number of responses received:

Table 1. Responses received from surveys sent in Fall 2022 and Spring 2023.

Semester	**Fall 2022**	**Spring 2023**
Responses with at least one Organic Chemistry course(s) taken/ongoing	73	54
Responses with no Organic Chemistry courses taken/ongoing	0	17
Total responses	73	71

The data were then divided into quantitative and qualitative data. The quantitative data were analyzed with SPSS (version 27), using ANOVA and *t*-tests with a significant level of 0.05 (Wainer & Robinson, 2003). The students were grouped in many ways: by their demographic data, their current performance in Organic Chemistry, and whether they had taken an Organic Chemistry course or not. Three students expressed interest in a follow-up interview and were invited to participate in a semi-structured interview for further exploration of their survey responses. The interview data were then combined with the quantitative data from the survey, analyzed using thematic coding (identifying common themes between interview scripts), and searched for any notable information.

Regarding instructors, emails were sent to instructors to invite them to a semi-structured interview. Faculty members at the Department of Chemistry who have taught Organic Chemistry and in the School of Pharmacy were selected, as these are professionals who have had academic experience with Organic Chemistry. Invitations were sent to more than 30 faculty members, but only 5 instructors responded and agreed to participate. The interview included questions on how the faculty members conduct their teaching in Organic Chemistry, what they think about students non-Chemistry majors learning the mechanisms, and their perceptions of whether these students need the mechanisms for their future classes and careers. The interview was then transcribed and analyzed to find information that provides answers to the research questions.

A summary of the responses from the survey in Fall 2022 is listed in Table 2.

Table 2. Demographic data of Fall 2022 survey responses.

	Response	Number of Respondents	Percent
Gender	Female	30	41.1%
	Male	23	31.5%
	Prefer not to say/Unanswered	20	27.4%
	Total	73	100%
Race	Asian	1	1.4%
	Middle East/North Africa (MENA)	1	1.4%
	White	50	68.5%
	Two or more races	2	2.7%
	Unanswered	19	27.4%
	Total	73	100%

The Spring 2023 survey responses had 22 (31.0%) entries that did not have demographic questions answered. A summary of the responses from the survey in Spring 2023 is listed in Table 3.

Table 3. Demographic data of Spring 2023 survey responses.

	Response	Number of Respondents	Percent
Gender	Female	29	53.7%
	Male	8	14.8%
	Non-binary	2	3.7%
	Prefer not to say/Unanswered	15	27.8%
	Total	54	100%
Race	Asian	4	7.4%
	Black or African American	0	0%
	Hispanic	3	5.6%
	Middle East/North Africa (MENA)	1	1.9%
	White	30	55.6%
	Two or more races	1	1.9%
	Unanswered	15	27.8%
	Total	54	100%

The demographic data of the Fall 2022 and Spring 2023 survey responses do not match WVU's demographic data, and as such, the responses are not representative. The sample size of each survey when divided into smaller groups is not sufficient (n < 30) to get a reliable result for statistical tests, so the data needs to be combined. Considering this and the fact that this study focused on students who have experiences with undergraduate Organic Chemistry, analyses were performed on each group and the combined group of Fall 2022 data (n = 73) with Spring 2023 data (n = 54).

3. Results and Discussion

3.1. What Are Students' Attitudes About Studying Organic Chemistry Reaction Mechanisms and Its Importance for Non-Chemistry-Major Students?

A summary of descriptive statistics of responses for Likert scale questions (importance of Organic Chemistry in life, importance of Organic Chemistry in your major, importance of Organic Chemistry mechanisms, and seven scales of attitudes using ASCIv2) is presented in Table 4.

Table 4. Descriptive statistics of responses for Likert scale questions.

Question	Scale	Fall 2022		Spring 2023		Combined Data	
		Mean	SD [1]	Mean	SD [1]	Mean	SD [1]
Importance of Organic Chemistry in life	1–5	3.04	1.207	2.59	1.055	2.85	1.162
Importance of Organic Chemistry in major	1–5	3.51	1.120	3.11	1.192	3.34	1.163
Importance of Organic Chemistry mechanism	1–5	2.99	1.124	2.59	1.325	2.82	1.224
ASCIv2: Easy–Hard	1–7	5.21	1.213	5.43	1.207	5.3	1.112
ASCIv2: Simple–Complicated	1–7	4.97	1.536	5.37	1.186	5.14	1.407
ASCIv2: Clear–Confusing	1–7	4.47	1.684	4.46	1.610	4.46	1.647
ASCIv2: Comfortable–Not comfortable	1–7	4.10	1.757	4.17	1.891	4.13	1.808
ASCIv2: Satisfying–Frustrating	1–7	3.93	2.016	3.78	2.080	3.87	2.037
ASCIv2: Not challenging–Challenging	1–7	5.56	1.354	5.96	1.027	5.73	1.237
ASCIv2: Pleasant–Unpleasant	1–7	4.16	1.667	4.22	1.890	4.19	1.758
ASCIv2: Organized–Chaotic	1–7	3.22	1.592	3.26	1.925	3.24	1.734

[1] Standard deviation.

The Cronbach's alpha values for the importance of Organic Chemistry during Fall 2022 and Spring 2023 are 0.726 and 0.716, respectively. The Cronbach's alpha values for the ASCIv2 questions are 0.835 for Fall 2022 and 0.896 for of Spring 2023. Considering these values, the score reliability is acceptable. Regarding the importance of Organic Chemistry, the results show that students tend to think of Organic Chemistry in real life and reaction mechanisms as of below average importance. Comparison between the Fall 2022 data and the Spring 2023 data shows that the Fall 2022 students think of Organic Chemistry as more important in life than the Spring 2023 students. However, when it comes to the importance of Organic Chemistry in their major, the results show a higher value (3.51 for Fall 2022, 3.11 for Spring 2023, and 3.34 in the combined data) than other questions of importance of Organic Chemistry. This suggests that students taking this course think of Organic Chemistry as just slightly more important to their major than in other aspects of their lives. The statistics of significant results are shown in Tables 5 and 6.

Table 5. Statistics of *t*-test comparing Likert scale questions between data set.

Variable	Data Set	*n*	$M_1 - M_2$	df	t	*p*	Cohen's d
Importance of Organic Chemistry in life	Fall 2022 Spring 2023	73 54	0.449	125	2.182	0.031	0.392
ASCIv2: Not challenging–Challenging	Fall 2022 Spring 2023	73 54	1.120	3.11	1.192	3.34	1.163

Table 6. Statistics of *t*-test comparing different perception of importance of OC within data set.

Data Set	Variable	*n*	$M_1 - M_2$	df	t	*p*
Fall 2022	Importance of OC * in major Importance of OC * in life	73	0.466	72	3.006	0.004
	Importance of OC * in major Importance of OC * mechanisms	73	0.521	72	5.325	<0.001
Spring 2023	Importance of OC * in major Importance of OC * in life	54	0.519	53	2.604	0.012
	Importance of OC * in major Importance of OC * mechanisms	54	0.519	53	5.109	<0.001
Combined data	Importance of OC * in major Importance of OC * in life	127	0.488	126	3.989	<0.001
	Importance of OC * in major Importance of OC * mechanisms	127	0.520	126	7.364	<0.001

* OC = Organic Chemistry.

Students' attitudes towards Organic Chemistry were mostly negative, as seen by the fact that most mean values (Table 4) are higher than 4, which is the neutral average for 7-point Likert scale questions. There was no answer that regarded Organic Chemistry as extremely easy (1 on the Easy–Hard scale) or not challenging (1 on the Not challenging–Challenging scale). The results in all three data groups suggest that most students think of Organic Chemistry as hard, complicated, and challenging. When compared between semesters, the Fall 2022 students tend to think Organic Chemistry is less challenging than the Spring 2023 students. The statistics of significant results are shown in Table 5.

Grouping students by gender or race did not give any significant results. When grouped by performances on Organic Chemistry 1, Fall 2022 students (results shown in Table 7) with higher grades (as self-reported by respondents in groups of 10% range) seem to think of Organic Chemistry as more important than lower-achieving students. Students with better performances also had more positive attitudes: They tended to view Organic Chemistry as easier, more comfortable, and more satisfying. The combined dataset gives a more complicated result (Table 8) but holds true with the idea that students who performed better showed more positive attitudes. It is unknown whether the attitude affects the performance, the performance affects the attitude, or it is a bidirectional relationship. It is also worth noting that the number of students who had an Organic Chemistry 1 grade lower than 70% was too low (less than 10 in each case) to be considered for statistical analysis.

On the other hand, grouping students by their performances on Organic Chemistry 2 did not give any significant result, which can be attributed to the low number of responses that have finished Organic Chemistry 2 ($n = 26$). The data from the Spring 2023 respondents were also omitted, as the component groups by either Organic Chemistry 1 grades or Organic Chemistry 2 grades had fewer than 13 entries. Similarly, analyses when grouped by instructors or majors do not give statistically significant results. The statistics of significant results are shown in Table 7 (Fall 2022) and Table 8 (combined data of both semesters).

Regarding effect size, Cohen's d values were calculated (Tables 7 and 8). The absolute values of Cohen's d ranged from about 0.3 to about 0.7, with some results boasting Cohen's d values of around 1.0. Considering this, students' attitudes and their perception towards the importance of Organic Chemistry might moderately affect the students' performance or vice versa.

Table 7. Statistics of *t*-test comparing students grouped by performances within Fall 2022 dataset.

Variable	OC * 1 Grade	*n*	$M_1 - M_2$	df	t	*p*	Cohen's d
Importance of Organic Chemistry in major	90% and above 70–79%	26 19	0.678	43	2.125	0.039	0.641
	80–89% 70–79%	22 19	0.766	39	2.296	0.027	0.719
ASCIv2: Easy–Hard	90% and above 70–79%	26 19	−1.073	43	−3.469	0.001	−1.047
ASCIv2: Comfortable–Not comfortable	90% and above 70–79%	26 19	−1.063	43	−2.206	0.033	−0.666
	80–89% 70–79%	22 19	−1.311	39	−2.448	0.019	−0.767
ASCIv2: Satisfying–Frustrating	90% and above 70–79%	26 19	−1.822	43	−3.425	0.001	−1.034
	80–89% 70–79%	22 19	−1.234	39	−2.141	0.039	−0.670

* OC = Organic Chemistry.

Table 8. Statistics of *t*-test comparing students grouped by performances within the combined dataset (data from both Fall 2022 and Spring 2023).

Variable	OC * 1 Grade	*n*	$M_1 - M_2$	df	t	*p*	Cohen's d
Importance of OC * in major	90% and above 70–79%	35 27	0.717	60	2.638	0.011	0.676
Importance of OC * mechanisms	90% and above 70–79%	35 27	0.705	60	2.509	0.015	0.643
ASCIv2: Easy–Hard	90% and above 70–79%	35 27	−1.109	60	−4.064	<0.001	−1.041
	80–89% 70–79%	34 27	−0.587	59	−2.221	0.030	−0.573
ASCIv2: Clear–Confusing	90% and above 70–79%	35 27	−0.852	60	−2.247	0.028	−0.576
ASCIv2: Comfortable–Not comfortable	90% and above 70–79%	35 27	−1.026	60	−2.407	0.019	−0.616
ASCIv2: Satisfying–Frustrating	90% and above 80–89%	35 34	−1.234	67	−2.797	0.007	−0.673
	90% and above 70–79%	35 27	−1.835	60	−4.130	<0.001	−1.058
ASCIv2: Not challenging–Challenging	90% and above 70–79%	35 27	−0.686	60	−2.078	0.042	−0.532
ASCIv2: Pleasant–Unpleasant	90% and above 80–89%	35 34	−0.955	67	−2.440	0.017	−0.588
	90% and above 70–79%	35 27	−1.366	60	−3.490	<0.001	−0.894

* OC = Organic Chemistry.

3.2. What Are Students' Perceptions of the Role That Learning Organic Chemistry Plays in Their Major and Future Career?

When interviewed, students mainly stressed the challenging nature of Organic Chemistry mechanisms. One student said, "The course content is... challenging to me right now, but I believe I can ace it"; and another student commented, "The mechanisms are really challenging. I find it very hard to remember everything". These notions agreed with the qualitative data from the first research questions that students mostly view Organic Chemistry as hard, complicated, and challenging. The three students also expressed the importance of Organic Chemistry for their majors and future careers, especially for further courses and as a requirement to get into medical school. They said "I think it will probably be helpful. I actually want to go to medical school", and "It's a requirement for med school, so it must be sort of important...". From these answers, it is reasonable to say that these students don't necessarily know why Organic Chemistry is important, but they thought that it must be crucial since it is a prerequisite. They also revealed that they were having good experiences with Organic Chemistry, which can be related to the positive attitudes that they had towards the course. A student expressed, "I think it's like pretty good, and it's set up in a way that pretty much anyone could pick it up and learn it and understand it". They all acknowledged that "Organic Chemistry... kind of... use, like everyday, like, like the foods we eat, and like the products we use like on our skin, or anything like that". However, they cannot give an example where Organic Chemistry or reaction mechanism is applied in their major or future work.

Overall, it can be said that the students in this study acknowledge the difficulty of Organic Chemistry mechanisms, but view Organic Chemistry more as a requirement than as a preparation for their future career.

3.3. What Are Instructors' Perceptions Towards Teaching and Learning Organic Chemistry Mechanisms?

Of the 5 instructors interviewed, 4 were instructors from the Department of Chemistry and had various experiences in teaching Organic Chemistry. The other instructor was a clinical pharmacist who also worked as an instructor at the School of Pharmacy, but he had no experience in teaching Organic Chemistry. Overall, the responses from the instructors were similar. They all agreed that Organic Chemistry provides fundamental knowledge for courses in Health Science, and that learning reaction mechanisms helps improve problem-solving skills. One of the instructors commented, "It's less about the mechanism itself and more about kind of a thought process and the problem solving that goes into how to approach a problem". Another instructor also stated, "They may not necessarily need the mechanisms for their future life, but I think it gives them a way to problem-solve and work through things". When asked to give examples of Organic Chemistry applications, they were all able to relate Organic Chemistry to fields other than chemistry, like "drugs that can be R and S, forms of drugs where one of them is a drug that you would want someone to take, whereas the other form is, you know, maybe cancer causing" and "how some are fat, soluble, and some are water soluble". They thought that reaction mechanisms may be applied in medical school, "It still is part of the MCAT, and it's because... the logic that goes into deducing mechanism in organic chemistry is similar to the logic that is going to be applied later in medical school or dental school". However, they found it difficult to give specific examples where knowledge of reaction mechanisms would be applied outside the field of chemistry.

To summarize, the faculty members interviewed in this study had a good understanding of how Organic Chemistry is related to other fields requiring the courses, but were unable to connect reaction mechanisms to those fields.

4. Limitations

One key limitation of this study is the low sample size that was surveyed. Although the population is large enough to get more data, it is likely that the students were not interested in filling out the surveys. Low sample size limits the statistical power of the qualitative analysis, making the results from this paper hardly generalizable. Thus, the qualitative results from this study are only applicable to the samples investigated, which are not representative of any other broader populations.

As this study includes quantitative data from interviews, it may also include the researcher's bias towards the topic. Although steps were taken to minimize bias, the potential for researcher's subjectivity remains. The authors' perspectives may have inadvertently influenced data interpretation, even though efforts were made to remain objective throughout the study. Another source of bias in this study is the self-report bias through the surveys. Students' opinions on the degree of importance of Organic Chemistry can be different, and the responses they selected may not objectively reflect their perspectives. Since the questions were based on Likert scales, there may be interpretation variability between respondents, and measurement errors can emerge from the differing backgrounds of the students.

5. Conclusions

This study can set the stage for future studies revolving around how Organic Chemistry is taught. Although there are minor differences between the answers of students and faculty members, they all converge towards a main theme: Organic Chemistry is related to health science (medicine and pharmacy), but reaction mechanisms are not frequently used outside the field of chemistry. Further studies need to be carried out to confirm or refute these findings; however, regardless of the results, lecturers and curriculum organizers should reconsider the setup of Organic Chemistry courses, and think about why reaction mechanisms are taught, and how to connect them to students' future courses and careers. More survey data and more interviews need to be conducted to further explore this idea, as the more data are present the more reliable the results will be.

If these findings are further confirmed with broader populations, the Organic Chemistry curriculum should be revised and changed accordingly, preferably splitting the subject into two different course paths: one for Chemistry majors, and one for non-Chemistry majors. The path for Chemistry majors should be kept as it is with a major focus on reaction mechanisms, but the path for non-Chemistry majors should be changed. From the results of this study, a suggestion worth considering is that Organic Chemistry for non-Chemistry majors should shift the focus away from reaction mechanisms and towards applications outside of Organic Chemistry. This could be achieved by introducing only basic concepts of a reaction mechanism instead of going into deep understanding of the content and incorporating more examples for different majors to relate Organic Chemistry to their intended careers. For example, on the topic of nucleophilic substitution, the course pathway for Chemistry majors may focus deeply on the detailed mechanisms, different reaction rates across different chemical species, and exploring uncommon reactivity such as Bredt's rule or anchimeric assistances. On the same topic for non-Chemistry majors, such as Pre-Pharmacy or Pharmacy majors, a brief overview of unimolecular (S_N1) and bimolecular (S_N2) nucleophilic substitution should be presented, but lecturers should not go into too much depth about the reactivities. Instead, the time spent on this content may be used to explore the applications of nucleophilic substitutions in the human body, such as how S-Adenosyl Methionine (SAM, an antidepressant) expressed its pharmaceutical activity (Lee et al., 2023) or how 3-MCPD (3-monochloropropane-1,2-diol, a food processing contaminant) is processed in the human body (Hamlet et al., 2002). These approaches, if applied correctly, will help students broaden their knowledge of the respective fields and may encourage students to be more interested in the subject. It is also worth noting that these suggestions are subjectively based on the authors' knowledge of Chemistry and Pharmacy. Teachers and instructors may find better approaches through their own experiences or from literature in the respective fields, making suitable changes considering the compositions of their students' majors.

Author Contributions: Conceptualization, N.T. and O.O.; methodology, N.T. and O.O.; formal analysis, N.T.; investigation, N.T. and O.O.; writing—original draft preparation, N.T. and O.O.; writing—review and editing, N.T. and O.O.; project administration, O.O.; funding acquisition, O.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of West Virginia University (IRB 2209647409: Students' perception of Organic Chemistry, October 2022 (PI–O.O)).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Anzovino, M. E., & Bretz, S. L. (2015). Organic chemistry students' ideas about nucleophiles and electrophiles: The role of charges and mechanisms. *Chemistry Education Research and Practice*, *16*(4), 797–810. [CrossRef]

Barr, D. A., Matsui, J., Wanat, S. F., & Gonzalez, M. E. (2010). Chemistry courses as the turning point for premedical students. *Advances in Health Sciences Education*, *15*(1), 45–54. [CrossRef]

Blackford, K. A., Greenbaum, J. C., Redkar, N. S., Gaillard, N. T., Helix, M. R., & Baranger, A. M. (2023). Metacognitive regulation in organic chemistry students: How and why students use metacognitive strategies when predicting reactivity. *Chemistry Education Research and Practice*, *24*(3), 828–851. [CrossRef]

Blackie, M. A. L., Arnott, G., & Kaschula, C. H. (2023). Engaging organic chemistry students in knowledge building. *Journal of Chemical Education*, *100*(9), 3302–3308. [CrossRef]

Brown, S. J., White, S., Sharma, B., Wakeling, L., Naiker, M., Chandra, S., Gopalan, R., & Bilimoria, V. (2015). Attitude to the study of chemistry and its relationship with achievement in an introductory undergraduate course. *Journal of the Scholarship of Teaching and Learning*, *15*(2), 33–41. [CrossRef]

Chang, J., & Menke, E. J. (2022). Measuring attitude toward chemistry, biology, and math at a hispanic-serving institution. *Journal of Chemical Education*, *99*(4), 1758–1765. [CrossRef]

Cheung, D. (2009). Developing a scale to measure students' attitudes toward chemistry lessons. *International Journal of Science Education*, *31*(16), 2185–2203. [CrossRef]

Collini, M. A., Miguel, K., Weber, R., & Atkinson, M. B. (2024). Investigating changes in students' attitudes towards organic chemistry: A longitudinal study. *Chemistry Education Research and Practice*, *25*, 613–624. [CrossRef]

Crimmins, M. T., & Midkiff, B. (2017). High structure active learning pedagogy for the teaching of organic chemistry: Assessing the impact on academic outcomes. *Journal of Chemical Education*, *94*(4), 429–438. [CrossRef]

Damo, K. L., & Prudente, M. S. (2019, January 10–13). *Investigating students' attitude and achievement in organic chemistry using interactive application*. 10th International Conference on E-Education, E-Business, E-Management and E-Learning (pp. 36–41), Tokyo, Japan. [CrossRef]

Dixson, L., Pomales, B., Hashemzadeh, M., & Hashemzadeh, M. (2022). Is organic chemistry helpful for basic understanding of disease and medical education? *Journal of Chemical Education*, *99*(2), 688–693. [CrossRef]

Dunnagan, C. L., Dannenberg, D. A., Cuales, M. P., Earnest, A. D., Gurnsey, R. M., & Gallardo-Williams, M. T. (2020). Production and evaluation of a realistic immersive virtual reality organic chemistry laboratory experience: Infrared spectroscopy. *Journal of Chemical Education*, *97*(1), 258–262. [CrossRef]

Dwyer, A. O., & Childs, P. E. (2017). Who says organic chemistry is difficult? Exploring perspectives and perceptions. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(7), 3599–3620. [CrossRef]

Edwards, B. I., Bielawski, K. S., Prada, R., & Cheok, A. D. (2019). Haptic virtual reality and immersive learning for enhanced organic chemistry instruction. *Virtual Reality*, *23*(4), 363–373. [CrossRef]

Galloway, K. R., & Bretz, S. L. (2015). Measuring meaningful learning in the undergraduate general chemistry and organic chemistry laboratories: A longitudinal study. *Journal of Chemical Education*, *92*(12), 2019–2030. [CrossRef]

Gupta, T., & Hartwell, S. K. (2019). Enhancing student retention in general and organic chemistry: An introduction. In *From general to organic chemistry: Courses and curricula to enhance student retention* (Vol. 1341, pp. 1–12). American Chemical Society. ISBN 978-0-8412-3748-3.

Hamlet, C., Sadd, P. A., Crews, C., Velisek, J., & Baxter, D. E. (2002). Occurrence of 3-chloro-propane-1,2-diol (3-MCPD) and related compounds in foods: A review. *Food Additives and Contaminants*, *19*(7), 619–631. Available online: https://www.researchgate.net/publication/11264350_Occurrence_of_3-chloro-propane-12-diol_3-MCPD_and_related_compounds_in_foods_A_review (accessed on 12 November 2024). [CrossRef]

Henary, M., Owens, E. A., & Tawney, J. G. (2015). Creative report writing in undergraduate organic chemistry laboratory inspires nonmajors. *Journal of Chemical Education*, *92*(1), 90–95. [CrossRef]

Henderleiter, J., Smart, R., Anderson, J., & Elian, O. (2001). How do organic chemistry students understand and apply hydrogen bonding? *Journal of Chemical Education*, *78*(8), 1126. [CrossRef]

Hermanns, J., & Kunold, H. (2022). Mechanism comics as a task in a written exam in organic chemistry for pre-service chemistry teachers. *Chemistry Teacher International*, *4*(3), 259–269. [CrossRef]

Higgins, T. S., & Reed, S. F. (2007). Changing premedical requirements. *JAMA*, *297*(1), 37–39. [CrossRef]

Kahveci, A. (2015). Assessing high school students' attitudes toward chemistry with a shortened semantic differential. *Chemistry Education Research and Practice*, *16*(2), 283–292. [CrossRef]

Lee, Y.-H., Ren, D., Jeon, B., & Liu, H. (2023). S-Adenosylmethionine: More than just a methyl donor. *Natural Product Reports*, *40*(9), 1521. [CrossRef] [PubMed]

Mooring, S. R., Mitchell, C. E., & Burrows, N. L. (2016). Evaluation of a flipped, large-enrollment organic chemistry course on student attitude and achievement. *Journal of Chemical Education*, *93*(12), 1972–1983. [CrossRef]

Nennig, H. T., Idárraga, K. L., Salzer, L. D., Bleske-Rechek, A., & Theisen, R. M. (2020). Comparison of student attitudes and performance in an online and a face-to-face inorganic chemistry course. *Chemistry Education Research and Practice*, *21*(1), 168–177. [CrossRef]

Paulson, D. R. (1999). Active learning and cooperative learning in the organic chemistry lecture class. *Journal of Chemical Education*, *76*(8), 1136. [CrossRef]

Rocabado, G. A., Kilpatrick, N. A., Mooring, S. R., & Lewis, J. E. (2019). can we compare attitude scores among diverse populations? An exploration of measurement invariance testing to support valid comparisons between black female students and their peers in an organic chemistry course. *Journal of Chemical Education*, *96*(11), 2371–2382. [CrossRef]

Schweiker, S. (2020). Engaging health student in learning organic chemistry reaction mechanisms using short and snappy lightboard videos. *Journal of Chemical Education*, *97*(10), 3867–3871. [CrossRef]

Shulman, J. I. (2013). Chemistry in the premedical curriculum: Considering the options. *Journal of Chemical Education*, *90*(7), 813–815. [CrossRef]

Szu, E., Nandagopal, K., Shavelson, R. J., Lopez, E. J., Penn, J. H., Scharberg, M., & Hill, G. W. (2011). Understanding academic performance in organic chemistry. *Journal of Chemical Education*, *88*(9), 1238–1242. [CrossRef]

Vilia, P. N., Candeias, A. A., Neto, A. S., Franco, M. D. G. S., & Melo, M. (2017). Academic achievement in physics-chemistry: The predictive effect of attitudes and reasoning abilities. *Frontiers in Psychology*, *8*, 1064. [CrossRef]

Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher*, *32*(7), 22–30. [CrossRef]

Xu, X., Alhooshani, K., Southam, D., & Lewis, J. E. (2015). Gathering psychometric evidence for ASCIv2 to support cross-cultural attitudinal studies for college chemistry programs. In M. Kahveci, & M. Orgill (Eds.), *Affective dimensions in chemistry education* (pp. 177–194). Springer. ISBN 978-3-662-45085-7. [CrossRef]

Xu, X., & Lewis, J. E. (2011). Refinement of a chemistry attitude measure for college students. *Journal of Chemical Education*, *88*(5), 561–568. [CrossRef]

Yearty, K. L., & Morrison, R. W. (2019). Engaging nonchemistry majors through application-based final projects in the elementary organic chemistry classroom. In *Communication in chemistry* (Vol. 1327, pp. 129–148). American Chemical Society. ISBN 978-0-8412-3510-6. [CrossRef]

Zhang, C., Kuncel, N. R., & Sackett, P. R. (2020). The process of attrition in pre-medical studies: A large-scale analysis across 102 schools. *PLoS ONE*, *15*(12), e0243546. [CrossRef] [PubMed]

Zotos, E. K., Tyo, J. J., & Shultz, G. V. (2021). University instructors' knowledge for teaching organic chemistry mechanisms. *Chemistry Education Research and Practice*, *22*(3), 715–732. [CrossRef]

education sciences

MDPI

Article

Exploring the Nature and Role of Students' Peer-to-Peer Questions During an In-Class Collaborative Activity

Tarah M. Dahl [1], Krystal Grieger [2], Soren Miller [3] and James Nyachwaya [4,*]

[1] School of Education, North Dakota State University, Fargo, ND 58102, USA; tarah.dahl@ndsu.edu
[2] Department of Chemistry and Biochemistry, North Dakota State University, Fargo, ND 58102, USA; krystal.grieger@ndsu.edu
[3] Chemistry Department, St Olaf College, Northfield, MN 55057, USA; miller80@stolaf.edu
[4] Department of Chemistry and Biochemistry, and School of Education, North Dakota State University, Fargo, ND 58102, USA
* Correspondence: james.nyachwaya@ndsu.edu

Abstract: During group activities, instructors expect that students will ask each other questions. Therefore, in this study, we looked at the nature and role of peer-to-peer questions during an in-class activity. During the activity, students worked collaboratively to respond to five prompts about an acid–base neutralization reaction. We examined the questioning behavior in groups and the nature and types of questions asked. We then looked specifically at the content questions, analyzing how they varied by prompt, as well as the level of those content questions using Bloom's taxonomy. Finally, we looked at the role that the peer-to-peer questions played as the students completed the activity. The results revealed that the students broadly asked each other social questions, process questions, and content questions, with content questions being the most frequently posed. The prompts that required students to make a prediction, sketch a graph, and explain their reasoning elicited most of the content questions asked. Furthermore, most of the peer-to-peer content questions asked across the five prompts ranked at the two lowest levels of Bloom's taxonomy. Finally, the posed peer-to-peer questions were found to play many roles in the discussion, including initiating and sustaining conversations, seeking consensus, challenging each other, and promoting social metacognition. The implications for instruction and research are discussed.

Keywords: collaborative learning; constructivism; metacognition; student questions; Bloom's taxonomy

1. Introduction

Carner (1963) noted that "the evidence that good teaching has taken place is reflected more in the kinds of questions students ask than the abundance of past answers they can produce" (p. 550). Questioning is central to guiding inquiry in science and is one of the science and engineering practices (SEPs) identified in the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013). Questioning plays a vital role in facilitating meaningful student learning experiences.

Questions in an educational setting can originate from students to teachers, teachers to students, students to other students, or even from a text to students. Integrating carefully crafted questions into the classroom can increase student motivation to learn, promote productive discussions, and support knowledge construction (Chin & Osborne, 2008; Yu, 2009). In fact, questions can serve a variety of functions, including confirming expectations, solving problems, and filling gaps in knowledge and understanding (Biddulph & Osborne, 1982). Student questions can provide feedback to teachers because they can reveal student conceptual understanding, as well as the nature and quality of their ideas (Woodward, 1992). Student questions can also influence the curriculum, especially when they highlight areas of student interest (Crawford et al., 2000; Watts et al., 1997). Despite the known benefits, research has shown that students seldom ask questions, and if they do, the questions are often low-level and require little to no cognitive processing (Chin et al., 2002).

Much of the extant research on student questions has been conducted in K-12 settings, looking, among other things, at the nature of the questions students asked during inquiry laboratories or during argumentation (e.g., Chin et al., 2002; Dogan & Yucel-Toy, 2022; Erdogan, 2017; Lai & Law, 2013). At the college level, research around student questions has mostly focused on written questions asked by students after reading a textbook chapter (Marbach-Ad & Sokolove, 2000) and students' written questions based on lecture material (Harper et al., 2003). Unfortunately, there is limited research looking at the peer-to-peer questions that arise spontaneously during a small-group in-class activity. Therefore, this study seeks to add to the existing research by answering the following four research questions:

1. What was the nature of questioning behavior within student groups?
2. What types of questions did students ask?
3. What was the nature of the content questions asked in the student groups?
4. What roles or functions did peer-to-peer questions play in completing the collaborative task?

2. Literature Review

Questions are central to meaningful learning, and using questioning techniques in the classroom can have many benefits for students. Research shows that intentionally delivered questions can increase student motivation to learn the content, promote lively and productive discussions, and guide students in their knowledge construction (Chin & Osborne, 2008; Yu, 2009). Questioning is a fundamental aspect of learning, particularly in collaborative contexts. Classroom discussions among students can stimulate questions that encourage relevant problem-solving strategies, thus fostering collaborative knowledge building and productive discourse (Chin et al., 2002). Teachers play a key role in encouraging questions in the classroom (van Zee et al., 2001). When teachers intentionally create collaborative learning environments that elicit student questions and promote peer discussions, students are more likely to generate their own questions during discourse (Penuel et al., 2004).

In a traditional science classroom, it is usually the instructor who poses the questions for students to answer (Nystrand et al., 2003). However, research has shown that providing opportunities for students to generate their own questions enables them to focus on the key aspects of the content and determine their level of understanding (Rosenshine et al., 1996). Therefore, a central goal of science education is to help students develop the ability to ask questions (Bybee, 2000; Chin & Osborne, 2008; Hakkarainen, 2003), since asking questions is fundamental to both science and scientific inquiry. Indeed, the ability to ask good scientific questions is an important aspect of scientific literacy (Miller & Osborne, 1988) and plays a central role in scientific discourse, especially through eliciting explanations, evaluating evidence, and clarifying doubts (Chin & Osborne, 2008).

The type of questions posed influences the knowledge produced and, consequently, the depth of understanding (Harper et al., 2003). Student-generated questions are an important aspect of both self- and peer assessment (Black et al., 2002). Furthermore, intellectually challenging questions lead to student engagement and learning (Zeegers &

Elliott, 2019). Peer-to-peer questions during group collaborative activities help promote productive discussions, which in turn lead to the co-construction of knowledge (Chin et al., 2002). Nevertheless, the ability to generate higher-order questions must be cultivated—it is not an inherent skill for all students. In fact, according to Chin et al. (2002), students do not naturally ask these types of questions. Thus, it is essential that teachers provide students with opportunities to engage in collaborative tasks in which they have opportunities to engage in questioning one another.

Furthermore, the tasks teachers use to engage their students can shape the questions that students generate. For instance, when students are asked to follow specific procedures, it often leads to the generation of factual questions, while when students are given open-ended tasks, it can encourage curiosity-driven questions (Chin et al., 2002). This is particularly important because the types of questions students pose can yield different learning outcomes. Explanation-seeking questions foster deeper understanding, which is more effective for knowledge construction than fact-based questions (Hakkarainen, 2003; Zhang et al., 2007). Research shows that students generating factual questions primarily build basic knowledge, while those posing explanation-seeking questions tend to develop more comprehensive explanations of the content (van Aalst, 2009). Therefore, to encourage the development of questions that seek deeper understanding, it is essential that instructors engage their students in learning tasks that are open-ended.

In contrast, peer-to-peer questioning can significantly contribute to social learning dynamics and the overall effectiveness of group interactions. According to Webb (2009), students that engage in questioning each other create an interactive environment that supports peer learning. Webb (2009) discovered that peer interactions of this nature not only helped students understand complex concepts but also increased a sense of community among the students. This research informs our study by highlighting the benefits of peer-to-peer questioning, specifically the increase in group cohesion and collective problem solving when asked to work in collaborative settings.

The research has proposed many ways of classifying cognitive questions. Most are based on Bloom's taxonomy (Blosser, 2000). Bloom's taxonomy has six categories: remember, understand, apply, analyze, evaluate, and create (1956). Remembering is the lowest level and requires that students recall or retrieve material already learned, while creating is the highest level. Therefore, questions increasingly demand higher-order thinking as one moves up the categories from remembering towards creating.

3. Theoretical Framework

This study is grounded in the theories of both social constructivism and social metacognition. Under social constructivism, knowledge and understanding of the world are developed by individuals, and meanings are developed in coordination with others. Social constructivists believe that meaningful learning occurs when individuals are engaged in social activities such as interaction and collaboration. Learning is therefore a social process, and meaningful learning occurs when individuals are engaged in a social activity that happens within a community (Adams, 2006). The development of ideas in this context is a collaborative process in which students work together and evaluate each other's ideas. Indeed, constructivist classrooms entail group work, dialogue, and shared norms (Smith et al., 2000). It can be argued that a collaborative environment, in which students work together—such as during problem-solving activities—is essential for building knowledge (Powell & Kalina, 2009). In the current study, students worked collaboratively to complete the assigned in-class activity. As they worked together, they proposed and shared ideas, asked each other questions, and answered their peers' questions.

Social metacognition, also called socially shared metacognition, involves awareness and control of other learners' thinking. In a group setting, such as during collaborative learning, students share ideas with their peers, invite others to critique and evaluate their ideas, evaluate their peers' ideas, and assess, change, and use each other's strategies (Goos et al., 2002; Van De Bogart et al., 2017). It is worth noting that while metacognition at the individual level involves being aware of and controlling one's own thinking, metacognition and social metacognition affect each other since social metacognition aids individual metacognition and individual metacognition supports social metacognition. Indeed, social metacognition shares the metacognitive demands among group members, thus making visible each person's metacognition and improving individual metacognition (Chiu & Kuo, 2009). According to Vygotsky, in a social classroom setting, students receive feedback that can help them monitor, evaluate, and adjust their performance (Vygotsky, 1978). Therefore, social metacognition allows students to help each other evaluate metacognitive strategies and learn new ones (Anderson, 2004).

Metacognition involves three skills: planning, monitoring, and evaluating. Planning involves assessing the problem at hand and the resources available, such as time. Monitoring at the individual and group level involves keeping track of one's own understanding, such as by checking with peers to see if one had the correct idea or monitoring someone else's statements for correctness. It can also involve a student checking to make sure the group has a shared understanding or a student asking their peers for more information beyond a stated response. Evaluation occurs at the levels of both the self and the group so that a student can assess their own thinking or solution as well as that of the group (Halmo et al., 2022).

4. Materials and Methods

4.1. Research Setting

This study was conducted within a General Chemistry (II) course for non-majors at a medium-sized research-intensive university in the midwestern United States. Students enrolled in this course are expected to have previously taken General Chemistry (I). This three-credit course was delivered through three weekly 50 min sessions over the 15-week semester. Instruction in the course consisted of a mixture of direct instruction (lecture), small-group discussions, whole-class discussions, and in-class collaborative activities, such as the one used in this study. Furthermore, clicker questions were used as formative assessments during each class meeting. The clicker questions were usually accompanied by small-group (turn-to-your-neighbor) as well as whole-class discussions. The activity that is the focus of this study, illustrated in Figure 1, was completed after finishing the chapter 'Aqueous Ionic Equilibrium', in which we covered the activity's concepts of neutralization, titrations, and pH curves (pH-metric titrations). Previously in the course, we also covered the concept of electrolytes. The concept of acid–base neutralization was first introduced in the General Chemistry (I) course. In this course, students completed a laboratory activity in which they reacted an acid with a base while using an indicator to monitor and determine the end point of the neutralization reaction.

During a lab activity on acids and bases, students were reacting sodium hydroxide solution and dilute hydrochloric acid. Students were instructed to add the base to the acid till the base was in excess. The progress of the reaction was monitored by measuring the amount of current conducted by the reaction mixture as sodium hydroxide was added to the acid.

a. Write a balanced equation for the acid-base reaction.
b. Write a net ionic equation for the neutralization reaction.
c. (i). Suppose you were measuring electrical conductivity of the reacting mixture as the base was added to the acid, predict how the conductivity would change.
(ii). Provide an explanation for your prediction.
d. Sketch a graph illustrating your prediction in 'c' above, clearly labelling the axes and indicating the endpoint.
e. Explain why you drew the graph as you did in 'd'.

Figure 1. Collaborative activity instructions and questions (Nyachwaya, 2016).

4.2. Data Collection

In this study, we sought to characterize and analyze peer-to-peer questioning among students during an in-class collaborative group activity. At the end of the unit on 'Aqueous Ionic Equilibrium', the students were assigned an in-class collaborative activity that addressed acid–base neutralization. The activity described a scenario involving the reaction between an acid and a base. The students were asked to work in self-selected groups of 2–4 to complete the activity shown in Figure 1. They recorded their conversations during the activity and shared the audio files with the instructor. The students had 20 min to complete the activity. There were 183 students who consented to take part in the study, comprising 63 groups.

It is worth noting that in the laboratory course, students completed a similar exercise in which they determined whether a reaction would occur when two solutions were mixed. Part of that process involves writing complete and net ionic equations, which requires understanding and applying the solubility rules.

4.3. Data Analysis

In this study, both oral and written data were collected. All audio files of the student conversations were transcribed verbatim, and the transcribed data were subsequently coded and analyzed alongside the written responses. Transcription was followed by fact checking (Tracy, 2013), which involved listening to the recordings while also reading transcripts to ensure accuracy. Corrections were made where necessary. During transcription, we listened to each student speaking in turn, making sure that we separated each speaker's vocalizations accurately. Each speaker was assigned a number based on the order in which they spoke (the first speaker was assigned a 1, the second speaker was assigned a 2, etc.). Written worksheets and the transcripts were anonymized by replacing student names with pseudonyms. We ensured that each group's transcript was separated before analysis began.

Coding of the data started with first identifying all inquiries in the student transcripts that could be classified as peer-to-peer questions. These inquiries were then quantified to obtain a sense of how many total peer-to-peer questions were asked across all the groups. We identified and counted instances when questions were asked within each group's transcript with 100% inter-rater agreement. Next, the team of three researchers co-developed codes that were informed by patterns emerging from the data as well as previous research. We tallied up occurrences of each coded category and looked for emergent patterns. The analysis focused on the types of questions asked, whether certain prompts elicited more content questions than others, the levels of content questions asked based on Bloom's taxonomy, and the role that the questions played as students worked on completing the collaborative activity.

To establish consistency in coding, the three researchers read through the transcripts from four groups and analyzed the student questions for the type of question, the number of content questions asked in each prompt, the level of content questions asked for each prompt (based on Bloom's taxonomy), and the role that the questions played during student collaborations as they completed the activity. We determined the type of questions asked by examining the objective of the question or the type of information the question sought. We then focused on content questions, first analyzing which prompts elicited more content questions. We expected that tasks students were familiar with would elicit fewer content questions, such as balancing an equation. We further analyzed the content questions using Bloom's taxonomy to ascertain the level and proportion of content questions asked for each prompt. Finally, we looked at the role each question played by looking at attributes such as what the question sought from peers or how it contributed to the overall problem-solving process. The average inter-rater agreement for the four transcripts was 94%. Differences were resolved through discussion. Two researchers then divided up and coded the remaining dataset.

5. Results

The goals of this study were threefold: characterize the peer-to-peer questions; analyze patterns in the peer-to-peer questions; and identify the roles or functions that these questions played during the collaborative activity. The results indicated that a total of 959 questions were asked across all 63 groups, with an average of 15 questions per group. Overall, 2 to 44 questions were asked within each group.

Furthermore, as shown in Figure 2, the analysis of the correctness of their submitted responses indicated that students did relatively better on algorithmic prompts or tasks with which they were familiar than on tasks that required the transfer of knowledge or reasoning, which is what we expected.

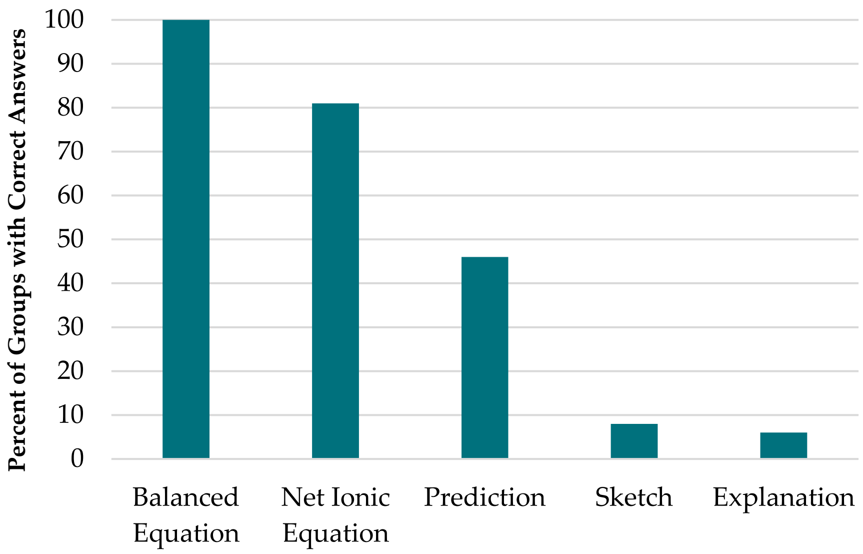


Figure 2. Percentage of groups with correct responses for each of the five prompts.

5.1. What Was the Nature of Questioning Behavior Within Groups?

To answer this question, we looked at patterns in the asking of questions within each group. In particular, we sought to determine whether all group members asked at least one question. Promisingly, in most groups, all the students engaged in asking each other questions. In fact, in 53 of the 63 groups, all group members asked at least one question. Of the 10 groups in which not everyone asked a question, there were 4 groups that had two members with only one that asked questions; 4 groups of three members with only two of the three group members that asked questions; and 2 groups with four members with three of the four participants that asked questions.

5.2. What Types of Questions Did Students Ask?

The students' peer-to-peer questions fell into three main categories: social, process, and content. Below, each category of questions is described, and a sample excerpt is used to illustrate the question type. Within the following excerpts, all the questions are italicized, and representative questions are both italicized and bolded.

5.2.1. Social Questions

The social questions were not about the assigned task or the content of acid–base neutralization; these questions did not help students understand the activity better, nor were they related to the content of the activity. The bolded questions in the excerpt below are examples of 'social' questions.

S1: C'mon. I'm the calligrapher. All right, if you have one, I'll take it.
S2: ***Calligrapher, do you want a pen? Do you want a fountain pen?***
S1: If you have one, I'll take it.
S2: I do have like not but like you played with before.
S1: I'll draw something pretty.
S2: ***Like a smiley face?***
S1: Yeah, like or whatever you want really.
S3: ***Oh, really?***
S1: Yeah.
S3: ***Like your doodle?***
S1: Yeah.
S2: I'm sure he'd appreciate that. ***Garfield?*** You should draw Garfield.
S1: I will. I will draw Garfield.
S2: ***Garfield cool?***
S3: No.

5.2.2. Process Questions

The process questions were those inquiries that sought out what was required in the activity; what a specific prompt was asking of students; or how to complete a process such as balancing an equation or sketching a graph. The excerpt below has a bolded example of a 'process' question.

S1: Net ionic equation.
S2: I'm gonna look something up.
S2: *Oh, is it adding the plus and minuses and aqueous?* Yeah.
S1: *Which ones how do we know if it's aqueous or not?*
S2: I don't know how we know that though.
S1: ***Well should we just write everything spaced out like that and then have the same on each side like that?***
S2: Well, that makes sense. *Is that kind of like what they have?*

S1: I'll just do that.

5.2.3. Content Questions

The content questions inquired about the activity's subject matter, which included acid–base neutralization and conductivity. In the excerpt below, the bolded questions are examples of content questions asked within a group.

S4: Is that electronegativity?
S2: Did we even talk about it?
S1: OK, it's like the stronger the acid or the stronger the base, the higher the conductivity?
S4: Wouldn't it become more electrically conductive, since they're splitting off into electrons?
S1: But won't it decrease since you're neutralizing it?
S2: Yeah, if.
S4: Oh yeah, that's true.
S2: It's weak electrical conductivity, It's a weak acid or base.
S1: Yeah, it's like you're getting closer to seven, which is neutral pH.
S3: Yeah.
S2: And sodium hydroxide, that's a strong base, right?

Figure 3 shows the relative proportions of each type of question posed. Most of the peer-to-peer questions were content questions, with social questions being the fewest.

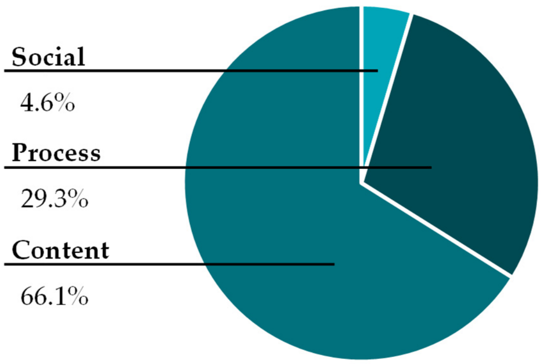


Figure 3. Proportion of the categories of questions asked.

Due to their prevalence, we chose to further analyze the content questions to uncover which of the activity prompts elicited the greatest number of questions and their Bloom's taxonomy classification. We assumed the prompts that were algorithmic in nature, such as writing and balancing an equation, would elicit fewer questions because we expected most students would already know how to perform these tasks. Similarly, we expected that the prompts that asked students to make a prediction, sketch a graph, or explain their reasoning would elicit the most content questions.

Figure 4 shows the proportion of content questions elicited by each prompt. Prompts A and B, which required writing and balancing of equations, elicited almost one-fourth of all the content questions. However, as expected, prompts C and D, which were the most challenging for the students, elicited most of the content questions.

Furthermore, we analyzed the types of content questions that students asked using Bloom's taxonomy (Bloom, 1956). We expected that the algorithmic-type prompts (such as writing and balancing an equation) would elicit questions at lower levels of Bloom's taxonomy, while prompts asking for predictions, sketching (creating), and explaining would elicit questions at higher levels of Bloom's taxonomy. Figure 5 shows the proportions of content questions asked at each of the levels of Bloom's taxonomy. Unfortunately, across all prompts, the largest proportion of the questions posed were at the two lowest levels of Bloom's taxonomy. However, it is encouraging to see that some of the peer-to-peer questions ranked high in Bloom's taxonomy.

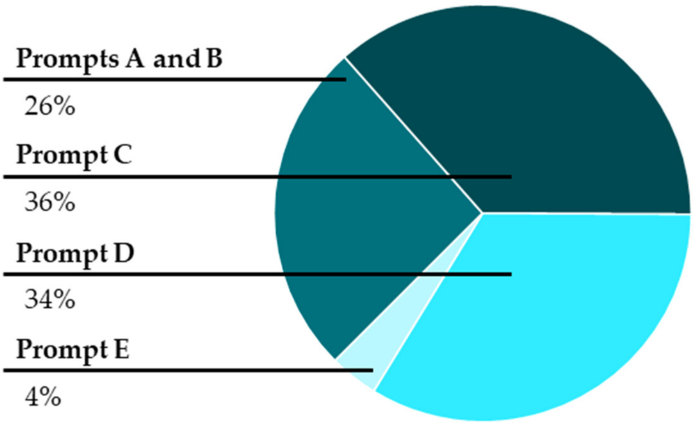


Figure 4. Percentage of content questions elicited by each of the activity's prompts.

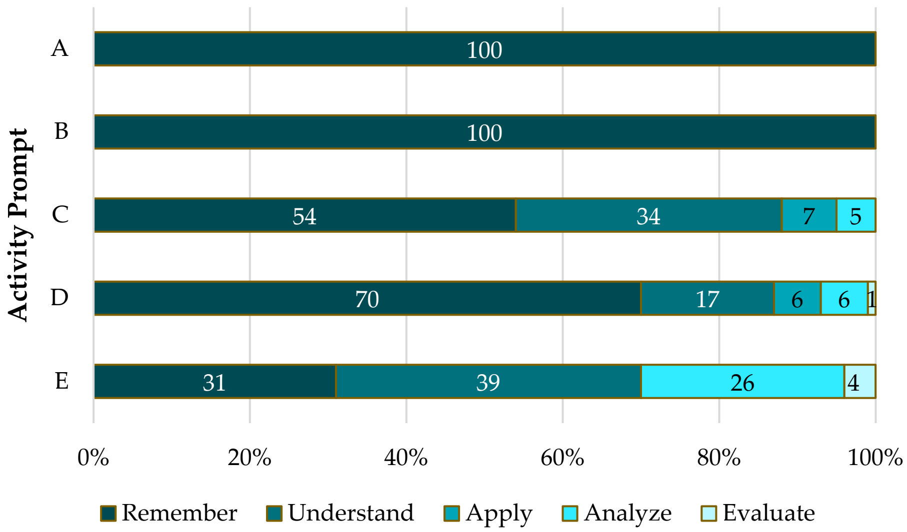


Figure 5. Proportion of content questions asked, ranked using Bloom's taxonomy.

5.3. What Roles or Functions Did the Peer-to-Peer Questions Play in Completing the Collaborative Task?

We found that the peer-to-peer questions helped to initiate and promote conversations and engagement in the activity, elicit additional ideas, challenge ideas raised in groups, express one's ideas or doubts, seek consensus, and promote social metacognition. Each of these roles/functions is described below, with illustrative examples provided.

5.3.1. Initiate and Promote Conversations and Engagement

We noted that in many groups, conversations started with a question from one of the group members. As in the excerpt below, the questions were followed by responses, which led to more questions and responses. In this way, the peer-to-peer questions both initiated and sustained conversations and therefore helped keep students engaged in their collaborative activity. The excerpt below is from a group in which a total of 44 peer-to-peer questions were asked.

*S1: **Isn't hydrochloric acid just HCl?***
S2: Yeah, just HCl.
S3: Write a balanced equation for the acid base reaction.
*S2: **Then you get water at the end too, right?***

S3: Yeah. It's NaOH + HCl.
*S2: **What is it supposed to equal?***
S1: Water and NaCl.
S2: It is sodium chloride and water.
S3: And then this is our balanced equation. And write a net ionic equation for neutralize.
S2: So, the net ionic equation. ***Isn't that just the hydrogens and the OH?*** Then we get the water.
*S3: **Would you cancel out the water?***
S2: So, I'm not sure.

5.3.2. Express One's Ideas

We noted that in some cases the students expressed their ideas as questions. It could be that these students were unsure of their ideas, or they were seeking confirmation of their ideas from their peers. In the bolded question below, the student seems to be commenting on their peer's statement but phrases their comment in the form of a question.

S2: And then I think it's supposed to be add base. *For our prediction, I think the conductivity would start positive because of the H pluses, right?*
S1: Yeah.
S2: So, start out positive charge from the H plus of the acid and then is slowly neutralized with the strong base.
*S1: **So, then the conductivity would become more negative?***
S2: Yeah, then the conductivity is negative as excess strong base is added.

5.3.3. Elicit Additional Ideas

The students asked questions to elicit additional ideas from their peers. For example, in the bolded question below, the student posed a question to see if their peers had any more ideas to add.

S1: To make it more or to have a higher level of electrical conductivity? Yes.
You're going to write about excess NaOH?
*S2: **So, anything else?***
S1: I think we're good. We've got a bit. I think we're at the equivalence point.
S2: Should we go at the equivalence point? Would it be zero?
S1: I don't know if it's like zero because I don't know how it's measured. But I'd say it's like neutral.
S2: Can conductivity be neutral?
S1: Yeah, I think so. This or maybe it is zero. I think either way it gets your point across.

5.3.4. Challenge Ideas Raised by Peers

Our results showed that the students challenged their peers' ideas by asking questions. The bolded question below is an example where a student offered a counter idea by posing a question.

S4: Is that electronegativity?
S2: Did we even talk about it?
S1: OK, it's like the stronger the acid or the stronger the base, the higher the conductivity?
S4: Wouldn't it become more electrically conductive, since they're splitting off into electrons?
S1: But won't it decrease since you're neutralizing it?
S2: Yeah, if.
S4: Oh yeah, that's true.
S2: It's weak electrical conductivity. It's a weak acid or base.

S1: Yeah, it's like you're getting closer to seven, which is neutral pH.
S3: Yeah.
S2: And sodium hydroxide, that's a strong base, right?

5.3.5. Seek Consensus

One of the ways that students checked for consensus was through asking questions to ensure that the whole group agreed. In the excerpt below, the bolded question illustrates a student checking whether or not their group agreed on the final response.

S1: So, to make a balanced equation, is it like NaOH plus HCl are those the two things we're working with?
S2: That's what I understood.
S1: OK, so what do we make with that?
S2: So, can you just go ok, so if you're making them an acid and base this is your, this is your acid.
S1: So, would it be NaCl and H_2O?
S2: That's what I would say.
S3: OK, are we in the first part?
S1: That's what we did. ***Ok, do we think that's right?***
S2: I think so.

5.3.6. Promoting Social Metacognition

Our results show that the peer-to-peer questions promoted social metacognition. We found that students in groups exhibited the metacognitive skills of monitoring and evaluation. For example, in the excerpt below, the bolded question shows a student checking with their peers to see if their reasoning is right.

S1: (Reads aloud the question from assignment about predicting conductivity)
S4: It would become less conductive conductivity would go lower, right?
S2: Yes.
S4: Because you are adding a base. ***Is that right?***
S2: Yea I think so.
S1: Base was added I think the, yeah it would go down. *Is what you said?*
S3: What did you say?
S1: Okay, if the base is added.
S2: The conductivity goes down.
S1: The conductivity will go down.

6. Discussion and Conclusions

This study sought to characterize peer-to-peer questions, analyze patterns in those questions, and explore the role that the questions played during the course of completing a collaborative activity. In a classroom setting where students are collaborating on an assigned activity, peer-to-peer questions are a form of student interaction as well as a way for students to engage with the activity (Fredricks et al., 2004). Such student engagement has cognitive, behavioral, and emotional dimensions (Naibert et al., 2022). More specifically, asking questions is considered to be a form of behavioral engagement (Naibert et al., 2022). As expected, in this study student engagement—as characterized by question frequency—varied with 2–44 questions posed per group; students in most of the groups asked at least one question. Therefore, through the lens of peer-to-peer questions, most students were engaged in the collaborative activity.

Moreover, from our results, one key role that the peer-to-peer questions played was that of initiating and sustaining conversations within the groups. In the sample excerpt used to illustrate this role, the conversation in that group started with a question, followed by conversations interspersed with questions.

As indicated above, this study identified three broad categories of peer-to-peer questions: social, process, and content questions. Social questions were the least commonly asked questions, while content questions were the most commonly asked. This result differs from the extant research in which most of the questions asked were procedural questions (Tiffany et al., 2023). It is worth noting here that the context of each activity may affect this finding. While social questions are not relevant to the subject matter of the activity, we suggest that they play a role in the social dynamics of a group. For example, students who engage in side talk would seem to be more 'free' or open with each other and are therefore more likely to contribute to the activity. Process questions help students obtain clarification on the various aspects of the activity, such as understanding the meaning of specific prompts or how to complete each part of the activity. Content questions, which were the majority of the questions in this study, could point to a number of things, such as gaps in what students know or what they are thinking about (Chin & Osborne, 2008).

When students enter a course, the instructors for the course assume the students possess a basic level of prior process and content knowledge. In this collaborative activity within General Chemistry II, we presumed that all students would be fluent in writing and balancing equations. Students were expected to transfer their understanding of acid–base neutralization to the scenario in the activity. We expected that there would be fewer or no content questions involving writing and balancing equations. At the same time, we expected content questions for the prompts involving predicting, sketching a graph, and explaining reasoning. Interestingly, there were still many content questions asked about the 'familiar concepts', providing evidence for the need to review relevant material covered in prior courses.

Some of the prompts in the collaborative activity required mere recall while others required conceptual understanding. The classification of the content questions revealed that most of the student content questions were at the lower levels of Bloom's taxonomy. As we anticipated, all the content questions posed while the students were actively writing and balancing equations were at the remember level of Bloom's taxonomy.

However, contrary to what we expected, the prompts that asked students to predict, create, and explain elicited more questions at the lower levels of Bloom's taxonomy than at the higher levels. The right kinds of questions could lead students to appropriate discussions, which in turn could lead them towards correct responses. When students ask lower-level questions, this may be a sign of a lack of conceptual understanding. It also provides further evidence that spontaneously generated student questions often tend to be lower-level questions. This finding lends support to the notion that there is a link between the level or type of question asked and the possible level of knowledge construction that occurs (Chin & Osborne, 2008).

Students were not explicitly instructed to ask each other questions. It is therefore encouraging to see that even though students were not required to ask each other questions, they did, and that these peer-to-peer questions played a role during the collaborative activity. For a number of students, their contributions were phrased as questions. While we cannot infer intent, the questions allowed students to run their ideas by their peers to confirm accuracy. In this way, these students were able to both contribute to the group and check their own ideas. Student statements phrased as questions could also be a form of self-questioning, which is considered to be a metacognitive skill, helping students engage in metacognition (Halmo et al., 2022).

Peer-to-peer questions were instrumental in promoting social metacognition. Two metacognitive skills evident in the student conversations were monitoring and evaluation

(Anderson, 2004). In the example excerpt shown above (see Section 5.3.6), student questions such as 'does that make sense'; 'does that look okay'; and 'did we do that correctly' allowed them to examine their thinking and their resulting solutions. At the same time, such questions also helped ensure there was consensus within the group, particularly when students asked questions such as 'are we on the same page' or 'do we all agree'.

7. Implications for Research and Teaching

Our results add to what is already known in science education about the benefits of student-centered instructional strategies such as collaborative learning. As part of the problem-solving process within their collaborative groups, the students asked each other questions that ultimately helped them complete the assigned activity. This occurred in the social setting created through the assigned activity. The focus on student peer-to-peer questions illuminated the fact that questions are a way through which students engage with both the assigned task and each other. Unlike in lecture settings, where students seldom ask questions, our study shows that the students in a group activity were engaged in asking many peer-to-peer questions. These peer-to-peer questions were beneficial because they not only supported student problem-solving efforts but also supported student social metacognition. Thus, for these benefits to be realized, it is necessary that we create environments that support students working together.

While our results showed that most students asked their peers a question during the activity, there were groups in which some students did not ask any questions. Furthermore, there was a wide range of questions asked within the various groups (2–44 questions), with an average of 15 questions per group. This leaves us to wonder why in one group only two questions were asked while in another 44 questions were asked.

Future research should therefore explore the causes for this variation in the number of questions posed. Are these specific results due to demographic factors such as gender, societal or classroom culture, group composition, student major, or year in college? Furthermore, in this study students were instructed to work with those who sat next to them. The groups were not pre-assigned; nevertheless, the students could not choose their group from the entire class roster. Could this factor have had any impact on the questioning behavior within the groups? How would the number and type of questions change across a semester if students consistently worked within the same group of students? How would the results differ if students were first instructed about the different types of questions and levels of Bloom's taxonomy before engaging in group work? Would it only take one instructional period about how to engage in questioning, or would it require multiple instructional periods throughout the semester to enhance student questioning behavior? Each of these questions warrants future studies in order to better understand how to best support students posing effective peer-to-peer questions.

We did not anticipate students having many content questions about writing and balancing equations because this is covered both in high school chemistry classes and first-semester general chemistry. Moreover, these content questions were at the lowest level of Bloom's taxonomy. Therefore, another area of future research could be whether students pose higher-level questions when solving familiar tasks or whether familiar tasks only elicit recall questions. It is interesting to note that students were most successful in these tasks and yet they had questions about the prompts' subject matter. As noted above, student questions can be a source of feedback for teachers (Woodward, 1992). One key takeaway from our results is that instructors should review relevant prior course material to ensure that all students have the fundamental knowledge to be successful in the course.

The results from our study revealed that students struggled to answer questions that required an explanation of their predictions, the sketching of a graph, or an explanation of why they drew their graphs the way they did. One major possibility to explain this observation is a lack of conceptual understanding of acid–base neutralization and conductometry. The assigned activity required students to transfer their knowledge of acid–base neutralization to the context of conductometry. Thus, the lack of conceptual understanding made this transfer difficult. This can also explain why there were very few peer-to-peer questions from the higher levels of Bloom's taxonomy. Future research could systematically study the connection between student conceptual understanding and the nature of student questions.

Our results showed that most of the posed content questions were in the two lowest levels of Bloom's taxonomy. Furthermore, we observed that most student responses to prompts that required higher-order thinking were incorrect, as shown in Figure 2 above. Did the fact that most of the content questions posed while answering these prompts were at the knowledge and remembering levels affect their ability to come up with correct responses? We hypothesize that when students pose questions at higher levels of Bloom's taxonomy, this should help them construct knowledge at higher levels of Bloom's taxonomy. Therefore, a future study should use Bloom's taxonomy to explore whether there is a causal relationship between the levels of the questions posed and the resulting correctness of the group responses.

Author Contributions: Conceptualization: T.M.D., K.G. and J.N.; Methodology: T.M.D., K.G., S.M. and J.N.; Transcribing T.M.D., K.G., S.M. and J.N.; Formal analysis: T.M.D., K.G., S.M. and J.N.; Writing—original draft preparation T.M.D., K.G., S.M. and J.N.; Writing—review and editing: T.M.D., K.G. and J.N.; Administration and supervision: J.N. All authors have read and agreed to the published version of the manuscript.

Funding: Funding to support Grieger was provided by the National Science Foundation (NSF funding number 2411805). Funding to support Miller's summer REU research was provided by the National Science Foundation (NSF funding number DUE1560142 and DUE 1852045).

Institutional Review Board Statement: The study was approved under Protocol No. IRB0004625 by the Institutional Review Board of North Dakota State University on 9 January 2023.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are not available due to privacy restrictions.

Acknowledgments: This study was supported in part by the National Science Foundation (2411805, DUE1560142, and DUE 1852045). The authors wish to thank the students who participated in this project and Kristina Caton from the NDSU Center for Writers for her help with revising the manuscript for clarity.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Adams, P. (2006). Exploring social constructivism: Theories and practicalities. *Education*, *34*(3), 243–257. [CrossRef]

Anderson, N. J. (2004). Metacognitive reading strategy awareness of ESL and EFL learners. *The CATESOL Journal*, *16*(1), 11–27. [CrossRef]

Biddulph, F., & Osborne, R. (1982). *Some issues relating to children's questions and explanations: Working Paper no. 106*. University of Waikato.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2002). *Working inside the black box: Assessment for learning in the classroom*. King's College London.

Bloom, B. S. (1956). *Bloom's taxonomy of educational objectives, handbook 1: Cognitive domain*. David McKay Company.

Blosser, P. E. (2000). *How to ask the right questions*. NSTA Press.

Bybee, R. W. (2000). Teaching science as inquiry. In J. Minstrell, & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 20–46). American Association for the Advancement of Science.

Carner, R. L. (1963). Levels of questioning. *Education*, *83*, 546–550.

Chin, C., Brown, D. E., & Bruce, B. C. (2002). Student-generated questions: A meaningful aspect of learning in science. *International Journal of Science Education*, *24*(5), 521–549. [CrossRef]

Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. *Studies in Science Education*, *44*(1), 1–39. [CrossRef]

Chiu, M. M., & Kuo, S. W. (2009). From metacognition to social metacognition: Similarities, differences and learning. *Journal of Education Research*, *3*(4), 1–19.

Crawford, T., Kelly, G. J., & Brown, C. (2000). Ways of knowing beyond facts and laws of science: An ethnographic investigation of student engagement in scientific practices. *Journal of Research in Science Teaching*, *37*(3), 237–258. [CrossRef]

Dogan, F., & Yucel-Toy, B. (2022). Students' question asking process: A model based on the perceptions of elementary school students and teachers. *Asia Pacific Journal of Education*, *42*(4), 786–801. [CrossRef]

Erdogan, I. (2017). Turkish elementary students' classroom discourse: Effects of structured and guided inquiry experiences that stimulate student questions and curiosity. *International Journal of Environmental and Science Education*, *12*(5), 1111–1137.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. [CrossRef]

Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, *49*(2), 193–223. [CrossRef]

Hakkarainen, K. (2003). Progressive inquiry in a computer-supported biology class. *Journal of Research in Science Teaching*, *40*(10), 1072–1088. [CrossRef]

Halmo, S. M., Bremers, E. K., Fuller, S., & Stanton, J. D. (2022). "Oh, that makes sense": Social metacognition in small-group problem solving. *CBE—Life Sciences Education*, *21*(3), ar58. [CrossRef] [PubMed]

Harper, K. A., Etkina, E., & Lin, Y. (2003). Encouraging and analyzing student questions in a large physics course: Meaningful patterns for instructors. *Journal of Research in Science Teaching*, *40*(8), 776–791. [CrossRef]

Lai, M., & Law, N. (2013). Questioning and the quality of knowledge constructed in a CSCL context: A study on two grade-levels of students. *Instructional Science*, *41*(3), 597–620. [CrossRef]

Marbach-Ad, G., & Sokolove, P. G. (2000). Can undergraduate biology students learn to ask higher level questions? *Journal of Research in Science Teaching*, *37*(8), 854–870. [CrossRef]

Miller, R., & Osborne, J. (1988). *Beyond 2000: Science education for the future*. King's College London.

Naibert, N., Vaughan, E. B., Lamberson, K. M., & Barbera, J. (2022). Exploring Student Perceptions of Behavioral, Cognitive, and Emotional Engagement at the Activity Level in General Chemistry. *Journal of Chemical Education*, *99*(3), 1358–1367. [CrossRef]

NGSS Lead States. (2013). Next generation science standards: For states, by states. In *Next generation science standards: For states, by states*. National Academies Press. [CrossRef]

Nyachwaya, J. M. (2016). General chemistry students' conceptual understanding and language fluency: Acid–base neutralization and conductometry. *Chemistry Education Research and Practice*, *17*(3), 509–522. [CrossRef]

Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, *35*(2), 135–198. [CrossRef]

Penuel, W. R., Yarnall, L., Koch, M., & Roschelle, J. (2004). Meeting teachers in the middle: Designing handheld computer-supported activities to improve student questioning. In *Embracing diversity in the learning sciences* (pp. 404–409). Routledge.

Powell, K. C., & Kalina, C. J. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. *Education*, *130*(2), 241–250.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*(2), 181–221. [CrossRef]

Smith, C. L., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction*, *18*(3), 349–422. [CrossRef]

Tiffany, G., Grieger, K., Johnson, K., & Nyachwaya, J. (2023). Characterizing students' peer–peer questions: Frequency, nature, responses and learning. *Chemistry Education Research and Practice*, *24*(3), 852–867. [CrossRef]

Tracy, S. J. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. Wiley-Blackwell.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning*, *4*(3), 259–287. [CrossRef] [PubMed]

Van De Bogart, K. L., Dounas-Frazer, D. R., Lewandowski, H. J., & Stetzer, M. R. (2017). Investigating the role of socially mediated metacognition during collaborative troubleshooting of electric circuits. *Physical Review Physics Education Research*, *13*(2), 020116. [CrossRef]

van Zee, E. H., Iwasyk, M., Kurose, A., Simpson, D., & Wild, J. (2001). Student and teacher questioning during conversations about science. *Journal of Research in Science Teaching*, *38*(2), 159–190. [CrossRef]

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Watts, M., Gould, G., & Alsop, S. (1997). Questions of understanding: Categorizing pupils' questions in science. *School Science Review*, *79*(286), 57–63.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, *79*(1), 1–28. [CrossRef]

Woodward, C. (1992). Raising and answering questions in primary science: Some considerations. *Evaluation & Research in Education*, *6*(2–3), 145–153. [CrossRef]

Yu, F.-Y. (2009). Scaffolding student-generated questions: Design and development of a customizable online learning system. *Computers in Human Behavior*, *25*(5), 1129–1138. [CrossRef]

Zeegers, Y., & Elliott, K. (2019). Who's asking the questions in classrooms? Exploring teacher practice and student engagement in generating engaging and intellectually challenging questions. *Pedagogies: An International Journal*, *14*(1), 17–32. [CrossRef]

Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9- and 10-year-olds. *Educational Technology Research and Development*, *55*(2), 117–145. [CrossRef]

education sciences

Article

Putting Inclusion into Practice: Five Commitments Toward Equity in Teaching

Brittland DeKorver

Department of Chemistry, Grand Valley State University, Allendale, MI 49401, USA; dekorveb@gvsu.edu

Abstract: Instructors make day-to-day decisions grounded in their own experiences, and this practice may be appropriate for the students who share similar experiences and backgrounds. But for students who come from a different socioeconomic status, nationality, racial or ethnic identity, gender or sexual identity, or ability status, the instructor's experiences may be insufficient to provide guidance for how to create an inclusive space for all learners. This manuscript describes interviews collected with students with disabilities regarding their experiences having their disability accommodations implemented in their courses. From these interviews and personal reflections on teaching, the author piloted and refined several teaching practices to improve the accessibility and inclusion in her own classroom. The author summarizes the principles underlying these pedagogical decisions as five commitments toward equity in teaching.

Keywords: action research; qualitative research; diversity; equity; inclusion; accessibility; disability; large lecture courses; science education

1. Introduction

You are halfway through the riveting (or so you thought) lesson that you prepared when you look over and see a student with their elbow propped on the desk, chin in hand, head lolling as they struggle to stay awake. A few minutes later, you glance in their direction again and notice they are clearly out. What do you do next?

(a) Wake them up and tell them they need to relocate because your classroom is not an appropriate place for napping.
(b) Wake them up and make a joke about it to cover up any awkwardness.
(c) Continue with the lecture without taking any action.
(d) Continue teaching, but then contact them afterward to see if everything is ok and if they were able to obtain satisfactory notes from a friend.

The course of action that you take is dependent on your goals and experiences. Humans, even professors, are logical creatures who pursue actions that they believe best align with their goals. In this article, we will assume that professors hold this goal in common, to facilitate the learning of each and every student in their classroom, and that the actions they take are in pursuit of that goal. This begs the question: How do professors choose the actions that are most likely to facilitate learning for all students?

Teachers tend to make pedagogical choices that are familiar to them, e.g., those that correspond to their experiences as students, teachers, or researchers (Oleson & Hora, 2014). For the students who have different traits, backgrounds, talents, and ability status than their teachers, those pedagogical decisions may fall short in supporting their learning. Therefore, teachers need to cultivate knowledge of students who are not like them to cultivate practices that will support those diverse students.

One potential area for mismatch between instructors' and students' experiences is in disability status: only an estimated 4% of university faculty members are disabled (Grigely, 2017), while students with disabilities make up nearly 20% of the undergraduate population (National Center for Education Statistics, 2023). Disabilities can be wide-ranging in both the way they impact a person and the degree to which a person is impacted. There are many categories of disability, such as physical, such as limb differences; cognitive, e.g., dyslexia; or emotional and behavioral, e.g., panic disorders. Disabilities may also be apparent to an observer, or they may be invisible. They may be temporary or chronic. The diversity of students' experiences with disability indicates an urgent need for faculty to learn more about various ways students may be impacted by disabilities, so that instructors are better equipped to meet students' educational needs.

When a student with a disability (SWD) matriculates, they must decide whether to disclose their disability(ies), navigate the process for receiving accommodations, and then implement those accommodations throughout their academic career. While there are several studies that explore the students' experiences throughout the first steps in the process (Ben-Simon et al., 2008; Cole & Cawthon, 2015; Harrison et al., 2013; Lightner et al., 2012; Lindstrom, 2007; Lovett et al., 2015; Salzer et al., 2008; Stevenson, 2010; Storrie et al., 2010), and the resulting academic outcomes for those students (Abreu et al., 2016; Blasey et al., 2023; Hollins & Foley, 2013; Nielsen, 2001), there is not much information about the interactions between the student and their instructors regarding the specific implementation of the accommodations. Therefore, the research questions guiding this article are as follows:

1. What are the experiences of students who try to implement their accommodations in individual classroom environments?
2. How can knowledge of student experiences inform and transform teaching practices?

2. Materials and Methods

The first research question was investigated by conducting interviews with SWDs according to a protocol designated as exempt by Grand Valley State University's Institutional Review Board (22-305-H-XXXX). The student participants were recruited by email through the university's Disability Support Resources (DSR) office at a large, public, primarily undergraduate institution in the Midwestern United States. Four students accepted the invitation to participate in semi-structured interviews ranging from 50–80 min in duration (Table 1). Interviews were conducted by video conference and transcripts were refashioned with narrative analysis and the resulting narratives were validated by member-checking with the interview participants (DeKorver et al., 2024). Themes from the narratives were discussed among researchers DeKorver, Brown, and Witcher, who brought a diversity of perspectives to the analysis. Brown was a former student of DeKorver, who upon learning about some of the struggles his peers with disabilities faced, approached DeKorver about the possibility of conducting research in the area. As an undergraduate researcher, Brown was the primary interviewer, as it was determined that a peer would be more well suited to elicit student experiences. Brown was also able to provide valuable perspective throughout the analysis, which heavily relied on interpreting the dynamics of student–instructor relationships. Brown and DeKorver recruited Witcher, the director of the Disability Services Resource center, to complete the research team. Witcher's role was to provide additional context and perspective on the process of student accommodations, as well as keeping the research findings grounded in the experiences of people with disabilities and minimizing unintended ableism and tokenism.

Table 1. Summary of SWDs interviewed about their experiences having accommodations implemented in their courses.

Pseudonym	Demographics	Necessary Accommodations
Lacey (she/her)	STEM major in her 6th year	extended time on assessments and student note-taker
Katie (she/her)	STEM major in her 4th year	extended time on assessments and student note-taker
Deb (she/her)	STEM major in her 3rd year	extended time on assessments, recorded lectures, instructor's slides in advance
Pam (she/her)	Humanities major in her 3rd year	closed captioned lectures

The data from this study were analyzed from a social model of disability (Shakespeare, 2006), which challenges the notion of disability as a personal characteristic and places the responsibility of dysfunction on existing social structures instead of on the individual. This suggests that university and course structures should function for students of all abilities, without the need for individual accommodations. However, DeKorver reflected that her course roster regularly included 5–10 students each semester who required accommodations to participate in her classes. This led DeKorver to scrutinize her own teaching practices through the social model of disability, documented through reflective memoing. These writings were inspired by critical reflection, where instructors examine their pedagogical practices through the lens of ethical considerations such as equity and inclusion, but also consisted heavily of deliberative reflection, weighing the practical considerations of instructor, student, and institutional needs (Larrivee, 2008; Valli, 1997). Her goal was to restructure her course policies to create greater equity in student learning outcomes and provide a more inclusive learning environment. This desire coincided with the exigencies of revising courses to respond to the COVID-19 pandemic, allowing much more freedom, motivation, and opportunities to implement drastic changes in course structure.

The courses undergoing these changes consisted of four different 100-level chemistry courses. Three of the four courses were a combination of lecture and laboratory sessions; the courses ranged from three to five credits (Table 2). Each class section enrolled up to 90 students, primarily students majoring in STEM fields, and [Redacted] taught some combination of two sections of these courses each semester as these changes were implemented during a time frame spanning from 2020 to 2024. After the initial changes were implemented, course data were reviewed to evaluate and inform further refinements. These data included end-of-term grades, mid-term and end-of-term student evaluations of teaching, degree and frequency of student accommodation requests, subsequent course grade outcomes were reviewed, and the experiences of faculty teaching the courses. Adjustments to course policies and the implementation of new policies continued in subsequent semesters based on the most recent data according to principles of action research (McNiff & Whitehead, 2002). Because this research was grounded in personal reflections and decision making, the results and discussion are reported in the first-person voice.

Table 2. Courses included in action research portion of this research.

Course Title	Number of Course Credits	Maximum Student Enrollment Per Section	Curricular Requirement for Majors and Programs
Preparatory Chemistry	3	20	not required
Introductory Chemistry	4	96	nursing, wildlife biology, allied health science, medical laboratory science, physical science general education
Principles of Chemistry I	4	66	chemistry, biology, pre-med, biomedical sciences, engineers, geology, physical science general education
Principles of Chemistry II	5	52	chemistry, biology, pre-med, biomedical sciences

Limitations

The interview data were collected from a relatively small number of participants. The experiences of this subset of participants cannot encompass the vast diversity of disabilities that impact students' education. In addition, all participants had similar racial, ethnic, and gender identifications, which further limits the diversity of the sample. As a student's other identities intersect with their identity as a person with a disability, this may lead to an even greater diversity of experiences that are not captured among these data.

3. Results

3.1. Classroom Engagement

Many instructors regard taking notes as quintessential to a student's engagement in the classroom (Wong & Chiu, 2021). Yet there are a variety of disabilities that act as a barrier to participating in this activity for students. Kate described her efforts to join in the practice of note-taking: "I take notes on literally everything because I can't tell what's important or not. I'm not very good with tone". It was difficult for Kate to interpret which points during her professors' lectures were salient, so try as she might, her own note-taking was not contributing to her learning. Kate's accommodations stipulated that the professor should recruit a note-taker to provide her with class notes. (Other institutions may have different policies and practices about providing accommodations but at this institution, the instructor is tasked with recruiting a note-taker). While this accommodation was useful when it was carried out, Kate found it difficult to have this implemented in all of her classes. Lacey faced a similar struggle in having her note-taking accommodation met, although her note-taking obstacle was related to a physical disability rather than cognitive.

> *Lacey: I generally have had trouble with receiving a note-taker. Professors ask for volunteers to take notes, but often no one volunteers. I think maybe because it's unclear what the notetaker does, and no one knows what to do next.*

Lacey resorted to asking her friends to take notes or taking photographs of the classroom projector screen and described these measures as inadequate. Deb, who did not have note-taking accommodations, attempted to aid her note-taking by taking photographs during a class and expressed her frustration at the outcome.

> *Deb: One time a professor called me out for taking a photo of a slide during class. "Hey, no taking pictures!" I had stitches in my right hand and couldn't write with it. I didn't have an accommodation for that because it was a temporary injury. I tried to write with my left hand, but was struggling to write legibly. It makes sense why they don't want us to take pictures, but I think sometimes students just struggle to write everything down. If they don't want us to take pictures, is there something else that they can do to help us? It would be nice if they could meet with us in office hours and find a way to make it work so that we don't break the rules or inconvenience them, but we still get what we need.*

Deb wished that her professors would provide a way for all students to have access to notes. Lacey, who was unfamiliar with Deb's account, made the same point.

> *Lacey: Even when the professor sends me all the slides, it's not as good as having notes, and when it comes time to homework or studying, I don't have adequate reference material to guide me. I think it would help if professors made their notes more available, for all students regardless of their DSR status.*

Kate had a professor who did make the notes, taken by a volunteer student note-taker, available to the entire class, saying, "She had a student anonymously take notes for us, and the notes were available to everyone. So not just me, not just another student who has a note-taking DSR memo, but everyone. And that was really helpful". The fourth student participant in this study, Pam, also made this recommendation, "I would rather that professors just have [lecture notes and slides provided] as part of the normal routine. It definitely helps me, but it would help a lot of other students, too". Even though these students had drastically different disabilities, each was limited by their note-taking and by the lack of accommodation, and each felt that note-taking should be more accessible for their able-bodied classmates, as well. I agreed with the points these students made and felt that I should be doing something to support note-taking, but I was not sure how to implement it.

My initial instinct was to de-emphasize note-taking in my classroom. Many instructors eschew lectures in favor of active-learning pedagogies. In these, the course is structured so that students frequently work together in small groups (e.g., flipped classrooms (Seery, 2015), cooperative learning (Bowen, 2000), and POGIL (Moog & Spencer, 2008)) rather than sitting and taking notes. Yet these forms of engagement are also not possible for some students with disabilities without accommodations or additional scaffolding. Students with clinical anxiety, an increasingly common condition among college students, may find their symptoms exacerbated by these sorts of activities (Cooper et al., 2018). Although the participants in this study did not specifically talk about group work, they did talk about their perceptions of their peers. Kate, who was diagnosed with autism, said she felt awkward around others, and that sometimes she perceived negative attention from her classmates: "it feels like everyone in that room knows she's got something going on up there". She also related that in one particular class, a group of classmates "made fun of me" and "kept a tally" of her interactions with the professor. Mandating that all students work with their peers and grading them on these efforts creates a situation that exacerbates existing inequities.

Furthermore, my reflections revealed that students had to be physically present to receive credit for participating in the in-class group activities. Classroom response systems or "clicker" are a popular tool for ensuring attendance and participation in active learning classrooms (Gibbons et al., 2017); I either used clickers or group worksheets to document students' attendance in my lectures. I found I was increasingly uneasy about awarding credit for in-class activities knowing that some of the students who provided DSR memos needed accommodations for frequent absences. I began to consider alternatives so that my course grades were less dependent on physical presence during scheduled class sessions.

I was able to address the accessibility of all three trouble areas (note-taking, group problem solving, and in-class attendance) by implementing one new course policy. I created an assignment category called "Community Learning" to replace the grades students had previously earned through group work and attendance. Community learning (CL) occurs throughout the semester and students are able to choose how they want to engage in CL. They may decide that they will share their notes on the class discussion board, participate in in-class discussions, form a study group, or provide a summary of class for a classmate who was absent from a session. They could create practice questions and post them to the class website. They could ask or answer questions on the class discussion board. I even encouraged them to facilitate virtual attendance for classmates who could not be physically present but still wanted to engage via video or teleconferencing.

Activities are valued as CL when they (1) are accessible to other classmates, (2) require sustained effort, (3) are a unique contribution, and (4) impact the student's own learning. Not every activity scores highly on all criteria. Posting notes for all to see is very accessible, but it likely would not impact the student's learning as much as hosting a review session for a small study group (which would rate as less accessible). Students assess themselves eight times throughout the semester on a rubric (see Figure 1) to earn points. The number of points required to earn full credit for this assignment category is much smaller than the number of points possible; that is, students can earn 100% of the points in this category even if they do not earn 100% each time they self-assess. During the first semester of implementation, I learned that students were surprisingly honest in their self-appraisals, and they chose a great variety of ways to contribute to CL. There was a healthy level of activity on the course discussion board, a wide variety of notes were promptly available after every class session, and there were robust study groups. I felt like this solution balanced my desire for students to practice communicating with others to build knowledge with the students' needs for autonomy and accessibility.

My community learning for this Learning Objective...	**Rating** **0: not true** **1: somewhat true** **2: very true**
...was accessible to a wide variety of my peers.	
...provided a unique contribution to the class.	
...improved my knowledge of the LO.	
...required effort throughout the duration of the LO.	
total	

Figure 1. The rubric for community learning as it appears in the course assignments.

3.2. Assessments

Perhaps unsurprisingly, the participants placed the greatest emphasis on how their disabilities impacted their completion of tests and quizzes, the portion of their classes that most heavily impacted their grades. Their accommodations for testing included stipulations for extended time, a quiet testing environment, and the ability to wear headphones. Unfortunately, the enactment of those accommodations sometimes negatively affected their ability to participate in other typical class activities.

> *Katie: Even though I got [testing accommodations] eventually, my professor made it a hassle. After that I interacted with the professor differently, I usually go to office hours. I love them. I live off them. But after that, I never went to that professor's office hours. I felt more comfortable missing class and I didn't really want to go. I lost a little respect for the professor.*

> *Lacey: Early in my college career, I had one professor who was very reluctant to give me extended time on tests. The class began with a quiz followed by lecture. She would have me leave the class to take the quiz, which cut into my class time, and made me miss the start of lecture.*

Katie and Lacey both missed out on other expected course activities due to the way their accommodations were implemented.

Some students' accommodations allow them to take the exam in a quiet area with minimal environmental distractions. These students are still able to choose this option by scheduling a proctored exam with our DSR office. Other students, preferring to remain in proximity to the instructor, rely on technology such as headphones to minimize distractions. Katie was one such student.

> *Katie: The only time I've felt singled out truly though was when I wasn't allowed to wear AirPods, and I have AirPods because: One, they are noise canceling which is fantastic; Two, they're really really low key, so you can fit them in your ear, and no one knows. I had a professor that was really uncomfortable with it because they're Bluetooth and they connect to my phone. So, I have a second set of headphones and it looks like I'm working in a construction yard when I wear them. They are huge. . .It's the only time where I truly felt singled out when it comes to accommodations.*

Even though Katie's accommodations specifically permitted her to use headphones, her instructor placed restrictions on the type of headphones she could use, making her self-conscious about her use of the accommodation.

Deb's account illustrated the challenges for students who have anxiety, but for whatever reason, do not have accommodations for it.

> *Deb: That first semester I had a panic attack during my exam. I had been getting a B in that class, but I failed the exam because I only filled in three answer bubbles, and there was no option to retake it. The professor couldn't do anything for me they said, because I did not have those accommodations. I might have passed that class the first time around if I had a DSR memo. It's just helpful to have a back-up plan so that the anxiety doesn't keep building as you try to figure out what to do about it.*

When students suffer from test anxiety, they are unable to demonstrate their knowledge. Yet, course assessments purport to be an accurate measurement of students' knowledge. If an analytical chemist discovers there is an environmental factor producing interference in their measurements, they do everything they can to reduce or eliminate that factor in order to make their measurements more valid, i.e., the measurements are only influenced by the factor(s) that they purport to measure. Therefore, to maximize the validity of our assessments, we should be doing all we can to minimize interference from student anxiety.

From the instructor's standpoint, we have to contend with the reality of physical limitations of time and space when meeting accommodations for assessments. My strategy to address this conflict was to radically revise my assessment strategy. I began by shortening my assessments to only one question for each stated learning objective. My mid-term exams went from 25 to 40 questions for a 50 min testing period to 3–15 questions (depending on the level of granularity of the learning objective and the type of assessment item). This meant that most students completed the exams within more than half of the time remaining in the class. This also meant that for the last 30 min of the exam period, the remaining students had a silent, nearly empty lecture hall in which to complete their work. Many students preferred to remain working for the entire 50 min period. I discovered that a few students wanted even longer to work, and because I taught two back-to-back sections of the same class in the same room, I was able to facilitate this. Students who wanted more than 50 min could begin their test in the first section and remain through the second section. One time, a student wanted additional time on a particular assessment, but their course schedule did not permit them to attend both sections; we were able to work out a different, mutually agreeable arrangement. I went from scheduling exam accommodations for 5–10 students at each exam to one instance of accommodation that entire semester. In addition to the convenience these exam modifications afforded me as an instructor in nearly eliminating the need for individual accommodations, students appreciated the emotional benefits of having extra time, expressing that it relieved their symptoms of test anxiety.

In addition to mental health issues, there are other chronic medical conditions that might be interfering with my students' ability to take exams. For example, I have taught students who were dealing with chronic migraines, Chron's disease, endometriosis, and ankylosing spondylitis. I did not want to assign a score of 0 if their symptoms prevented them from attending a scheduled exam; I wanted their score to be an indicator of their learning rather than their health. Yet in a large lecture course, scheduling make up exams would be prohibitively challenging. To solve this issue, I moved to a standard-based grading format, where each exam is given multiple times throughout the semester and students can earn full credit even if they are absent (or fail) the first attempts (Talbert & Clark, 2023). The shorter format of my exams meant that writing and grading multiple iterations of the exams did not impose additional hardship on me. Implementing these additional assessments in person does require extra course time; I made room for this in the course schedule by recording videos of some of the material that I would have covered by an in-class lecture. These videos were assigned to students to view as homework in the course learning management system, similar to the distribution of content in flipped instructional methods (Seery, 2015).

Another way to make classroom assessments more valid would be to make sure they ask students to demonstrate the knowledge and skills expected of practicing scientists (Stowe & Cooper, 2019). The primary mode of assessment for burgeoning scientists is through summative exams, yet practicing scientists do not take written exams as part of their professional activities. This means that we are evaluating students' science knowledge and skills by a metric that is not authentic to the work of real world science—again, a threat to the validity of our assessments. Katie recognized this weakness.

> *Katie: "Another one of my math classes was based on portfolio style homework. So this portfolio style way of doing things had no quizzes, tests, or things like that, and the homework is optional, only turned in for feedback. We do check-ins. Like "How's it going? What do you need from us that you need to be successful? What can we do to help you?" I appreciate every bit of that. What they "grade" if you will, is the portfolio you create of all these math problems that you do. I knew early that I didn't need my DSR memo because it wasn't that kind of class. At the end of the semester, I asked for an A and I got an A, because my portfolio was good. It was the professor's philosophy that this method of teaching better prepared us for the real world. There are no timed tests and graded assignments in the real workforce. There are projects and problems you work on and receive feedback. You should be prepared for that".*

In Katie's math class, the portfolio assessment allowed her to showcase her understanding, as well as developing the "soft skills" of working on a project throughout the semester, making continuous revisions after feedback. This flexibility that this type of assessment allows is much more accessible and, as Katie points out, is more aligned with the types of tasks students will be expected to perform in a professional setting. Unfortunately, I have not reached that level of authentic assessment in my own teaching, although it remains one of my goals as I continue to revise my teaching methods and materials. One way I have sought to make the evaluation process more authentic while remaining in the framework of exams is by writing assessment items that require students to use their knowledge to engage in science practices such as drawing conclusions from data or using models (Stowe & Cooper, 2019). Another strategy I have employed is to allow students to use reference material on their assessment, ranging from a relatively paltry notecard measuring 3 inches by 5 inches to including fully open notes, open textbook, or even allowing them to use online resources. After all, practicing scientists are generally permitted to use references in order to solve the tasks set before them.

Additional details about the course policies I have enacted are provided in the Supplemental Information. Table 3 summarizes the key classroom policies that I revised in the semesters following the transition to remote teaching in 2020 and the subsequent transition back to face-to-face instruction. Over this period, the proportion of my students formally disclosing their disabilities has dwindled to about 1% each semester, and these remaining students have indicated they were disclosing their disability status in an effort to provide me with information, but did not need me to take any further action to meet their necessary accommodations.

Table 3. Comparison of selected classroom characteristics for general chemistry, before and after implementing commitments to inclusive teaching.

Classroom Policies	Previous Implementation	Revised Implementation
Assessment	3–4 mid-term exams per semester, 25–35 questions, multiple choice and short answer, limited to the time of the class period, closed-note, individual effort.	10 mid-term exams per semester, 3–15 questions, multiple choice and short answer, flexible time, hand-written reference sheet is permitted, individual effort.
Attendance	Attendance is included in grade calculations and assessed by clicker or by group work participation. Allowances for occasional absences are made by setting some threshold (e.g., 80–90% attendance) to earn full credit for this grade.	Attendance is not graded. Students are encouraged to provide individualized support for absent peers as part of their community learning. Virtual attendance, facilitated by classmates, is permitted.
Group work	Students complete group problem solving exercises for credit during class sessions. Lowest in-class assignment scores are dropped from the grade calculation at the end of term.	Students engage in solving practice problems during class, but they can choose whether to participate with classmates. Participation in small groups during class or study groups external to class sessions can be used to earn credit for community learning.
Note-taking	No broad policy implemented; note-takers recruited as needed per accommodation memos.	Students are encouraged to share their personal notes on the course website as a way to engage in community learning.
Student use of technology	No specific policy.	Students are encouraged to bring devices to class to use online resources and facilitate virtual attendance. Devices are prohibited on assessments, with the exception of headphones, which students may use to reduce environmental distractions.

4. Discussion

Providing accommodations for students who have documented and disclosed disabilities is necessary to achieve equity, but it is not sufficient. One participant pointed out the following.

> *Deb: I have an appointment in a couple weeks in August that should help me get that ADHD diagnosis and the auditory processing diagnosis, but I know my access to healthcare is like a privilege so a lot of students who can't afford it, they are not getting the accommodations they need to succeed in those classes.*

Deb's comment highlights the problem with providing accommodations only for students with documented disabilities. I wondered how many of my students had undiagnosed chronic illnesses, were not able to access healthcare in order to complete the disability support bureaucracy, had other barriers to official accommodations, or did not even know that accommodations were available.

Deb: I didn't meet up with anyone in DSR during my first semester in college. I didn't really know that was something that I could access with anxiety, depression, or bipolar. I thought that was just for physical disabilities that you could see. My disabilities are invisible, relating to my digestive system and mental health.

I knew that the students who struggled to obtain necessary accommodations might have their difficulties compounded by additional factors, such as first-generation status, low socioeconomic status, or membership in other disadvantaged groups.

The idea that classroom materials and methods should be accessible for all students, not just those with a registered disability, is encompassed by the Universal Design for Learning (UDL) movement (CAST, 2018). The goal of the UDL framework is "learner agency that is purposeful & reflective, resourceful & authentic, strategic & action oriented". Despite the lack of explicit reference to inclusion, the UDL framework offers actionable guidelines for instructors who wish to design their course to be accessible for all learners. The guidelines are categorized according to three principles: engagement, representation, and action and expression. Each principle has themes, and each theme has recommendations. For example, in representation, one of the design options pertains to "Language and Symbols", with specific recommendations for ensuring multiple forms of representation are available and using respectful language in the classroom. In my own teaching, these recommendations were manifested as curricular changes such as ensuring my materials have appropriate captioning and are accessible by screen readers, and cultural changes such as replacing "you guys" with "you all" in my daily speech.

The UDL guidelines were extremely useful to me while thinking about how to make my teaching more inclusive, but from the interviews with students, I knew that implementation of inclusive policies was not enough. I needed to explicitly address inclusion with my students. The following are Deb and Katie's words.

Deb: I have had some professors who made it clear they would accommodate you if you struggled. They made a note in the syllabus about their willingness to help, even with things like mental health and well-being. They made it clear they were there for their students. They gave details about how to communicate them about our DSR memo... When they show that they are ready to take my memo, I can be confident that they will make those accommodations. I've appreciated those professors so much, because it just gave reassurance that I wasn't going to be on my own.

Katie: One professor did a great job taking away that feeling of being a burden. On the first day of class, she mentioned that any student who needs accommodations and has a DSR memo should let her know and she'd work with us to get it set up... That announcement let me know she would be willing to work with me, and was happy to work with me. I was glad that professor was so transparent about being willing to help. It made it easier to deliver my DSR memo to her.

The students with disabilities appreciated professors who explicitly stated their desire to be inclusive. Providing a written statement about my values regarding inclusion would not only provide confidence to my students, but it would also give me a way to explain to them (and to my colleagues) why and how I had arrived at the particular instructional choices I have made. Finally, by enumerating my philosophy regarding inclusion, I would be better able to engage in honest self-appraisal and be held accountable by students and colleagues. With these benefits in mind, I engaged in deliberate reflections about inclusion, why I pursue it in my teaching and how I see it implemented in my classrooms. Those reflections and the pedagogical choices I have made because of those reflections became the basis for this manuscript. Another product of the reflections was a set of explicit principles that I have called my "5 Commitments to Inclusion". I first drafted these commitments while attending an academic conference to summarize my reflections on a series of keynote speeches about racial equity in education. They were further informed by and refined for application to my teaching, and then broadened so that they could apply to all facets of my daily life. The version below (Figure 2) was written in August of 2022 for inclusion in my syllabus for the fall semester and has appeared in all of my course syllabi since then. Specific people and works that helped to shape individual commitments are cited in the following paragraphs.

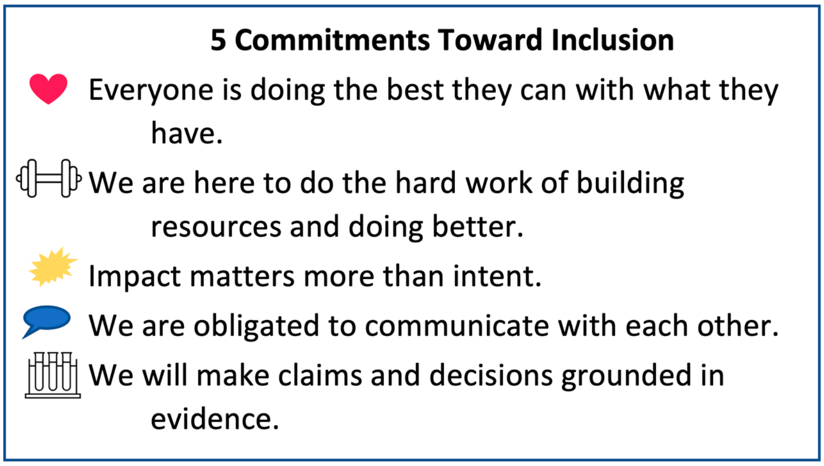


Figure 2. The five commitments as they appear in a course syllabus.

4.1. Everyone Is Doing the Best They Can with What They Have

This commitment is borrowed from the Collaborative Proactive Solutions parenting philosophy (Greene, 2017). Although teaching adult students differs from parenting children, it does require a similar level of compassion and empathy, and this mantra has helped me tremendously in both roles. Putting this commitment into practice requires reframing my beliefs about students' participation in my class. My students are not procrastinating or lazy; they are making choices based on their priorities—many of which are unknown to me. My students are not ignoring my helpful study reminders; they have barriers to carrying out the task or I haven't provided enough support yet. When students disparage themselves as "not getting it" or "terrible at chemistry", I remind them, "You're doing the best you can right now". This mantra was picked up by my students: in one class session, I was attempting to write with an old dry erase before rejecting it and muttering, "This marker is terrible". A student quipped, "It's doing the best it can!" to the laughter of their classmates. Greene, the creator of the parenting method, refers to this as a "lens shift". When we embrace the notion that the people around us are doing the best that they can, it primes us to respond with empathy and care.

4.2. We Are Here to Do the Hard Work of Developing Our Resources and Doing Better

Recognizing that students are giving me their best does not mean that I cannot set high expectations or ask them for more. My second commitment provides the rationale for our mutual engagement in our course. As the instructor, I am there to do the hard work of providing resources and improving my teaching. Students are there to do the hard work of learning chemistry and developing metacognitive skills. This commitment paves the way for introducing students to a growth mindset, goal setting, and other metacognitive strategies (McGuire, 2015). Orienting our mutual hard work as growth toward our goals, also allows me to empathize with students who are struggling, while reassuring them that the frustration they feel is an indication of the learning process (Bjork et al., 2011).

4.3. Impact Matters More than Intent

The saying that "intent is not equal to impact" is often used in a social justice context by activists to call people with privilege to consider their actions from the perspective of marginalized people (DiAngelo, 2018). This credo was applied in education to help instructors understand how their assumptions might be harming their students and to guide instructors on facilitating classroom discourse on sensitive subjects or when instructors need to mediate microaggressive events in their classroom (Meadows & Wickner, 2020). But this statement also affords an opportunity to extend the metacognitive practices and growth mindset of commitment two. Working hard or doing your best is not a guarantee that goals will be met. No matter how excited I am or how much time I spend on reforming my curriculum, if the students are not learning from it, I did not meet my goals. Some students may spend an inordinate amount of time and effort studying, with very little improvement on their understanding of the material. When our impact and intent are not aligned, this gives us a signal that we need to refer to commitment number two and do the hard work of revising our beliefs, changing our actions, and making a new plan to close the gap between intent and impact. What can we use to guide those changes? How will we know if our impact is misaligned with our intent? Only through communication.

4.4. We Are Obligated to Communicate with Each Other

To be an effective instructor, I must communicate adequately with my students. Presenting these commitments on my syllabus is my first strategy for communicating my expectations and motivations to my students. On the other hand, students are encouraged to communicate their needs to me and provide feedback informally via email or face-to-face, or formally through structured mid-term teaching evaluations if my efforts are not sufficient. Further, communication goes beyond providing information. It also requires receiving and processing the information. Both students and faculty should listen to each other with empathy and an open mind. This does not however mean that all communication is equally valued in my classroom.

4.5. We Present Our Claims and Make Decisions Based on Evidence

This commitment underscores my identity as a scholar. From academic matters such as assigning students' grades to social justice issues like familiarizing myself with the experiences of marginalized people to everyday decisions like choosing what to plant in my yard, I strive to collect evidence, relying on outside experts wherever necessary, and then arrive at my decisions. It also aligns with the scientific practices that I want my students to adopt in their careers (NGSS Lead States, 2013). And it supports each previous commitment: We know what "doing our best" looks like and can optimize "growth" when we rely on evidence-based practices. We can better understand "impact" when we listen to the evidence presented by those on the receiving end of our "intentions". When we communicate, we have a responsibility to provide evidence and receive new evidence, and we will engage in deliberation with the data at hand. Practicing scientists know that evidence-based decisions are also supported by theoretical models, i.e., abstract principles that are unobservable, but offer explanatory power (Windschitl et al., 2008). This final commitment, along with the previous four, offers a theoretical framework that allows me to reflect, discuss, and act on issues of inclusion and equity.

5. Conclusions

Throughout this manuscript, I have provided accounts from my research with students with disabilities, described my own reflections on accessibility and inclusion, and given examples of how these ideas were translated into practice in my own courses. You may believe that my motivation to undertake this work comes from a deep sense of justice and equity, and I do ascribe to those lofty values. However, my purpose also comes from my own experiences needing (and not having available) accommodations. Throughout my life, I have suffered from a sleep disorder. However, it was not diagnosed until adulthood. During my university days, all I knew was this: regardless of how much rest I had gotten previously, I frequently found myself drifting off to sleep. It would happen while watching television, reading a book, or sitting in class. Several times each week, I found myself dozing off during a lecture. One professor chose option "c" from the list above: he ignored my naps and continued lecturing. I know he noticed, though, because he would often make it a point to call on me to answer questions, even when my hand was not raised. I was grateful for his merciful non-response. A different professor chose option "a", requesting that the student leave. After being un-invited from his class once, I decided it would be best if I only attended on exam days. My commitments tell me this professor was doing the best he could to provide learning opportunities to his students. Yet, I wonder if I had been able to communicate my situation, or if he would have known how his brusque dismissal would impact my engagement with the course, perhaps he would have been able to do better. For instructors who teach large classes, for students who are reluctant to disclose their disability status, this kind of communication may be a barrier. For this reason, instructors should seek out accounts of people with disabilities in order to raise their general awareness of the ways that people with disabilities experience the world. Lists of books written by or about people with disabilities are a good place to start (Gettysburg College Musselman Library, 2024; Kirker, 2022). It is in this spirit of sharing experiences as evidence, forthright communication, and understanding impact and accountability in the process of reflection and growth, I present these ideas to you, my peers, with the understanding that we are all doing the best that we can.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci15010084/s1, Sample course syllabus.

Funding: This research received no external funding.

Institutional Review Board Statement: The human subject protocol was determined to be exempt by the Institutional Review Board of Grand Valley State University (22-305-H-GVSU, 21 June 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author to protect the privacy of research participants.

Acknowledgments: I am grateful to David Asai, Scott Milam, and Zakiya Wilson-Kennedy for their contributions to the public discourse about inclusive teaching. I owe a debt of gratitude to the students who participated in the research, for their bravery and honesty in sharing their stories, and to Gordon Brown and Shontaye Witcher, my research collaborators on the project. Finally, many thanks to Sarah Clark, my colleague who was willing to test and refine many of these classroom practices with me, and whose empathy for her students is a continual source of inspiration.

Conflicts of Interest: The author declares no conflicts of interest.

References

Abreu, M., Hillier, A., Frye, A., & Goldstein, J. (2016). Student experiences utilizing disability support services in a university setting. *College Student Journal*, *3*(6), 323–328.

Ben-Simon, A., Beyth-Marom, R., Inbar-Weiss, N., & Cohen, Y. (2008, September 7–12). *Regulating the diagnosis of learning disability and the provision of test accommodations in institutions of higher education*. 34th International Association for Educational Assessment Annual Conference, Cambridge, UK.

Bjork, E. L., Bjork, R., Roediger, H. L., Mcdermott, K. B., & Mcdaniel, M. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the real world: Essays illustrating fundamental contributions to society* (Vol. 2, pp. 59–68). Worth Publishers.

Blasey, J., Wang, C., & Blasey, R. (2023). Accommodation use and academic outcomes for college students with disabilities. *Psychological Reports*, *126*(4), 1891–1909. [CrossRef] [PubMed]

Bowen, C. W. (2000). A quantitative literature review of cooperative learning effects on high school and college chemistry achievement. *Journal of Chemical Education*, *77*(1), 116. [CrossRef]

CAST. (2018). *Universal design for learning guidelines*. Available online: https://udlguidelines.cast.org/ (accessed on 13 January 2025).

Cole, E. V., & Cawthon, S. W. (2015). Self-disclosure decisions of university students with learning disabilities. *Journal of Postsecondary Education and Disability*, *28*(2), 163–179.

Cooper, K. M., Downing, V. R., & Brownell, S. E. (2018). The influence of active learning practices on student anxiety in large-enrollment college science classrooms. *International Journal of STEM Education*, *5*(1), 23. [CrossRef]

DeKorver, B., Brown, G., & Witcher, S. (2024). Student xperiences of implementation of accommodations. *Journal of Disability Studies in Education 50*, 1–25. [CrossRef]

DiAngelo, R. (2018). *White fragility: Why it's so hard for white people to talk about racism*. Beacon Press.

Gettysburg College Musselman Library. (2024, November 4). *Anthologies and collections.* Diversity, Equity, Inclusion, and Belonging Resource Guide: Disabled Voices. Available online: https://libguides.gettysburg.edu/diversity/disabled-voices (accessed on 13 January 2025).

Gibbons, R. E., Laga, E. E., Leon, J., Villafañe, S. M., Stains, M., Murphy, K., & Raker, J. R. (2017). Chasm crossed? clicker use in postsecondary chemistry education. *Journal of Chemical Education*, *94*(5), 549–557. [CrossRef]

Greene, R. (2017). *Raising human beings: Creating a collaborative partnership with your child*. Scribner.

Grigely, J. (2017). *The neglected demographic: Faculty members with disabilities*. The Chronicle of Higher Education.

Harrison, J. R., Bunford, N., Evans, S. W., & Owens, J. S. (2013). Educational accommodations for students with behavioral challenges: A systematic review of the literature. *Review of Educational Research*, *83*(4), 551–597. [CrossRef]

Hollins, N., & Foley, A. R. (2013). The experiences of students with learning disabilities in a higher education virtual campus. *Educational Technology Research and Development*, *61*(4), 607–624. [CrossRef]

Kirker, K. (2022, December 7). *Winter reading list 2022: Stories from authors with disabilities.* Rick Hansen Foundation Access & Inclusion. Available online: https://www.rickhansen.com/news-stories/blog/winter-reading-list-2022-stories-authors-disabilities (accessed on 13 January 2025).

Larrivee, B. (2008). Development of a tool to assess teachers' level of reflective practice. *Reflective Practice*, *9*(3), 341–360. [CrossRef]

Lightner, K. L., Kipps-Vaughan, D., Schulte, T., & Trice, A. D. (2012). Reasons university students with a learning disability wait to seek disability services. *Journal of Postsecondary Education and Disability*, *25*(2), 145–159.

Lindstrom, J. H. (2007). Determining appropriate accommodations for postsecondary students with reading and written expression disorders. *Learning Disabilities Research & Practice*, *22*(4), 229.

Lovett, B. J., Nelson, J. M., & Lindstrom, W. (2015). Documenting hidden disabilities in higher education: Analysis of recent guidance from the association on higher education and disability (AHEAD). *Journal of Disability Policy Studies*, *26*(1), 44–53. [CrossRef]

McGuire, S. (2015). *Teach Students How to Learn*. Routledge.

McNiff, J., & Whitehead, J. (2002). *Action research: Principles and practice, second edition*. RoutledgeFalmer. Available online: http://www.actionresearch.net (accessed on 13 January 2025).

Meadows, E., & Wickner, D. (2020). *Antiracism: Intent vs. impact*. The International Educator.

Moog, R. S., & Spencer, J. N. (2008). POGIL: An overview. In *Process oriented guided inquiry learning (POGIL)* (Vol. 994, pp. 1–13). ACS Symposium Series; Available online: https://pubs.acs.org/sharingguidelines (accessed on 13 January 2025).

National Center for Education Statistics. (2023). Number and percentage distribution of students enrolled in postsecondary institutions, by level, disability status, and selected student characteristics: Academic year 2019–2020. In *Table311.10*; U.S. Department of Education, Institute of Education Sciences. Available online: https://nces.ed.gov/programs/digest/d22/tables/dt22_311.10.asp (accessed on 13 January 2025).

NGSS Lead States. (2013). *Next generation science standards*. National Academies Press.

Nielsen, J. A. (2001). Successful university students with learning disabilities. *Journal of College Student Psychotherapy*, *15*(4), 37–48. [CrossRef]

Oleson, A., & 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. *Higher Education*, *68*(1), 29–45. [CrossRef]

Salzer, M. S., Wick, L. C., & Rogers, J. A. (2008). Familiarity with and use of accommodations and supports among postsecondary students with mental illnesses. *Psychiatric Services*, *59*(4), 370–375. [CrossRef]

Seery, M. K. (2015). Flipped learning in higher education chemistry: Emerging trends and potential directions. In *Chemistry Education Research and Practice* (Vol. 16, pp. 758–768). Royal Society of Chemistry. [CrossRef]

Shakespeare, T. (2006). The social model of disability. *The Disability Studies Reader*, *2*(3), 197–204.

Stevenson, M. (2010). "If they can't stand the heat...": Supporting the academic development of higher education students with anxiety and depression disorders. *The Open Rehabilitation Journal*, *3*, 41–46.

Storrie, K., Ahern, K., & Tuckett, A. (2010). A systematic review: Students with mental health problems-A growing problem. *International Journal of Nursing Practice*, *16*(1), 1–6. [CrossRef] [PubMed]

Stowe, R. L., & Cooper, M. M. (2019). Assessment in chemistry education. *Israel Journal of Chemistry*, *59*(6), 598–607. [CrossRef]

Talbert, R., & Clark, D. (2023). *Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education*. Taylor & Francis.

Valli, L. (1997). Listening to other voices: A description of teacher reflection in the United States. *Peabody Journal of Education*, *72*(1), 67–88. [CrossRef]

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, *92*(5), 941–967. [CrossRef]

Wong, B., & Chiu, Y. L. T. (2021). Exploring the concept of 'ideal' university student. *Studies in Higher Education*, *46*(3), 497–508. [CrossRef]

education sciences

MDPI

Article

Who Learns from Reading Texts in General Chemistry?

René W. Buell [1] and Samuel Pazicni [2,*]

[1] Department of Chemistry and Biochemistry, Gonzaga University, Spokane, WA 99258, USA; buell@gonzaga.edu
[2] Department of Chemistry, University of Wisconsin–Madison, Madison, WI 53705, USA
* Correspondence: sam.pazicni@chem.wisc.edu

Abstract: Understanding how student characteristics affect learning in General Chemistry can influence the pedagogical strategies employed by instructors. Previous studies have investigated the effects of characteristics including prior knowledge, math ability, and motivations on course performance. Student characteristics can also influence study strategies employed by students. Few studies, however, have focused on the role of language and reading comprehension skill in learning in chemistry. This work explores the effects of prior knowledge and reading comprehension skill on learning from reading texts about two chemistry concepts. Linear regression analyses were utilized to establish relationships between predictors and test scores after reading to determine whether reading comprehension skill influenced learning gains after reading texts. A meta-analysis of four large-scale studies showed that prior knowledge and reading comprehension correlated with post-test scores and that an effect called expertise reversal may help low prior knowledge students close the post-test score gap if they read texts with certain readability characteristics. However, our findings also suggest that texts with similar readability characteristics can hinder the learning of those with higher prior knowledge.

Keywords: first-year undergraduate chemistry; textbooks; reading comprehension; prior knowledge; linear regression; meta-analysis

1. Introduction

General Chemistry courses are foundational to many science-related fields, serving as a launching point for more specialized studies in chemistry, biology, engineering, medicine, and other scientific disciplines. Traditionally, textbooks have served as one of the principal learning resources in chemistry courses, as they provide students with structured content, representations, explanations, and practice exercises. A recent review of research on chemistry textbooks by Thompson, Bunch, and Popova [1] highlights nearly forty years of work, which includes characterizations of various aspects of texts (e.g., sequencing of topics, presentation of representations, and gender/racial representation) and how chemistry textbooks are written, perceived, or used (e.g., students' and teachers' perceptions and use of textbooks). Of note, Thompson et al. cited only three studies that assessed student learning from chemistry texts [2–4]. Each of these studies was performed in the secondary education space and found that conceptual difficulties remained after engaging with texts. In response to the limited work in this space, this review recommended that "research is needed to evaluate the effectiveness and use of higher-level postsecondary chemistry textbooks" (p. 2891). The present contribution aims to address this need by exploring learning from university-level chemistry texts.

The effectiveness of a common text resource among a diverse group of students may be limited, as both student and text characteristics vary considerably. Students in General Chemistry are diverse in terms of prior knowledge and reading skill. The effect of both individual differences has been investigated in the context of overall performance in General Chemistry courses [5]. This work found that both students' reading skill and prior knowledge were correlated with exam scores and course performance; findings also suggested that reading skill can partially compensate for low prior knowledge. In addition, the language used in General Chemistry textbooks spans a wide range of readability between different textbooks and even within the same textbook [6]. In this work, the five most widely used General Chemistry textbooks of 2012 were analyzed and compared to a best-selling popular novel. On average, General Chemistry textbooks used longer, more complex sentences than novels and are therefore more difficult to read. However, the General Chemistry texts had wide ranges of referential cohesion (how often the text uses overlapping words and ideas to make connections between ideas for the reader) and more limited ranges of deep cohesion (the extent to which the text makes causal and intentional connectives between ideas).

Crucially, both student and text characteristics have been shown to affect comprehension of science texts. For example, in the context of a General Biology course, students with low and high prior knowledge were given reading material where text cohesion was varied [7]. This work found that both prior knowledge and reading ability acted as predictors for performance on comprehension questions. In addition, a statistically significant interaction between text cohesion and reading ability was also found, suggesting that more cohesive texts benefit learners of lower reading skill. The work presented here contributes to this evidence base by exploring the extent to which different types of chemistry students learn from texts like those found in General Chemistry textbooks. Moreover, understanding how student and text characteristics interact with learning in this space is critical to the development of equitable learning materials for General Chemistry.

2. Theoretical Framework

This work is grounded in two theoretical perspectives: cognitive load theory and misconceptions constructivism. Cognitive load theory models how working memory functions during learning tasks (like reading) and informs not only our experimental design but also the interpretation of our findings. Misconceptions constructivism frames our ideas and assumptions about knowledge and learning and informs aspects of our experimental design.

2.1. Working Memory and Cognitive Load Theory

Working memory is a cognitive structure that plays an important role in the short-term retention of information and transfer of information to long-term memory [8]. Cognitive Load Theory is a model of how working memory interacts with long-term memory during learning activities, including reading [9,10]. During a learning event, the limited capacity of working memory moderates the extent to which information is transferred to long-term memory. “Cognitive load” is the term given to describe the amount and complexity of information being processed in short-term memory [10]. Two categories of cognitive load are germane to the present work: intrinsic cognitive load (determined by the inherent complexity of the material being learned) and extraneous cognitive load (extra cognitive burden caused by poor instructional design). Regardless of the source, the greater the amount and/or complexity of the information being processed, the higher the cognitive load. Long-term memory also plays a role in this process, as greater background knowledge of what is being learned results in a heightened ability to process information in working memory [11].

While learning from a text is influenced by the working memory’s ability to process and integrate new information into long-term memory, aspects of the text itself interact with working memory. Highly cohesive materials (i.e., those that provide cues linking phrases and sentences and relate information across the text [12]) are known to reduce cognitive load for the reader [13]. However, text cohesion can also influence the extent to which a learner can leverage background knowledge [14]. Learners with lower background knowledge tend to benefit from texts with high cohesion, while learners with higher background knowledge tend to benefit from texts with low cohesion. Learners with higher background knowledge appear to experience a larger cognitive load when engaging with highly cohesive texts, as the referential cues in these texts present as extraneous information. However, texts of low cohesion require the learner to generate inferences using their background information, permitting the learner with higher background knowledge to more actively process the text. That said, texts of low cohesion can increase cognitive load for learners of low background knowledge, as they lack the knowledge to draw upon when making inferences and establishing meaning.

2.2. Misconceptions Constructivism and Conceptual Change

This work is situated generally in a constructivist philosophy, that "knowledge is constructed in the mind of the learner" [15] (p. 873) rather than being transmitted to a learner by an expert. The cornerstone of any constructivist framework is the critical role of a learner's prior knowledge in constructing new understanding. However, there are several different perspectives regarding the nature of this prior knowledge and how a learner's understanding changes over time, notably misconceptions constructivism and fine-grained constructivism [16,17]. From a misconceptions constructivist perspective [18–21], prior knowledge consists of large, stable ideas—some of which are scientifically nonnormative (i.e., "misconceptions"). The conceptual change process typically associated with misconceptions constructivism is "confront and replace" [16] (p. 483), where a learner is confronted with evidence that contradicts their misconception and guided to recognize that a more canonical idea (and not their misconception) leads to more consistent and productive thinking [22,23]. For example, of most relevance to this work are conceptual change texts [24–26], where common misconceptions are explicitly pointed out, and then a subsequent discussion explains why the misconceptions do not accurately represent the concept. In contrast, a fine-grained constructivist perspective views prior knowledge as a dynamic constellation of small cognitive units that are activated in more or less appropriate ways within a given context [27–29]. Within a fine-grained framework, conceptual change is not concerned with identifying and replacing non-normative ideas with more accepted ones; rather, conceptual change is a matter of restructuring—refining fragmented knowledge elements into a coherent and coordinated structure. Historically, chemistry education research has framed student knowledge from a misconceptions constructivist perspective (e.g., [30–32]); however, there is a rising trend in using fine-grained constructivist models in the field [33].

For this work, we adopt a misconceptions constructivist perspective. We do so principally due to the experimental limitations of using a fine-grained constructivist model. Learning interventions within a fine-grained framework tend to be associated with responsive teaching (in science education) [34] or teacher noticing (in mathematics education) [35] and occur at the classroom level via teacher professional development facilitated by research teams (e.g., [36]), while assessing student knowledge and conceptual change is typically a qualitative venture, one that employs knowledge analysis [28] as the principle methodology. The nature of this work necessitated a small-scale learning intervention unassociated with a particular course as well as a quantitative measure of student knowledge. These constraints are well met by a misconceptions constructivist perspective, where learning interventions can focus on confronting and replacing noncanonical conceptions with texts, and student knowledge can be measured with concept inventories [37], which assume that students' ideas are context-independent and of sufficiently large grain size to be evaluated by fixed-choice items.

3. Research Questions

Because textbooks (both physical and web-based) are such ubiquitous reference materials in General Chemistry classrooms, it is imperative to understand whether learning gains can be achieved by reading and to what extent reading texts helps a student learn General Chemistry concepts. As undergraduate students exhibit a range of reading comprehension abilities and prior knowledge about chemistry concepts, understanding how students with different backgrounds learn from typical General Chemistry texts can inform authors and instructors about how texts can be tailored to promote learning for all students. The goal of understanding how student prior chemistry knowledge and reading skill affect learning gains that students can achieve by engaging with text-based materials leads to the following research questions:

1. To what extent does reading about a General Chemistry concept promote learning gains?
2. To what extent do individual student characteristics (prior knowledge and reading comprehension skill) moderate learning from reading a General Chemistry concept text?

4. Materials

4.1. Text Passages on General Chemistry Concepts

The chemistry concepts chosen for this study were bonding representations and redox reactions. We developed text passages for these concepts, simulating writing found in textbooks in terms of content and text readability measures. However, even though representations are common in chemistry textbooks, we chose not to include them in the simulated text passages, as it would not be possible to control for how/if students engage with the representations in this study. While research on the number, type, and presentation of images and representations in chemistry textbooks is an active area of inquiry in chemistry education [1], exploring how text and representations interact to facilitate learning was beyond the scope of this work.

Consistent with the theoretical framing of this work, the texts were also designed to be conceptual change texts and include topics that students often hold incorrect or unproductive ideas about to test whether learning occurs upon reading the text passage. The information included in the text passages was chosen based on published literature on students' ideas of bonding [38–42] and redox reactions [43–48]. In all experiments, unrelated expository text passages about loons or psychedelic mushrooms were used for comparison. The text passages developed for this study are provided in Appendix A.

All passages were limited to approximately 600 words, and each scored at the College level on commonly used readability tests (Table 1). Additionally, the readability of the passages was analyzed using Coh-Metrix [49,50]. Values for the five readability dimensions of both texts fell within 1.5 standard deviations of the mean readability measures for commonly adopted General Chemistry textbooks [6] (Table 2).

Table 1. Word counts and readability test scores for the text passages developed for this study.

Attribute	Bonding Representations	Redox Concepts
number of words	456	605
Flesh reading ease	35.6	40.3
Flesh-Kincaid grade level	13.0	13.4

Table 2. Coh-Metrix measures of readability dimensions for the text passages developed for this study, as compared to median values found for five General Chemistry textbooks.

Dimension	Bonding Representations	Redox Concepts	Mean for GC Textbooks (St. Dev) [1]
narrativity	5% [2]	10%	21.9% (13.8)
syntactic simplicity	75%	49%	52.0% (20.7)
word concreteness	7%	23%	32.0% (24.3)
referential cohesion	86%	96%	65.4% (27.6)
deep cohesion	33%	83%	54.8% (25.8)

[1] Taken from supporting information of ref [6]. [2] Percentages represent the percentile at which the measured dimension falls.

4.2. Measure of Reading Comprehension Skill

The standardized Gates-MacGinitie reading test (Comprehension 10/12—Form S 4th edition, GMRT) was used to measure students' reading comprehension skills [51].

The test is timed and comprised of a series of short text passages followed by four or five questions about each passage (48 items total). GMRT results can be compared to normalized reading levels based on national (U.S.) standards; the GMRT Comprehension 10/12—Form S was designed to be taken by high school students, and norms for this population were published for the fall, winter, and spring terms during the students' final year. As most of the participants in the study were first-year university students, GMRT scores for this study were compared to the published spring term norms [52]. The grand mean for the GMRT scores across this study's five experimental samples was 63%, which falls at the 49th percentile of graduating high school seniors; the mean GMRT scores across experimental samples were not statistically significantly different from each other. Descriptive statistics for all GMRT data are reported in Section S1 of the Supplementary Materials.

4.3. Measures of Conceptual Understanding for Bonding Representations and Redox Reactions

To measure student understanding of bonding representations and redox reactions, previously published concept inventories were used. The Bonding Representations Inventory (BRI) [53] and Redox Concept Inventory (ROXCI) [48,54] were completed by students at the beginning of each semester of study (pre-test) and approximately four weeks later (post-test). The BRI includes items referring to periodic trends, electrostatic interactions, surface features, the octet rule, and how those concepts are represented in chemical bonding. Bonding concepts are often introduced during the first semester of General Chemistry. The ROXCI includes items about oxidation numbers, electrostatics and bonding, electron transfer, surface features, and roles of spectator ions, and these concepts are often introduced during the second semester of General Chemistry. Descriptive statistics for all BRI and ROXCI data are reported in Sections S2–S5 of the Supplementary Materials.

5. Methods

5.1. General Experimental Design

The same general between-subjects pre-/post-test design [55] was used for the five experiments that comprised this study. At the beginning of a semester during which an experiment took place, participants completed assessments to measure their reading comprehension skill (GMRT) and their level of prior knowledge for one of two General Chemistry concepts (BRI or ROXCI). Three weeks later, an intervention and post-test were administered. The intervention consisted of reading a text passage about the topic of the pre-test. A short time later (~1–3 h), a post-test (identical to the pre-test) was administered. In each experiment, a comparison group of students (randomly selected from the larger sample) was given an unrelated text passage that had similar text characteristics as the chemistry concept text passages. All experiments occurred before formal instruction on the two General Chemistry concepts occurred.

5.2. Participants and Setting

Participants of this study were enrolled in a traditional two-semester General Chemistry course sequence and in a one-semester General Chemistry course for engineering majors at a four-year public research university in the northeastern United States. Experiments with five different samples of students comprised this study. The first four experiments (Exps. 1A–1D) were exploratory in nature and conducted with large samples of students who had a range of prior knowledge of the concepts. The fifth experiment (Exp. 2) tested a finding from Exps. 1A–1D and was conducted with a sample of lower prior knowledge students (additional details to follow). Demographic information for study participants was collected from school records and provided in Section S6 of the Supplementary Materials.

In all experiments, students completed the GMRT and either the BRI or ROXCI at the beginning of the semester. Students in the first-semester General Chemistry course completed the BRI, students in the second-semester General Chemistry course completed the ROXCI, and students in the one-semester course designed for engineering majors completed either the BRI or ROXCI. The experiments comprising this study are summarized in Table 3. For Exps. 1A–1D, students were allotted time at the beginning of a three-hour laboratory session to read the text passage intervention (or associated comparison text), and the post-test was administered at the end of the lab session. For Exp. 2, the intervention and post-test were performed in a conference room, with the post-test being given approximately one hour after the reading of the text passage.

Table 3. Summary of the five experiments comprising this study.

Experiment Label	Course	Concept Inventory	Sample Size
1A	GC [a] 2	ROXCI	290
1B	Engineers GC	BRI	143
1C	GC 1	BRI	396
1D	Engineers GC	ROXCI	145
2	GC 1	BRI	56

[a] General Chemistry

6. Data Analysis

6.1. Data Processing

IBM SPSS statistics software was used for the processing and analysis of all data. In all cases, GMRT scores and BRI/ROXCI scores were transformed into centralized *z*-scores to generalize results and compare values with different scales, means, and standard deviations [56] (p. 75). All multivariate outliers were eliminated from analyses using a Mahalanobis distance criterion of $p < 0.001$ [57] (p. 99). Five outliers were omitted from Exp. 1A, no outliers were found in Exp. 1B, twenty-five outliers were omitted from Exp. 1C, twenty outliers were omitted from Exp. 1D, and two outliers were omitted from Exp. 2. Following the removal of outliers, the data for each experiment were inspected for deviations from normality, employing a hypothesis testing method using kurtosis and skew values [58].

6.2. Regression Analyses

Regression analyses were used to explore significant predictors of post-test scores (dependent variable). The following linear regression model was tested in each experiment:

$$z_{\mathrm{post}} = \beta_0 + \beta_1(txt_{\mathrm{rel}}) + \beta_2(z_{\mathrm{GMRT}}) + \beta_3(z_{\mathrm{pre}}) + \beta_4(txt_{\mathrm{rel}} \times z_{\mathrm{GMRT}}) + \beta_5(txt_{\mathrm{rel}} \times z_{\mathrm{pre}}) + \beta_6(z_{\mathrm{GMRT}} \times z_{\mathrm{pre}}) + \beta_7(txt_{\mathrm{rel}} \times z_{\mathrm{GMRT}} \times z_{\mathrm{pre}}) \quad (1)$$

where independent variables tested to be predictors of post-test scores (z_{post}) included pre-test *z*-scores (z_{pre}), reading comprehension skill *z*-scores (z_{GMRT}), participation in the intervention of reading a text passage about the topic (txt_{rel}: dummy scored, did read relevant text = 1, comparison group = 0), and all possible interactions between main effects.

Variables were entered stepwise based on how much the model R^2 increased by including the variable; variables that increased R^2 the most were entered into the regression equation first. If a variable became not meaningful or not statistically significant to the model after the addition of another variable, it was removed from the model [56] (pp. 560–651). Given the exploratory nature of Exps. 1A–1D, variables were considered potentially meaningful if they uniquely explained more than 1% of variance in the data; this threshold was chosen based on known difficulties in detecting moderating effects [59]. Given the confirmatory nature of Exp. 2, the threshold for statistical significance was set to $\alpha = 0.05$. The extent to which the regression analyses reported here satisfy the five assumptions underlying multiple regressions [60] (pp. 45–47) is discussed in Appendix B.

Both standardized (β) and unstandardized coefficients (B) are reported here, but the discussion of the models focuses on the unstandardized coefficients (as predictor variable data had been standardized using *z*-scores). Squared semipartial correlations (sr^2) are reported and represent the variance of outcome scores that could be uniquely predicted by each variable. For a predictor, sr^2 can be used to interpret effect size when all other predictors were statistically controlled. The semipartial correlation (*sr*) was used when applying Cohen's effect size guidelines [61] (pp. 79–81).

6.3. Meta-Analysis: Experiments 1A–1D

When large-scale experiments are used to test effects of interventions, results are often untested with additional studies. If the results are investigated with a new sample, the outcomes and effects may be different, and the researchers then face the difficult task of reconciling disparate or conflicting results. Meta-analysis is a technique for comparing and combining results from multiple studies. By analyzing the effect sizes of predictors in different experiments, a weighted average effect size can emerge to produce a more generalizable effect size that may not have been apparent in any one study [62–65]. However, meta-analyses may be hindered by publication bias, as studies that do not present marked results tend not to be published [66]. A small-scale meta-analysis can be conducted using results from several experiments carried out by a research team. Using unpublished results will eliminate the danger of publication bias. Another criticism of meta-analysis is the heterogeneity of studies investigated. Including studies conducted by the same research group and with the same methodology will eliminate the concerns raised by heterogeneity of methodologies and analysis techniques.

The data used in this meta-analysis emerged from the four large-scale experiments (Exps. 1A–1D): the effect sizes of each predictor (the semipartial correlations, *sr*) and sample sizes. As there was no reason to suspect heterogeneity among experiments [67], the effects were entered as fixed into ESCI (Estimation Statistics with Confidence Intervals) in Excel [68]. For each predictor, the output of the ESCI in Excel analysis includes 95% confidence intervals (CIs) for the effect size from each experiment, the weight percentage of each study, and the overall weighted effect of the predictor (and accompanying confidence intervals) [65].

7. Results: Experiments 1A–1D

As discussed above, two research questions framed this study; these research questions were explored in four separate large-scale experiments, sampling from three different General Chemistry courses, and using two different chemistry concepts: bonding representations (assessed using the BRI) and redox reactions (assessed using the ROXCI). A summary of regression analysis results for Exps. 1A–1D is provided in Section S7 of the Supplementary Materials.

7.1. Experiment 1A

Exp. 1A employed a sample of 290 participants in the second semester of a traditional two-semester General Chemistry sequence. Here, the ROXCI was used as the pre- and post-test measure of student understanding. The final regression model (adjusted $R^2 = 0.20$, $F(3, 286) = 10.98$, $p < 0.001$) explained approximately 20% of the variance in post-test ROXCI scores, and it included the main effects of the GMRT z-score ($p = 0.04$, $sr^2 = 0.01$) and pre-test *z*-score ($p < 0.001$, $sr^2 = 0.08$). On average, students scored a *z*-score of zero on the post-test (by definition of a *z*-score), but students scoring one standard deviation above the mean on the pre-test were predicted to score 0.45 standard deviations above the mean on the post-test when controlling statistically for other variables. Reading skill and prior knowledge uniquely predicted 1% and 8% of the variation in post-test ROXCI scores, respectively; these effects are considered to have small to medium effect sizes. In addition, an interaction between reading comprehension skill score and pre-test score was found to be potentially meaningful ($p = 0.05$, $sr^2 = 0.01$) and have positive value ($\hat{\beta}_6 = 0.17$). Thus, in this sample of students, those who were more skilled readers were predicted to score higher on the post-test than average readers (controlling statistically for other variables), and higher prior knowledge students were also predicted to perform better than average students. Moreover, given the positive interaction between reading comprehension skill score and pre-test score, the main effect of prior knowledge was more pronounced for more highly skilled readers (and vice versa). Interestingly, no main effect of reading a relevant text was detected in this sample (i.e., whether students read about redox concepts or read an unrelated text had no effect on the post-test ROXCI score). The final regression model for Exp. 1A was as follows:

$$z_{ROXIpost} = 0.18(z_{GMRT}) + 0.45(z_{ROXIpre}) + 0.17(z_{GMRT} \times z_{ROXCIpre}) \quad (2)$$

7.2. Experiment 1B

The second large-scale experiment was conducted with participants of a one-semester General Chemistry course designed for engineering majors. This sample (n = 143) completed the BRI. The results of the linear regression analysis for Exp. 1B differed from Exp. 1A. The final model (adjusted $R^2 = 0.28$, $F(2, 141) = 8.64$, $p < 0.001$) explained 28% of the variance in BRI post-test scores and included BRI pre-test scores and the interaction between reading a related text and BRI pre-test scores as the two meaningful ($sr^2 \geq 0.01$) predictors. On average, students scored a z-score of zero on the post-test, but those with higher prior knowledge were predicted to score positive z-scores ($p < 0.001$, $sr^2 = 0.13$). Students scoring one standard deviation above the mean on the pre-test were predicted to score 0.57 standard deviations above the mean on the post-test when controlling for other variables statistically.

The interaction between prior knowledge and reading the text was also found to be meaningful ($p = 0.08$, $sr^2 = 0.02$). The negative association between this interaction and post-test scores ($\hat{\beta}_5 = -0.29$) suggests that for those who did read the text about bonding, gains were made for students with lower prior knowledge (pre-test z-scores that are negative, or lower than the mean), and losses were experienced for students with higher prior knowledge (pre-test z-scores above the mean). This interaction had a small effect size based on the squared semipartial and uniquely predicted 2% of the variance of post-test scores. The final model for Exp. 1B was as follows:

$$z_{BRIpost} = 0.57(z_{BRIpre}) - 0.29(txt_{rel} \times z_{BRIpre}) \quad (3)$$

For students scoring one standard deviation above the mean on the pre-test and who read the bonding text as a learning intervention, the predicted post-test z-score would be 0.28 standard deviations above the mean. For students who similarly scored one standard deviation above the mean on the pre-test but did not read the text, the predicted post-test z-score would be 0.57 standard deviations above the mean. Based on this model, students with higher prior knowledge would not have learning gains upon reading the text and would be predicted to score below those who did not read an expository text about the chemistry topic. However, students with lower prior knowledge would be helped by reading the text. A student scoring one standard deviation below the mean on the pre-test (and who did not read text) would have a predicted post-test score of 0.57 standard deviations below the mean. By reading the text, however, this student would be predicted to score 0.29 standard deviations closer to the mean average on the post-test (a final post-test z-score of -0.28).

7.3. Experiment 1C

The sample for Exp. 1C consisted of 396 students enrolled in the first-semester course of a traditional two-semester General Chemistry sequence; these students were tested on bonding representations using the BRI. The final linear regression model for these data included all three main effects as potentially meaningful ($sr^2 \geq 0.01$) predictors, with no meaningful interactions (adjusted $R^2 = 0.19$, $F(3, 395) = 12.83$, $p < 0.001$). Approximately 17% of the variance in BRI post-test scores was explained by this model. All three main effects had positive associations with BRI post-test scores. Students who read the bonding text were predicted to score 0.30 standard deviations higher than the mean on the post-test compared to students who did not read the text when reading ability and prior knowledge were controlled statistically ($p < 0.001$, $sr^2 = 0.02$). The effect size of this predictor was small. Reading comprehension skill was found to meaningfully predict post-test scores, and the effect size was also small based on the squared semipartial (sr^2 = 0.01). Participants who were one standard deviation above the rest of the sample in reading skill would be predicted to score 0.16 standard deviations above the class mean on the BRI post-test, controlling for prior knowledge and reading the intervention text (p = 0.03). Finally, prior knowledge was found to have the largest effect on BRI post-test scores, with students one standard deviation above the mean on the pre-test predicted to score 0.41 standard deviations above the mean on the post-test (p < 0.001, sr^2 = 0.06). The final regression model for Exp. 1C was as follows:

$$z_{\text{BRIpost}} = -0.18 + 0.30(\text{txt}_{\text{rel}}) + 0.16(z_{\text{GMRT}}) + 0.41(z_{\text{BRIpre}}) \quad (4)$$

7.4. Experiment 1D

The final large-scale experiment was conducted with a one-semester General Chemistry course designed for engineering students (n = 145), and the concept inventory used was the ROXCI. The final model included two meaningful ($sr^2 \geq 0.01$) predictors: whether a relevant text was read and the interaction between reading a relevant text and prior knowledge. The overall model, however, was not statistically significant (adjusted R^2 = 0.01, $F(2, 144)$ = 1.26, p = 0.28). Only 1% of the variance in ROXCI post-test scores was predicted by the variables available in this study. That said, the variables that are potentially meaningful are discussed below.

The effect of reading a related text was meaningful (p = 0.08, sr^2 = 0.02), with a small effect size based on the squared semipartial. The text intervention, however, was opposite of the effect found in Exp. 1C, as reading in this sample had a negative association with the ROXCI post-test ($\hat{\beta}_1 = -0.24$). Students who read the text, on average, scored 0.24 standard deviations below the mean (controlling for other variables), whereas students who did not read the text were predicted to score the mean on the post-test. However, when prior knowledge as measured by the ROXCI pre-test was accounted for, an interaction was found between reading the related text and prior knowledge (p = 0.05, sr^2 = 0.03). The effect was similar to the effect of the interaction found in Exp. 1B, as there was a negative coefficient for the interaction ($\hat{\beta}_5 = -0.37$). This interaction predicts that for a high prior knowledge student (who scores one standard deviation above the mean on the pre-test), reading the text about redox concepts will result in a predicted post-test score 0.37 standard deviations lower than the mean. Interestingly, the main effect of prior knowledge based on pre-test scores was not found to be meaningfully predictive of post-test scores (sr^2 < 0.01), even though this predictor was meaningful in Exps. 1A–1C. The final regression model for Exp. 1D was as follows:

$$z_{\text{ROXIpost}} = -0.24(\text{txt}_{\text{rel}}) - 0.29(\text{txt}_{\text{rel}} \times z_{\text{ROXCIpre}}) \quad (5)$$

7.5. Meta-Analysis of Experiments 1A–1D

Because the findings from Exps. 1A–1D were inconsistent among themselves, a meta-analysis was performed to synthesize the results from these large-scale studies. The four effects that emerged most often from Exps. 1A–1D were included in the meta-analysis: whether a relevant text was read, reading comprehension skill level (GMRT z-scores), prior knowledge (pre-test z-scores), and the interaction between reading a related text and prior knowledge. The complete results of the meta-analysis are provided in Section S8 of the Supplementary Materials, and a forest plot of the results is presented in Figure 1. In a forest plot, if a CI overlaps with a semipartial correlation (sr) of zero, then that effect is interpreted as being not statistically significant at the threshold of confidence employed (α = 0.05). All effect sizes were interpreted using Cohen's guidelines [61] (pp. 79–81).

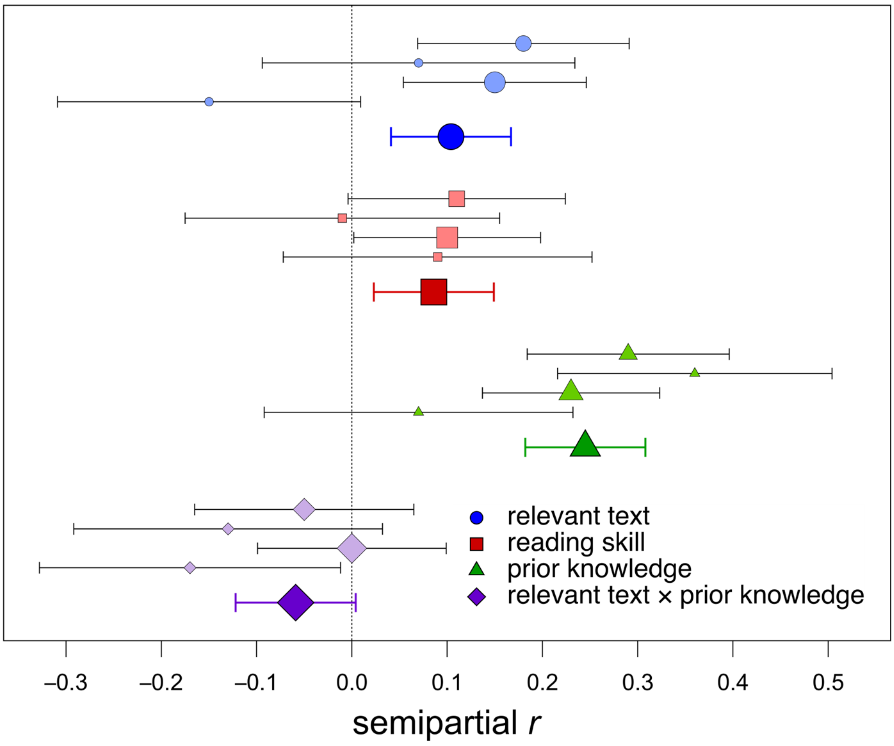


Figure 1. Forest plots for the four predictors included in the meta-analysis. The effects are (blue circles) whether a relevant text was read, (red squares) reading comprehension skill, (green triangles) measure of prior knowledge (pre-test scores), and (violet diamonds) the interaction between reading relevant text and prior knowledge. The weighted mean *sr* for each effect is displayed below each set of four experiments with confidence intervals in color.

Overall, reading a relevant text appeared to have a small effect on post-test scores: *sr* = 0.10, 95% CI [0.04, 0.17]. This finding suggests that simply providing reading material about a concept will have a slim effect on learning gains when not accounting for other student characteristics. Additionally, reading comprehension skill emerged to have a statistically significant (but small) effect on post-test scores in the fixed effects model: *sr* = 0.09, 95% CI [0.02, 0.15]. Students with higher reading comprehension skill levels were predicted to have positive gains on post-test scores, controlling for other variables. Prior knowledge had the largest effect size (medium) on post-test scores: *sr* = 0.25, 95% CI [0.18, 0.31]. It is no surprise that pre-test score has a positive correlation with post-test score, statistically controlling for other variables. Students who have coherent understanding of a concept during a pre-test will likely maintain that understanding during a post-test. The final effect tested was the interaction between reading the relevant concept text and prior knowledge. This effect was not statistically insignificant at $\alpha = 0.05$: *sr* = −0.06, 95% CI [−0.12, 0.004], but these results do have implications that will be addressed in the discussion section.

8. Results: Experiment 2

Given the exploratory findings of Exps. 1A–1D, it was clear that student prior knowledge played a critical role in learning from text materials, both as a main effect and a potential moderating effect with reading a relevant text. To gain further insight into how prior knowledge influences learning from texts, Exp. 2 was designed to constrain the sample to students of lower-than-average prior knowledge. This was done for two reasons: (1) to provide additional evidence for the negative interaction effect of prior knowledge and reading a relevant text observed in Exps. 1A–1D (i.e., reading a related text should result in higher post-test scores for students of low prior knowledge) and (2) to confirm that chemistry concept passages possessing high cohesion (similar to General Chemistry textbooks, vide supra) can aid students of lower prior knowledge.

For Exp. 2, 459 participants completed the BRI pre-test (M = 8.8, SD = 3.4). Following an invitation from the researchers, 56 students who scored below average on the BRI pre-test (M = 8.2, SD = 2.5) consented to participate in the subsequent component of this experiment (a reading intervention, followed by BRI post-test). During this subsequent experiment phase, the intervention group read the text about bonding concepts, while the comparison group read a text about loons.

Like Exps. 1A–1D, a linear regression model (Equation (1)) was tested to determine whether any variables emerged as statistically significant predictors of BRI post-test z-scores. The analysis indicated that three predictors were statistically significant at 95% confidence: reading the bonding text passage, the interaction between BRI pre-test scores and reading the relevant text, and the interaction between BRI pre-test scores and GMRT scores (adjusted R^2 = 0.45, $F(3, 52) = 8.72$, $p < 0.001$). The results of this analysis are presented in Table 4. The final model explained approximately 45% of the variance in BRI post-test scores. The mean BRI post-test z-score for the sample was −0.71 ($p < 0.001$), controlling for BRI pre-test scores, reading comprehension skill, and whether the bonding text was read. Notably, BRI pre-test score was not a statistically significant main effect, likely because the students in this controlled experiment all scored below the mean of the larger sample on the BRI pre-test, and so the range of pre-test scores in this sample was small.

Table 4. Regression analysis results for Experiment 2.

Variable	B	SE	β	t	sr^2	p
intercept (β_0)	−0.71	0.18		−4.04		<0.001
read bonding text (β_1) (1 = did read; 0 = did not read)	1.11	0.22	0.52	5.16	0.27	<0.001
read bonding text × BRI pre-test score (β_5)	0.54	0.13	0.42	4.12	0.17	<0.001
BRI pre-test score × GMRT score (β_6)	−0.20	0.09	−0.22	−2.16	0.05	0.04

R^2 = 0.48, adjusted R^2 = 0.45, R = 0.69, $F(3, 52) = 8.72$, $p < 0.001$.

Participants who read the text about bonding representations scored, on average, 1.1 standard deviations higher on the post-test than students who did not read the text ($p < 0.001$, sr^2 = 0.27). To put this large effect size into context, approximately 27% of the variance in BRI post-test scores could be uniquely predicted by whether the participant read the relevant text, controlling for all other variables. Thus, engaging with a relevant text had a large effect on learning gains for this sample of lower-than-average prior knowledge students. This outcome can be visualized by inspecting the distributions of pre- and post-test BRI scores, separated by whether students engaged with the relevant text passage (Figure 2).

The second statistically significant predictor of BRI post-test z-scores was the interaction between prior knowledge and engaging with the bonding text passage. Students who read the text and scored one standard deviation above the mean on the BRI pre-test (i.e., near average when considering the entire class sample) were predicted to score 0.54 standard deviations above the mean compared to similar students who did not read the text ($p < 0.001$, sr^2 = 0.17). A total of 17% of variance in post-test scores was uniquely explained by the interaction. The interaction was of medium effect size, based on the semipartial correlation. This finding suggests that average prior knowledge students were aided more by reading the text than those with very low prior knowledge, regardless of reading comprehension ability.

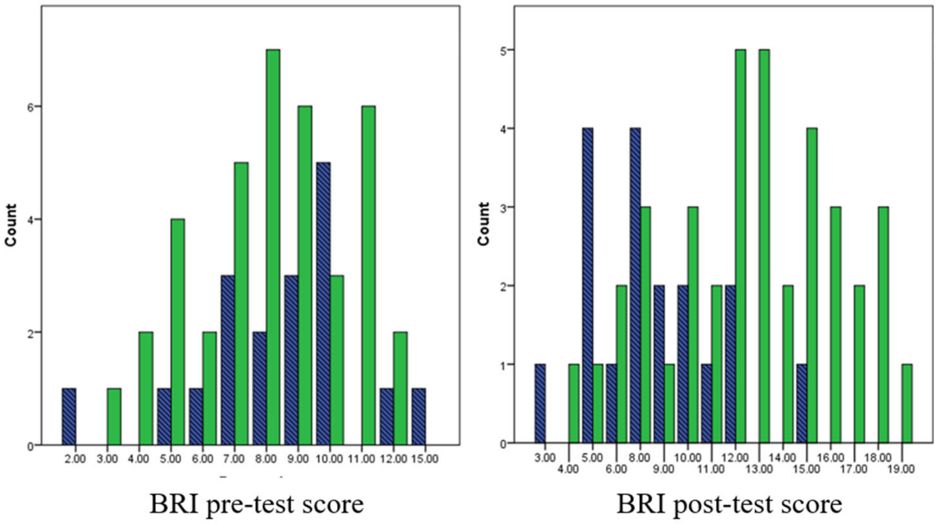


Figure 2. (**left**) BRI pre-test scores for students who did engage with the text on bonding (green) and who did not (blue patterned); (**right**) BRI post-test scores for students who did engage with the text on bonding (green) and who did not read the text (blue patterned). All students participating in this experiment had lower BRI pre-test scores than the rest of the class from which they were recruited. Independent samples *t*-tests confirm that the difference in mean BRI score is statistically significant for the post-test score, $t(56) = 4.0$, $p < 0.001$, Hedge's $g = 1.2$.

Lastly, the third statistically significant predictor of BRI post-test *z*-scores was the interaction between prior knowledge and reading comprehension skill ($p = 0.04$, $sr^2 = 0.05$). The estimate of this interaction effect was negative ($\hat{\beta}_6 = -0.20$), meaning that while the relationship between GMRT score and BRI post-test *z*-scores was negative for students of higher prior knowledge, the relationship was positive for students of lower prior knowledge. To interpret this in the context of Exp. 2, for students who read the related bonding text, the gap in BRI post-test scores between those of low prior knowledge and those of higher prior knowledge closes as GMRT score increases. That is, we observed that reading comprehension skill partially compensated for lower prior knowledge in this sample of students. That said, this interaction was of small effect size, based on the semi-partial correlation. The final linear regression model for Exp. 2 was as follows:

$$z_{\text{BRIpost}} = -0.71 + 1.1(\text{txt}_{\text{rel}}) + 0.54(\text{txt}_{\text{rel}} \times z_{\text{BRIpre}}) - 0.20(z_{\text{GMRT}} \times z_{\text{BRIpre}}) \quad (6)$$

9. Discussion

9.1. Inconsistent Results from Individual Experiments

When the five experiments are investigated separately, the resulting predictors of learning gains (as suggested by the pre-/post-test design) differ from experiment to experiment. In fact, no two experiments resulted in the same group of meaningful or statistically significant predictors of post-test score. Once experiments were compared, however, some predictors were meaningful more often than others, including whether students engaged with a related text passage (Exps. 1C, 1D, and 2), prior knowledge (Exps. 1A, 1B, and 1C), reading comprehension skill (Exps. 1A and 1C), and an interaction between prior knowledge and reading the relevant text (Exps. 1B, 1D, and 2).

Prior knowledge and reading comprehension skill have been found to be predictive of chemistry course performance [5,69]. The interaction between reading a relevant text and prior knowledge, however, has interesting implications for pedagogical strategies. As illustrated in Figure 3, students who read a related text passage did not uniformly experience the same gains in learning as assessed by the BRI or ROXCI—this relationship appears to be moderated by prior knowledge such that reading a text passage resulted in learning gains for those students with lower prior knowledge. Indeed, Exp. 2 confirmed that reading a related text passage could improve post-test scores for students of lower prior knowledge, consistent with the models presented in Figure 3. Of note, this interaction was observed when different chemistry concepts were tested; the effect is not specific to just one concept inventory or one text passage, but results appear to be generalizable to multiple concepts. Interestingly, this interaction was observed in samples of the General Chemistry course designed for engineers; meaningful interactions were not observed in samples of the traditional General Chemistry courses. To better understand the results of the linear regression analyses, effect sizes can be compared through meta-analysis.

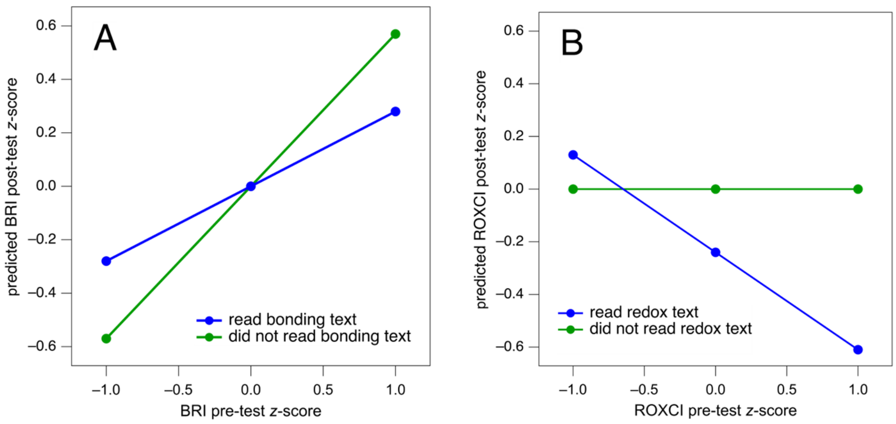


Figure 3. The effect of reading a relevant text on post-test scores for the BRI (**Panel A**, Exp. 1B) and ROXCI (**Panel B**, Exp. 1D). Based on the regression models reported here (Equations (3) and (5)), lower prior knowledge students are predicted to have higher post-test scores if they read the relevant text passage, whereas higher prior knowledge students are predicted to perform worse than their peers if they read the relevant text passage.

9.2. The Effect of Prior Knowledge

The variable that emerged from the meta-analysis of Exps. 1A–1D with the largest effect size (medium, sr = 0.25, 95% CI [0.18, 0.31]) for predicting post-test scores was prior knowledge. The meta-analysis performed using ESCI in Excel uses the degree of overlap between sr confidence intervals for a predictor [65]. Confidence intervals of 95% were computed for the correlation of prior knowledge scores with the post-test scores of each large-scale experiment. As shown in Figure 1, the confidence intervals for prior knowledge overlap for all four studies. When there is significant overlap between confidence intervals of several studies, the power of the calculated mean correlation becomes larger. That is, the probability of committing a Type II error (when a researcher fails to reject a null hypothesis that is actually false) decreases. Even though Exp. 1D did not find prior knowledge to be a meaningful predictor, the confidence interval for the semipartial correlation was wide enough to overlap with the confidence intervals of Exps. 1A–1C, thus enhancing the power of the sr statistic in the meta-analysis. The pooled confidence interval for prior knowledge can be interpreted in the following way: If 100 additional samples were taken, the true mean sr between prior knowledge and post-test score would fall between 0.18 and 0.31 95 times.

Overall, this result comes as no surprise, as linear regression analyses of three of the four large-scale studies found this main effect to be meaningful with medium effect sizes. Moreover, this result is consistent with previous work concerning the effect of prior knowledge on quantitative learning outcomes, both in chemistry [5,70,71] and in more general scenarios [72]. Given this effect across Exps. 1A–1D, it was somewhat surprising that no statistically significant effect for prior knowledge was detected in Exp. 2. However, this result is likely due to the controlled sampling for this experiment, where participants all scored below the mean of their General Chemistry class on the BRI pre-test.

9.3. The Effect of Reading Comprehension Skill

Reading comprehension skill was found to be a meaningful positive predictor of post-test score across Exps. 1A–1D (with a small effect size) but not in Exp. 2. Again, there was sufficient overlap among all four *sr* confidence intervals between reading skill and post-test scores that the power of the statistic was enhanced by comparing the effect from each study. So, although the effect of reading comprehension skill was not meaningful in half the exploratory studies, the confidence intervals computed at 95% confidence overlapped to a degree where the probability of not finding an effect when there is an effect decreases. The interpretation from this work is that although a correlation between reading skill and post-test score was found to be not meaningful in individual studies, the probability of committing a Type II error may be large for each individual experiment. However, when the results of the studies are pooled, the effect of reading comprehension skill emerges as significant, as the probability of committing a Type II error was reduced via meta-analysis.

Previous studies have shown a positive relationship between science achievement and reading skill [5,71,73,74]. This study contributes to those conclusions, showing a positive, though modest, effect of reading comprehension ability on performance on chemistry concept inventories. A model of comprehension ability put forward by Gernsbacher and colleagues [75] posits that students who are more adept at comprehending text can more easily make connections between ideas while suppressing errant or irrelevant information and thus can begin to structure understanding of concepts more efficiently. Readers with lower comprehension skill can be hindered by the language used in the text (or on assessments, as has been summarized for English language learners by Lee and Orgill [76]) and will expend more effort to build connections between topics and concepts.

In Exp. 2, reading comprehension skill was not a statistically significant predictor of BRI post-test scores. This is likely because the effect of GMRT score on BRI post-test scores was opposite for different subgroups of students. Given the negative interaction observed between BRI pre-test score and GMRT score in this experiment, the relationship between reading comprehension and BRI post-test score was positive for students of lower prior knowledge, while it was negative for students of higher prior knowledge. This compensatory effect of reading comprehension skill for prior knowledge has been observed in other work [74], including General Chemistry [5]. Interestingly, a meaningful, though positive, interaction of GMRT score on post-test scores was observed in Exp. 1A. We speculate that this is due to the controlled sampling of Exp. 2, where only students of lower-than-average prior knowledge were invited to participate.

9.4. Expertise Reversal: The Interaction Between Prior Knowledge and Reading a Relevant Text

A key variable in all experiments of this study was whether a text related to the General Chemistry concept being assessed was read. While reading a related text was found to be a meaningful positive predictor of post-test score across Exps. 1A–1D, the overall effect emerging from the metanalysis was small: $sr = 0.10$, 95% CI [0.04, 0.17]. This small main effect of reading a related text was surprising; after all, how else could students perform better on the post-test versus the pre-test if not for reading a text related to the concepts being tested? This small effect may be due to the relationship between reading a related text and post-test score being different for different subgroups of students. Indeed, the interaction between pre-test z-score and reading a related text, though not found to be statistically significant at 95% confidence, begins to emerge as a possibly meaningful predictor in the meta-analysis. The 95% CI for this effect [−0.12, 0.006] passes $sr = 0$ only slightly. With more measurements, this effect may become statistically significant at 95% confidence; thus, the implication of such an interaction warrants discussion.

The meta-analysis of Exps. 1A–1D suggests that prior knowledge may differentially affect students engaging with a related text passage. Students who perform lower than average on the pre-test and then read the text passage about the chemistry topic will have a small gain in their post-test score, whereas higher performing students who read the text passage will have a small reduction in their post-test score, controlling for all other variables. Though the interaction does not have a large enough effect ($sr = -0.06$) for reading the text to completely compensate for low prior knowledge, the implication is that reading a text passage about a concept can aid learning for students disadvantaged by lower prior knowledge. Additional evidence for this implication is provided by Exp. 2, where for a sample of low prior knowledge students, reading a related text had a large effect ($sr = 0.52$; Figure 2) on post-test score when other variables were controlled statistically.

This effect (termed expertise reversal) has been observed in previous studies [77,78]. In the context of cognitive load theory, which describes how working memory load can affect a learner's ability to process and retain information [9,79], expertise reversal refers to when an instructional strategy that is beneficial for novice learners becomes ineffective or even detrimental to more experienced learners. This distinction can be understood by comparing intrinsic and extraneous cognitive load—essentially, what is considered "intrinsic" for a novice can become "extraneous" for those with more background knowledge. Text passages contain new vocabulary, examples, definitions, and arguments that the reader must navigate to process the main concepts presented. Highly cohesive and redundant materials are more useful for novice learners, but those with high prior knowledge may experience a higher cognitive load when encountering extraneous information (i.e., that which they already know) [13,80,81]. Students with high prior knowledge can experience higher extraneous cognitive load because of redundancies in the text and therefore may be less engaged with the text. So, to lessen cognitive load, they may have the tendency to miss new information upon reading. Students with low prior knowledge will experience these redundancies as intrinsic cognitive load and may then be able to retain more nuanced information from the cohesive text. Thus, the characteristics of a text passage need to be designed with the level of prior knowledge of students in mind. The text passages in this study were written such that the readability dimensions matched the average characteristics of passages found in popular General Chemistry textbooks (Table 2) [6]. However, there was a very high level of referential cohesion in both text passages relative to traditional chemistry textbook passages. This level of cohesion may have benefitted those with low prior knowledge more than those with high prior knowledge in Exps. 1A–1D, leading to the small expertise reversal effect that begins to emerge from the data upon meta-analysis. For the students of lower prior knowledge in Exp. 2, however, the effect of reading a related text was unequivocal.

10. Conclusions

Regarding research question 1, our findings suggest that reading a related text like that found in a General Chemistry textbook promotes learning (experimentally operationalized as gains measured via concept inventories) only to a small extent. However, regarding research question 2, learning from text resources appears to be moderated by individual characteristics: namely, prior knowledge. Across the four experiments comprising our meta-analysis, prior knowledge had the strongest relationship with post-test scores (a medium effect size). An expertise reversal effect, however, may have a small effect on post-test outcomes; this is an interaction between prior knowledge and engaging with a related text. This effect, though not statistically significant at 95% confidence in the meta-analysis, is implied in the results using a sample of students of low prior knowledge. Thus, our findings suggest that the learning disadvantages experienced by students with lower prior knowledge can be lessened when they read highly cohesive text passages. While the expertise reversal effect has been documented previously, the likelihood of its presentation in this study is notable, given its use of learning materials that are similar to those found in common General Chemistry textbooks.

11. Limitations

A limitation of this study was the very high level of cohesion found in both concept texts. This text characteristic may have enhanced an expertise reversal effect that may not be present with all General Chemistry texts. However, the high cohesion may have highlighted the expertise reversal effect, which may not have otherwise been detected. Future studies should include different text types to further investigate whether text cohesion differently affects students based on level of prior knowledge.

Another limitation is comparing the GMRT results to learning outcomes based on reading. The text passages used in this study have very low narrativity and are more expository in nature. However, more passages in the 7/9th grade GMRT test were found to be narrative in nature than expository [82]. It is likely that the 10/12th grade version of the GMRT follows the same trend, and so using a reading comprehension measure that focuses more on narrative text to analyze the relationship between reading skill and learning using an expository text may limit the interpretation of results.

12. Implications

12.1. Chemistry Textbook Organization

Before designing the simulated texts used for this study, several textbooks were surveyed to find short text passages about bonding representations and redox concepts. However, no textbook was found to have concise and centralized text about these concepts. Textbooks often introduce ideas pertaining to a concept in different chapters throughout the text. The results of this work imply that this type of organization (which lacks referential cohesion) may make it difficult for students to make the connections between the ideas, unless the student has a high level of prior knowledge. Including more global summaries of topics in university-level chemistry textbooks may not only help students with lower prior knowledge but also help those students who are looking to use the textbook as a resource for review.

12.2. Expertise Reversal and Text Cohesion

The implications of an expertise reversal effect using text to learn about a concept are that individuals must be aware of their own characteristics to choose the best and most effective materials for studying. If the same study materials are provided to all students, both high prior knowledge and low prior knowledge learners would suffer. Having a variety of texts that differ in text cohesion would be one way to provide the optimal learning opportunities for a wide range of students. It must be noted, however, that for a student to choose the best material, they must be aware of their level of knowledge about the concept. Providing an opportunity for students to take a pre-test or to complete a concept inventory to assess their level of prior knowledge would be necessary.

12.3. Meta-Analysis as a Research Tool

The power of meta-analysis to increase the statistical power of effects has been used for large-scale studies that include tens or hundreds of literature sources. However, the technique is not limited to such a large scale. In this study, only four experiments were compared to find statistically significant effects that would have been lost in the noise of variance without the meta-analysis. Though a researcher may be able to spot meaningful trends in data, such trends do not always pan out to be statistically significant when analyzed through traditional methods. By effectively increasing the overall sample size, statistical power can be increased, permitting otherwise non-significant effects to emerge. This can be a particularly powerful tool when variables have small effect sizes and when several studies have disparate results. In addition, the method of using meta-analysis emphasizes the importance of effect size over p values when evaluating the significance of effects. In very large samples, predictors can be considered statistically significant even when contributing very little to the magnitude of the outcome. However, effect sizes measured by Pearson's r, Cohen's d, or semipartial correlations display more clearly the magnitude of difference of means or the magnitude of the phenomenon.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/educsci14121287/s1: Section S1: Demographic data for each experiment (Tables S1 and S2); Section S2: Descriptive statistics for reading comprehension skills scores (Table S3); Section S3: Descriptive statistics for bonding representations inventory pre-test (Table S4); Section S4: Descriptive statistics for BRI pre-test scores (Table S5); Section S5: Descriptive statistics for BRI post-test scores (Table S6); Section S6: Descriptive statistics for ROXCI post-test scores (Table S7); Section S7: Summary of regression results for experiments 1A–1D; Section S8: Summary of meta-analysis results for experiments 1A–1D.

Author Contributions: Conceptualization, S.P.; methodology, R.W.B. and S.P.; validation, R.W.B. and S.P.; formal analysis, R.W.B. and S.P.; investigation, R.W.B.; resources, R.W.B.; data curation, R.W.B.; writing—original draft, R.W.B.; writing—review and editing, S.P.; visualization, R.W.B. and S.P.; supervision, S.P.; project administration, R.W.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The Institutional Review Board for the Protection of Human Subjects in Research (IRB) of the University of New Hampshire reviewed and approved this study as Exempt as described in Title 45, Code of Federal Regulations, Part 46, Subsection 101(b) (IRB #6072, 4 September 2014).

Informed Consent Statement: All subjects gave their informed consent for inclusion before they participated in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors would like to thank the teaching assistants who cooperated with us in mounting these experiments in their laboratory sections.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Text Passages Developed for This Study

Bonding representations. The text about bonding representations included brief discussions about several topics that students are often found to have misconceptions about. These topics include the ideas that bonds are not necessarily purely ionic or covalent, that polar covalent bonds exist because of differences in electronegativities of atoms bound together, that there is a tendency for metals to become cations and nonmetals to become anions by losing or gaining electrons, and that covalent compounds tend to form discrete molecules, while ionic compounds form extended 3D arrays of alternating ions. The reasoning for including these ideas in the text is that it is hypothesized that if a student holds a misconception about a topic prior to reading the text passage, the student may start to overcome the misconception if confronted with the scientifically accepted understanding of the topic while reading [83,84]. Some of the items in the BRI address those same common misconceptions.

Redox concepts. The text about redox concepts included discussions about the difference between oxidation numbers and charge, the movement of electrons from one species to another, and the definition of reductants and oxidants. These ideas were included because students have been found to hold misconceptions about these topics. The ROXCI includes some items that refer to these concepts.

Appendix A.1. Text Passage on Bonding Representations

In order to make compounds, atoms and ions come together to form chemical bonds. The components of a chemical bond (i.e., atoms or ions) interact via electrostatic attractions, with positively charged particles attracted to negatively charged particles. The bonds of most compounds fall into one of two categories: ionic and covalent. It must be noted, however, that purely covalent and purely ionic interactions are two extreme cases of chemical bonding, and in reality, all compounds exhibit features of both types of bonding to some degree.

As suggested by the name, ions form ionic bonds. In ionic species, oppositely charged ions are attracted to each other and form a solid three dimensional array (or lattice) of particles. Metals, with low ionization energies, have a tendency to lose electrons; nonmetals, with high electron affinities, have a tendency to gain electrons. Once a metal has ionized, it becomes a positively charged cation, while a nonmetal becomes a negatively charged anion. When a collection of oppositely-charged ions are close to each other, they will be attracted by electrostatic forces, and ionic bonds will form. The resulting solid is made up of an ordered arrangement of ions interacting with one another—each cation is attracted to several neighboring anions, and each anion interacts with several neighboring cations. The interaction between ions in an array is not limited to just monoatomic ions. Ionic bonds can also be generalized to include ions that are made up of several atoms, or polyatomic ions.

Unlike ionic bonding cases, none of the atoms involved in a covalent bond have gained or lost electrons. Instead, covalent compounds are composed of neutral atomic species and not ions. A covalent bond can be described as a "sharing" of electrons between the nuclei of two atoms, and compounds composed of covalent bonds are commonly called "molecules". Molecules are discreet networks of covalently bonded atoms that range in size according to how many atoms comprise the compound; only in very rare cases do covalent compounds contain enough atoms to rival the infinite lattice structures of ionic compounds.

Though electrons involved in a covalent bond are localized between two atoms, the extent to which the electrons are "shared" varies. That is, one of the atoms in the bond may attract the electrons more to itself because the atom possesses a greater electronegativity relative to the other atom involved in the bond. So, electrons between atoms of unequal electronegativity may in fact lie closer to the more electronegative atom. In this case, the electrons are still shared between the two nuclei, but in an unequal manner; this scenario is known as a polar covalent bond. If the covalently bonded atoms have the same electronegativity value, the electrons between them will be shared equally.

Appendix A.2. Text Passage on Redox Concepts

Many chemical reactions can be categorized as one of three major classes: Precipitation, acid-base, or oxidation-reduction (redox) reactions. This passage will focus on the latter class.

Redox reactions involve the transfer of electrons between chemical species, and these processes can occur among ionic or covalently bonded compounds as well as pure elemental species. Electron transfer occurs in many important applications, including batteries, combustion, photosynthesis, electroplating, and cellular respiration.

In order to identify a reaction as a redox process, electrons must move from one reagent to another. A bookkeeping method has been devised by chemists to keep track of whether an electron has been "gained" or "lost" by an atom in a reaction by assigning oxidation numbers (or oxidation states) to individual atoms. This is not to be confused with assignment of charge to an atom or molecule. The oxidation state is not a true physical charge which can be measured, but instead is the theoretical charge the atom would have if the atom was ionically bonded with the other atoms in the species. This scheme is followed for molecules exhibiting the whole spectrum of bonding, from purely ionic to purely covalent, and is used simply for the ease of keeping track of electrons.

Atoms in their pure elemental form have an oxidation number of zero, but if the atoms are charged ions or are bound to other atoms, then a positive or negative oxidation number can be assigned. The oxidation number of an atom in its ionic form will be the same as the charge the ion carries. If the atom is a group 1 metal, it will have an oxidation number of +1, and atoms in group 2 will have an oxidation number of +2. When the atom is a halogen, the oxidation number for that atom will be −1. Oxygen will usually have an oxidation number of −2, while hydrogen typically is assigned to an oxidation number of +1. The sum of all the oxidation numbers assigned to atoms in a compound must be equal to the overall charge of the compound. For example, in phosphate (PO_4^{3-}) the oxygen atoms are each assigned to a −2 O.N. This will make the phosphorous atom have a +5 oxidation number so that overall the phosphate ion will have a 3− charge.

When a redox process occurs, the oxidation states for at least two atoms will change during the course of the reaction. If the oxidation state of an atom increases to become more positive (or less negative) after the reaction, then the species containing that atom will have been oxidized (or there will have been a loss of electrons from that species). If the oxidation state of the atom becomes less positive (or more negative), then that species has been reduced, or has gained electrons during the reaction. Commonly, oxidation is referred to as a loss of electrons, while reduction is referred to as a gain of electrons. The species which undergoes oxidation is called the "reductant" or "reducing agent", and the species which is reduced is called the "oxidant" or "oxidizing agent". The reductant and oxidant work together in a redox reaction.

The chemical equation must be balanced to determine the number of electrons transferred between the reductant and oxidant. Balancing a redox reaction involves splitting the chemical equation into two half-reactions, where one half-reaction describes the oxidation process, and the other half-reaction describes the reduction process. The electrons transferred in each half reaction are found independently, and then the total number of electrons involved in the overall redox process is the common multiple of the number of electrons involved in the separate half-reactions.

Appendix B. Discussion of the Assumptions of Linear Regression

Five main assumptions underlying multiple regressions should be considered when analyzing data using this technique (in order of decreasing importance, according to Gelman and Hill): (1) validity; (2) additivity and linearity; (3) independence of errors; (4) equal variance of errors; and (5) normality of errors [60] (pp. 45–47). While data used in empirical research rarely meet all these criteria precisely, addressing the extent to which a study satisfies these assumptions lends trustworthiness to a study's findings. In this section, each of these assumptions is discussed in the context of the present study.

Appendix B.1. Validity

In the case of regression, validity refers to the extent to which the data being analyzed map to the research questions framing the study. In general, this study takes on an "Is it the intervention or the students?" [85] orientation. By this, we mean that our research questions frame an investigation of whether student performance on an outcome measure is related more to an intervention (reading a related text) or characteristics of the student (reading comprehension skill and prior knowledge) and (if performance on the outcome measure is related to the intervention) whether that effect is different for different student characteristics. Thus, the nature of the study aligns well with employing multiple linear regression as an analytic technique. Additionally, there exists previous empirical evidence [5,71,74] for the main predictors employed in our analyses (prior knowledge and reading comprehension skill). Indeed, while other predictors may be "relevant", prior knowledge and reading comprehension skill were most germane to addressing the research questions posed by the study. On the other hand, regarding the outcome variable, a model using post-test scores on a concept inventory will provide limited insight into the complex construct of student "learning". However, given the general constraints of quantitative work, a pre-/post-test design provided one possible way to address the research questions posed.

Appendix B.2. Additivity and Linearity

The fundamental assumption of a multiple linear regression model is that the dependent variable is described by linear functions of separate predictor variables. There is no compelling reason (other than convenience) to assume linear relationships between the outcome variable (post-test scores) and predictor variables in this work. However, previous work has provided evidence for a linear relationship between "achievement" in science and both reading comprehension skill and prior knowledge [5,71,74]. Thus, assuming linearity between the outcome and predictor variables in this study has been satisfied empirically, although (to our knowledge) no work has tested non-linear relationships among these variables. Additionally, inspection of scatterplots of the dependent variable and each independent variable revealed roughly linear relationships. In models of achievement that only include reading comprehension skill and prior knowledge as predictors, the assumption of additivity is violated; prior knowledge is known to moderate the relationship between reading comprehension skill and achievement [5,74,86]. Thus, in all models presented in this work, interaction terms between predictor variables are included.

Appendix B.3. Independence of Errors

A linear regression model assumes that there is no relationship between the residuals and the predictor variables. To check this assumption, we inspected scatterplots of residuals vs. fits for each regression model. For each experiment, there was no visible relationship between the plotted residuals and fits.

Appendix B.4. Equal Variances of Errors

In a linear regression model, the variance in residuals should be the same across all values of the predictor variables. This assumption can be checked by examining the same "residuals versus fits" scatter plots that were used to check the independence of errors assumption above. For each experiment, there was no visible pattern in these plots that would suggest that the residuals were not consistent across all values of the x-axis.

Appendix B.5. Normality of Errors

This assumption can come into play if there are ceiling effects present in measures employed in the study [85]; that is, if the measure is artificially limiting the higher end of the score distribution. Based on the descriptive statistics presented in Sections S1–S5 of the Supplementary Materials, there were no instances of negative skew in any of the score distributions. Moreover, the means of each data set were quite distant from the maximum score possible for each measure. Thus, there is no evidence of ceiling effects in our analyses. Additionally, this normality of errors assumption can be checked by examining histograms of residuals for each regression analysis. Doing so revealed that for each of the five experiments, the residuals were roughly normal in distribution.

References

1. Thompson, B.; Bunch, Z.; Popova, M. A Review of Research on the Quality and Use of Chemistry Textbooks. *J. Chem. Educ.* **2023**, *100*, 2884–2895. [CrossRef]
2. Tsaparlis, G.; Pappa, E.T.; Byers, B. Teaching and Learning Chemical Bonding: Research-Based Evidence for Misconceptions and Conceptual Difficulties Experienced by Students in Upper Secondary Schools and the Effect of an Enriched Text. *Chem. Educ. Res. Pract.* **2018**, *19*, 1253–1269. [CrossRef]
3. Gegios, T.; Salta, K.; Koinis, S. Investigating High-School Chemical Kinetics: The Greek Chemistry Textbook and Students' Difficulties. *Chem. Educ. Res. Pract.* **2017**, *18*, 151–168. [CrossRef]
4. Nimmermark, A.; Öhrström, L.; Mårtensson, J.; Davidowitz, B. Teaching of Chemical Bonding: A Study of Swedish and South African Students' Conceptions of Bonding. *Chem. Educ. Res. Pract.* **2016**, *17*, 985–1005. [CrossRef]
5. Pyburn, D.T.; Pazicni, S.; Benassi, V.A.; Tappin, E.E. Assessing the Relation between Language Comprehension and Performance in General Chemistry. *Chem. Educ. Res. Pract.* **2013**, *14*, 524–541. [CrossRef]
6. Pyburn, D.T.; Pazicni, S. Applying the Multilevel Framework of Discourse Comprehension To Evaluate the Text Characteristics of General Chemistry Textbooks. *J. Chem. Educ.* **2014**, *91*, 778–783. [CrossRef]
7. Ozuru, Y.; Dempsey, K.; McNamara, D.S. Prior Knowledge, Reading Skill, and Text Cohesion in the Comprehension of Science Texts. *Learn. Instr.* **2009**, *19*, 228–242. [CrossRef]
8. Baddeley, A. Working Memory: Looking Back and Looking Forward. *Nat. Rev. Neurosci.* **2003**, *4*, 829–839. [CrossRef]
9. Sweller, J. Cognitive Load During Problem Solving: Effects on Learning. *Cog. Sci.* **1988**, *12*, 257–285. [CrossRef]
10. Sweller, J.; van Merriënboer, J.J.G.; Paas, F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31*, 261–292. [CrossRef]
11. Ericsson, K.A.; Kintsch, W. Long-Term Working Memory. *Psychol. Rev.* **1995**, *102*, 211–245. [CrossRef] [PubMed]

12. Graesser, A.C.; McNamara, D.S.; Louwerse, M.M. What Do Readers Need to Learn in Order to Process Coherence Relations in Narrative and Expository Text. In *Rethinking Reading Comprehension*; Sweet, A.P., Snow, C.E., Eds.; Solving Problems in the Teaching of Literacy; Guilford Press: New York, NY, USA, 2003; pp. 82–98.
13. McNamara, D.S.; Kintsch, E.; Songer, N.B.; Kintsch, W. Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning From Text. *Cogn. Instr.* **1996**, *14*, 1–43. [CrossRef]
14. Smith, R.; Snow, P.; Serry, T.; Hammond, L. The Role of Background Knowledge in Reading Comprehension: A Critical Review. *Read. Psychol.* **2021**, *42*, 214–240. [CrossRef]
15. Bodner, G.M. Constructivism: A Theory of Knowledge. *J. Chem. Educ.* **1986**, *63*, 873. [CrossRef]
16. Elby, A. What Students' Learning of Representations Tells Us about Constructivism. *J. Math. Behav.* **2000**, *19*, 481–502. [CrossRef]
17. Rodriguez, J.-M.G.; Towns, M.H. Analysis of Biochemistry Students' Graphical Reasoning Using Misconceptions Constructivism and Fine-Grained Constructivism: Why Assumptions about the Nature and Structure of Knowledge Matter for Research and Teaching. *Chem. Educ. Res. Pract.* **2021**, *22*, 1020–1034. [CrossRef]
18. McCloskey, M. Naive Theories of Motion. In *Mental Models*; Gentner, D., Stevens, A., Eds.; Erlbaum: Hillsdale, NJ, USA, 1983; pp. 299–324.
19. McCloskey, M. Intuitive Physics. *Sci. Am.* **1983**, *249*, 122–130. [CrossRef]
20. Clement, J. Students' Preconceptions in Introductory Mechanics. *Am. J. Phys.* **1982**, *50*, 66–71. [CrossRef]
21. Hawkins, D. Critical Barriers to Science Learning. *Outlook* **1978**, *29*, 3–23.
22. Strike, K.A.; Posner, G.J. A Conceptual Change View of Learning and Understanding. In *Cognitive Structure and Conceptual Change*; West, L.H.T., Pines, A.L., Eds.; Academic Press: New York, NY, USA, 1985; pp. 221–231.
23. Posner, G.J.; Strike, K.A.; Hewson, P.W.; Gertzog, W.A. Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change. *Sci. Educ.* **1982**, *66*, 211–227. [CrossRef]
24. Wang, T.; Andre, T. Conceptual Change Text versus Traditional Text and Application Questions versus No Questions in Learning about Electricity. *Contemp. Educ. Psychol.* **1991**, *16*, 103–116. [CrossRef]
25. Tippett, C.D. Refutation Text In Science Education: A Review Of Two Decades Of Research. *Int. J. Sci. Math Educ.* **2010**, *8*, 951–970. [CrossRef]
26. Ültay, N.; Durukan, Ü.G.; Ültay, E. Evaluation of the Effectiveness of Conceptual Change Texts in the REACT Strategy. *Chem. Educ. Res. Pract.* **2015**, *16*, 22–38. [CrossRef]
27. diSessa, A.A. A Friendly Introduction to "Knowledge in Pieces": Modeling Types of Knowledge and Their Roles in Learning. In *Invited Lectures from the 13th International Congress on Mathematical Education*; Kaiser, G., Forgasz, H., Graven, M., Kuzniak, A., Simmt, E., Xu, B., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 65–84.
28. diSessa, A.A.; Sherin, B.L.; Levin, M. Knowledge Analysis: An Introduction. In *Knowledge and Interaction*; diSessa, A.A., Levin, M., Brown, N.S.J., Eds.; Routledge: New York, NY, USA, 2015; ISBN 978-1-315-75736-0.
29. Hammer, D.; Elby, A.; Scherr, R.; Redish, E.F. Resources, Framing, and Transfer. In *Transfer of Learning from a Modern Multidisciplinary Perspective*; Mestre, J.P., Ed.; Information Age Publishing: Greenwich, CT, USA, 2006; pp. 89–119.
30. Barke, H.-D.; Hazari, A.; Yitbarek, S. *Misconceptions in Chemistry*; Springer: Berlin/Heidelberg, Germany, 2009; ISBN 978-3-540-70988-6.
31. Taber, K.; Pack, M.J. *Chemical Misconceptions: Prevention, Diagnosis and Cure: Classroom Resources, Volume 2*; Royal Society of Chemistry: London, UK, 2002; ISBN 978-0-85404-381-1.
32. Taber, K. *Chemical Misconceptions: Prevention, Diagnosis and Cure: Theoretical Background, Volume 1*; Royal Society of Chemistry: London, UK, 2002; ISBN 978-0-85404-386-6.
33. Rodriguez, J.-M.G. Using Analytic Autoethnography to Characterize the Variation in the Application of the Resources Framework: What Is a Resource? *J. Chem. Educ.* **2024**, *101*, 3676–3690. [CrossRef]
34. Robertson, A.D.; Atkins, L.J.; Levin, D.M.; Richards, J. What Is Responsive Teaching? In *Responsive Teaching in Science and Mathematics*; Robertson, A.D., Scherr, R.E., Hammer, D., Eds.; Teaching and Learning in Science; Routledge: New York, NY, USA, 2016; pp. 1–35.
35. Sherin, M.G.; Russ, R.S. Accessing Mathematics Teachers' In-the-Moment Noticing. In *Mathematics Teacher Noticing*; Routledge: London, UK, 2011; pp. 109–124.
36. Carpenter, T.P.; Fennema, E.; Peterson, P.L.; Chiang, C.-P.; Loef, M. Using Knowledge of Children's Mathematics Thinking in Classroom Teaching: An Experimental Study. *Am. Educ. Res. J.* **1989**, *26*, 499–531. [CrossRef]
37. Sands, D.; Parker, M.; Hedgeland, H.; Jordan, S.; Galloway, R. Using Concept Inventories to Measure Understanding. *High. Educ. Pedagog.* **2018**, *3*, 173–182. [CrossRef]
38. Peterson, R.; Treagust, D.; Garnett, P. Identification of Secondary Students' Misconceptions of Covalent Bonding and Structure Concepts Using a Diagnostic Instrument. *Res. Sci. Educ.* **1986**, *16*, 40–48. [CrossRef]
39. Zoller, U. Students' Misunderstandings and Misconceptions in College Freshman Chemistry (General and Organic). *J. Res. Sci. Teach.* **1990**, *27*, 1053–1065. [CrossRef]
40. Taber, K.S. Student Understanding of Ionic Bonding: Molecular versus Electrostatic Framework? *Sch. Sci. Rev.* **1997**, *78*, 85–95.
41. Harrison, A.G.; Treagust, D.F. Learning about Atoms, Molecules, and Chemical Bonds: A Case Study of Multiple-Model Use in Grade 11 Chemistry. *Sci. Educ.* **2000**, *84*, 352–381. [CrossRef]

42. Özmen, H. Some Student Misconceptions in Chemistry: A Literature Review of Chemical Bonding. *J. Sci. Educ. Tech.* **2004**, *13*, 147–159. [CrossRef]
43. Garnett, P.J.; Treagust, D.F. Conceptual Difficulties Experienced by Senior High School Students of Electrochemistry: Electric Circuits and Oxidation-Reduction Equations. *J. Res. Sci. Teach.* **1992**, *29*, 121–142. [CrossRef]
44. de Jong, O.; Acampo, J.; Verdonk, A. Problems in Teaching the Topic of Redox Reactions: Actions and Conceptions of Chemistry Teachers. *J. Res. Sci. Teach.* **1995**, *32*, 1097–1110. [CrossRef]
45. Stains, M.; Talanquer, V. Classification of Chemical Reactions: Stages of Expertise. *J. Res. Sci. Teach.* **2008**, *45*, 771–793. [CrossRef]
46. Österlund, L.-L.; Ekborg, M. Students' Understanding of Redox Reactions in Three Situations. *Nord. Stud. Sci. Educ.* **2009**, *5*, 115–127. [CrossRef]
47. Barke, H.D. Two Ideas of the Redox Reaction: Misconceptions and Their Challenge in Chemistry Education. *Afr. J. Chem. Educ.* **2012**, *2*, 32–50.
48. Brandriet, A.R.; Bretz, S.L. Measuring Meta-Ignorance through the Lens of Confidence: Examining Students' Redox Misconceptions about Oxidation Numbers, Charge, and Electron Transfer. *Chem. Educ. Res. Pract.* **2014**, *15*, 729–746. [CrossRef]
49. Graesser, A.C.; McNamara, D.S.; Louwerse, M.M.; Cai, Z. Coh-Metrix: Analysis of Text on Cohesion and Language. *Beh. Res. Meth. Instr. Comp.* **2004**, *36*, 193–202. [CrossRef]
50. McNamara, D.S.; Graesser, A.C.; McCarthy, P.M.; Cai, Z. *Automated Evaluation of Text and Discourse with Coh-Metrix*; Cambridge University Press: Cambridge, UK, 2014; ISBN 978-0-521-19292-7.
51. MacGinitie, W.H.; MacGinitie, R.K.; Maria, K.; Dreyer, L.G.; Hughes, K.E. *Gates MacGinitie Reading Test Fourth Edition Forms S and T-Paper-Pencil*; Riverside: Chicago, IL, USA, 2000.
52. MacGinitie, W.H.; MacGinitie, R.K.; Maria, K.; Dreyer, L.G.; Hughes, K.E. *Gates MacGinitie Reading Tests Fourth Edition Levels 7/9 & 10/12 Forms S&T: Manual for Scoring and Interpretation*; Riverside: Chicago, IL, USA, 2000.
53. Luxford, C.J.; Bretz, S.L. Development of the Bonding Representations Inventory To Identify Student Misconceptions about Covalent and Ionic Bonding Representations. *J. Chem. Educ.* **2014**, *91*, 312–320. [CrossRef]
54. Brandriet, A.R.; Bretz, S.L. The Development of the Redox Concept Inventory as a Measure of Students' Symbolic and Particulate Redox Understandings and Confidence. *J. Chem. Educ.* **2014**, *91*, 1132–1144. [CrossRef]
55. Charness, G.; Gneezy, U.; Kuhn, M.A. Experimental Methods: Between-Subject and within-Subject Design. *J. Econ. Behav. Organ.* **2012**, *81*, 1–8. [CrossRef]
56. Warner, R.M. *Applied Statistics: From Bivariate Through Multivariate Techniques*, 2nd ed.; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2012; ISBN 978-1-4129-9134-6.
57. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, 6th ed.; Pearson: Boston, MA, USA, 2013; ISBN 978-0-205-84957-4.
58. Kim, H.-Y. Statistical Notes for Clinical Researchers: Assessing Normal Distribution (2) Using Skewness and Kurtosis. *Restor. Dent. Endod.* **2013**, *38*, 52–54. [CrossRef]
59. Evans, M.G. A Monte Carlo Study of the Effects of Correlated Method Variance in Moderated Multiple Regression Analysis. *Organ. Behav. Hum. Decis. Process.* **1985**, *36*, 305–323. [CrossRef]
60. Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*; Analytical Methods for Social Research; Cambridge University Press: Cambridge, UK, 2006; ISBN 9780511790942.
61. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Routledge: New York, NY, USA, 1988; ISBN 978-0-203-77158-7.
62. Glass, G.V. Primary, Secondary, and Meta-Analysis of Research. *Educ. Res.* **1976**, *5*, 3–8. [CrossRef]
63. Fitz-Gibbon, C.T. Meta-Analysis: An Explication. *Br. Educ. Res.* **1984**, *10*, 135–144. [CrossRef]
64. Rosenthal, R.; DiMatteo, M.R. Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. *Annu. Rev. Psychol.* **2001**, *52*, 59–82. [CrossRef] [PubMed]
65. Cumming, G. *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*; Routledge: New York, NY, USA, 2011; ISBN 978-0-203-80700-2.
66. Borenstein, M.; Hedges, L.V.; Higgins, J.P.T.; Rothstein, H.R. Criticisms of Meta-Analysis. In *Introduction to Meta-Analysis*; John Wiley & Sons, Ltd.: New York, NY, USA, 2009; pp. 377–387, ISBN 978-0-470-74338-6.
67. Dettori, J.R.; Norvell, D.C.; Chapman, J.R. Fixed-Effect vs Random-Effects Models for Meta-Analysis: 3 Points to Consider. *Glob. Spine J.* **2022**, *12*, 1624–1626. [CrossRef]
68. Cumming, G.; Calin-Jageman, R. Esci (Estimation Statistics with Confidence Intervals). Available online: https://thenewstatistics.com/itns/esci/ (accessed on 26 August 2024).
69. Akbaşlı, S.; Şahin, M.; Yaykiran, Z. The Effect of Reading Comprehension on the Performance in Science and Mathematics. *J. Educ. Pract.* **2016**, *7*, 108.
70. Seery, M.K. The Role of Prior Knowledge and Student Aptitude in Undergraduate Performance in Chemistry: A Correlation-Prediction Study. *Chem. Educ. Res. Pract.* **2009**, *10*, 227–232. [CrossRef]
71. Pyburn, D.T.; Pazicni, S.; Benassi, V.A.; Tappin, E.M. The Testing Effect: An Intervention on Behalf of Low-Skilled Comprehenders in General Chemistry. *J. Chem. Educ.* **2014**, *91*, 2045–2057. [CrossRef]
72. Shapiro, A.M. How Including Prior Knowledge As a Subject Variable May Change Outcomes of Learning Research. *Am. Ed. Res. J.* **2004**, *41*, 159–189. [CrossRef]
73. Reed, D.K.; Petscher, Y.; Truckenmiller, A.J. The Contribution of General Reading Ability to Science Achievement. *Read. Res. Q.* **2017**, *52*, 253–266. [CrossRef]

74. O'Reilly, T.; McNamara, D.S. The Impact of Science Knowledge, Reading Skill, and Reading Strategy Knowledge on More Traditional "High-Stakes" Measures of High School Students' Science Achievement. *Am. Educ. Res. J.* **2007**, *44*, 161–196. [CrossRef]
75. Gernsbacher, M.A.; Varner, K.R.; Faust, M.E. Investigating Differences in General Comprehension Skill. *J. Exp. Psychol. Learn. Mem. Cogn.* **1990**, *16*, 430–445. [CrossRef]
76. Lee, E.N.; Orgill, M. Toward Equitable Assessment of English Language Learners in General Chemistry: Identifying Supportive Features in Assessment Items. *J. Chem. Educ.* **2022**, *99*, 35–48. [CrossRef]
77. Kalyuga, S. Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction. *Educ. Psychol. Rev.* **2007**, *19*, 509–539. [CrossRef]
78. Kalyuga, S.; Renkl, A. Expertise Reversal Effect and Its Instructional Implications: Introduction to the Special Issue. *Instr. Sci.* **2010**, *38*, 209–215. [CrossRef]
79. Sweller, J. Cognitive Load Theory, Learning Difficulty, and Instructional Design. *Learn. Instr.* **1994**, *4*, 295–312. [CrossRef]
80. O'Reilly, T.; Mcnamara, D.S. Reversing the Reverse Cohesion Effect: Good Texts Can Be Better for Strategic, High-Knowledge Readers. *Discourse Process.* **2007**, *43*, 121–152. [CrossRef]
81. McNamara, D.S.; Kintsch, W. Learning from Texts: Effects of Prior Knowledge and Text Coherence. *Discourse Process.* **1996**, *22*, 247–288. [CrossRef]
82. Rowe, M.; Ozuru, Y.; McNamara, D. An Analysis of Standardized Reading Ability Tests: What Do Questions Actually Measure? In Proceedings of the 7th International Conference on Learning Sciences, Bloomington, IN, USA, 27 June–1 July 2006; International Society of the Learning Sciences: Bloomington, IN, USA, 2006; pp. 627–633.
83. Kenedeou, P.; Rapp, D.N.; van den Broek, P. The Influence of Readers' Prior Knowledge on Text Comprehension and Learning from Text. In *Progress in Education*; Nata, R., Ed.; Nova Science Publishers, Inc.: New York, NY, USA, 2003; Volume 13, pp. 189–209, ISBN 1-59454-090-X.
84. Nakhleh, M.B. Why Some Students Don't Learn Chemistry: Chemical Misconceptions. *J. Chem. Educ.* **1992**, *69*, 191. [CrossRef]
85. Theobald, R.; Freeman, S. Is It the Intervention or the Students? Using Linear Regression to Control for Student Characteristics in Undergraduate STEM Education Research. *CBE Life Sci. Educ.* **2014**, *13*, 41–48. [CrossRef] [PubMed]
86. Adams, B.C.; Bell, L.C.; Perfetti, C.A. A Trading Relationship between Reading Skill and Domain Knowledge in Children's Text Comprehension. *Discourse Process.* **1995**, *20*, 307–323. [CrossRef]

education sciences

MDPI

Article

Implementing Mastery Grading in Large Enrollment General Chemistry: Improving Outcomes and Reducing Equity Gaps

Joshua D. Hartman * and Jack F. Eichler *

Department of Chemistry, University of California, Riverside, CA 92521, USA
* Correspondence: joshua.hartman@ucr.edu (J.D.H.); jack.eichler@ucr.edu (J.F.E.)

Abstract: Specifications and mastery grading schemes have been growing in popularity in higher education over the past several years, and reports of specifications grading and other alternative grading systems are emerging in the chemistry education literature. The general goal of these alternative grading approaches is to reduce the reliance on high-stakes exams and give students a more transparent pathway to achieving the course learning outcomes. More importantly, relying less on infrequent high-stakes exams may help reduce historical equity gaps in introductory gateway STEM courses. Herein, we describe the implementation of two versions of mastery grading systems in large enrollment general chemistry courses at a public R1 institution. Class-wide course outcomes, equity gaps in performance on a common final exam, and student feedback on their experience navigating these grading schemes are presented. We show that combining mastery grading with interactive courseware tools improved the average performance on a common final assessment for under-represented minority (URM) students by 7.1 percentage points relative to an active control course that used infrequent high-stakes exams.

Keywords: mastery grading; testing effect; second-change testing; alternative grading

1. Introduction

Mastery grading has been discussed in the education research literature since the 1960s, notably in Bloom's proposal to develop broader mastery learning curricula [1]. Though several meta-analyses corroborate the general efficacy of mastery learning approaches, broader adoption of this assessment approach has not been observed in higher education, and the American higher educational system generally relies on high-stakes exams [2]. Research studies that specifically examine the impact of the mastery outcomes approach in higher education STEM remain limited but are beginning to emerge in the literature [3]. A mastery outcomes structure that utilized second-chance testing in an undergraduate engineering course resulted in significantly improved final exam performance relative to a course that used traditional high-stakes exams, and students in the mastery outcomes course earned twice as many As and half the number of failing grades. Most importantly, traditionally underrepresented students performed on par with non-URM students [3].

Alternative grading and assessment models have been explored across numerous STEM disciplines, including chemistry education [4]. For example, a learner-centered grading method using a standards-based assessment structure for general chemistry has been shown to improve grading transparency. This implementation did not quantify the impact on student learning outcomes, focusing instead on observational data related to the generally positive student learning experience [5]. In another study involving high school chemistry students, mastery learning improved performance and attitude toward learning [6]. Although the research is limited, these studies highlight the generally positive shift in student perspective when moving from a traditional high-stakes grading system to a mastery approach [7,8].

Another outcomes-based assessment approach that has gained recent attention in the chemistry education community is specifications grading [9]. The specifications grading system differs from the mastery outcomes (second chance testing) approach in that specifications grading allows students to demonstrate mastery by completing bundles of assignments or tasks (e.g., a letter grade of A can be earned by completing 9/10 components in the bundle, a letter grade of B can be earned by completing 8/10 components in the bundle, etc.). Within the chemistry education literature, specifications grading has been implemented predominantly in laboratory courses [7,8]. We speculate that this is due to this type of grading structure being well suited to lab courses that focus on skills and completion of tasks [7], and specifications systems can lead to mixed results with respect to student satisfaction [8]. Nevertheless, implementing the specification grading system in a large enrollment organic chemistry laboratory setting notably improved final letter grades [8].

Implementing a specifications grading system in a lecture setting represents a dramatic overhaul in course design from traditional points-based systems, where grades are largely determined by a single attempt, to high-stakes summative assessments [2]. As a result, traditional points-based grading remains the norm for lecture courses despite mounting evidence that points-based models increase student stress levels, decrease equity, and de-emphasize the acquisition of content knowledge [10]. Several innovative strategies have been examined for implementing specifications grading in general chemistry [11], organic chemistry [12–14], analytical chemistry [15], and upper-division chemical biology [16] lecture courses. Many of these studies primarily focus on qualitative aspects of the specifications grading implementations, highlighting reduced self-reported anxiety and generally positive feedback from professors. Hollinsed et al. did report an increase in the conversion of B students to A students. However, the specifications grading model did not significantly impact the number of lower-performing students [11].

The promise of improving the learning environment for students and professors while creating a more equitable grading system continues to motivate the development of alternative grading models. Recently, Noell et al. implemented a hybrid-specs grading system introducing an element of second-chance testing to shift the emphasis toward content mastery without a full course redesign [17]. The hybrid-specs system increased the conversion of B to A grades, but there was a small increase in the DFW rates using this hybrid-specs model. The success of the hybrid-specs model is likely tied to the testing effect. The testing effect is a framework based on research linked to retrieval practice, which is a critical component of the learning process [18]. However, for students to benefit from the second-chance testing model, they must have the skills, resources, and metacognitive strategies required to fill knowledge gaps, thereby improving scores on subsequent tests [19]. For this reason, recent literature suggests that the testing effect can decrease or even disappear as the complexity of the learning materials increases [20]. Yet, when regular testing is coupled with additional tools and resources, student performance gains have been realized for advanced topics directly related to chemistry [21].

Successful engagement with the mastery grading model is closely linked to a student's background, particularly through metacognitive development and familiarity with effective learning strategies. Therefore, comparing student performance across different demographics, such as familial education history, ethnicity, and income, can provide valuable insight into the design and implementation of alternative grading models. Previous reports did not study the specific impacts of these alternative grading strategies on traditionally underrepresented students [3,7,8,12,15,16]. A previous study implementing mastery learning in high school general chemistry did find a particularly pronounced positive effect on learning outcomes and attitudes from students who are struggling with the chemistry content [6]. However, the results were not disaggregated by ethnicity or familial education level.

To the authors' knowledge, the present work represents the first study that assesses the impact of a mastery grading system in large enrollment general chemistry courses relative to an active control course that used traditional high-stakes exams. This work highlights the importance of coupling second-chance testing with interactive courseware designed to promote asynchronous active learning and metacognitive development [22]. Particular emphasis is placed on exploring correlations between a student's familial education history, ethnicity, and socioeconomic status by looking at disaggregated student performance data. This study shows that supporting a mastery grading model with robust interactive courseware improved the average student performance for the entire class population on a common final assessment by 6.9 percentage points relative to a control using infrequent high-stakes exams. The improvement was more pronounced (11.6 percentage points) for first-generation college students with an URM background receiving financial aid.

1.1. Theoretical Frameworks

In this project, we implement a mastery outcomes assessment approach rooted in the theoretical frameworks of the testing effect [18] and mindset theory [23]. The testing effect is linked to the phenomenon of retrieval practice, in which it has been found that the act of retrieving information is, in some cases, a more impactful learning event than an information coding event, where information encoding refers to learning new knowledge [24]. The mastery outcomes approach implemented in this study naturally leverages the positive impacts of the testing effect by providing more frequent self-assessment scenarios. Furthermore, the mastery grading approach naturally facilitates the incorporation of metacognitive strategies.

Student feedback on mastery assessments is directly linked to interactive courseware content with built-in tools for helping students monitor the learning process by identifying gaps in their understanding and addressing them through targeted practice. The assessment, reflection, and practice cycle is designed to provide a pattern of engagement that promotes a growth mindset. Mindset theory is based on research indicating students who believe that intelligence is malleable often experience more positive learning outcomes. Students with a growth mindset tend to view initial failure as an opportunity for improvement rather than a predictor of future negative outcomes [23].

1.2. Research Questions and the Current Study

1. How do student performance outcomes differ between courses incorporating a mastery grading/test-retake system and a course using infrequent high-stakes exams?
2. How do courses incorporating a mastery grading/test-retake system affect equity gaps compared to a course using infrequent high-stakes exams?
3. What is the general qualitative student affective response to courses using mastery grading/test-retake system?

The present work employs a second-chance testing strategy through weekly unit mastery assessments (denoted as Mastery). However, we have chosen a mastery-focused model that dispenses with the token economy and allows every student a fixed number of scheduled retakes for a given mastery assessment [7,9]. This mastery grading approach directly fosters a growth mindset in students and demonstrates a commitment from instructors that students possess the capacity to improve through persistent effort. The cues hypothesis states that instructors who espouse a fixed mindset create threatening situational cues that can demotivate students, especially traditionally under-represented students [25]. By adopting a mastery outcomes assessment structure, instructors will show a commitment to a student growth mindset that should lead to improved cognitive and affective outcomes.

Coupling the mastery grading model with metacognitive coaching and interactive courseware designed to promote asynchronous active learning (Mastery+OLI) is a crucial complement to the mastery learning model. Suppose that students are not given guidance on evaluating and reflecting upon their learning. In that case, it is unlikely that having multiple attempts on the various content assessments will lead to meaningful learning gains [20]. The General Chemistry curriculum available through the Open

Learning Initiative (OLI) at Carnegie Mellon University was selected for the present work. OLI General Chemistry provides a rich, interactive learning environment built upon the OpenStax Chemistry textbook with embedded problems that provide extensive hints and feedback [26,27]. The structure of OLI General Chemistry is based on the literature findings, which suggest students learn more by doing interactive problems rather than reading text or watching videos [28].

2. Materials and Methods

2.1. Implementation Details

The study involved the second course in a three-quarter sequence for general chemistry, which is the required introductory chemistry sequence for all students in the UCR College of Natural and Agricultural Sciences (CNAS). The topics are separated into six units: gases, thermochemistry, liquids and solids, solutions chemistry, thermodynamics, and kinetics. A quasi-experimental study compared two versions of mastery outcomes grading system to a teaching-as-usual course that used traditional high-stakes exams. Performance on a common final assessment administered to the three sections was used to compare student learning outcomes.

The study took place during the winter 10-week quarter at a large public research university in Southern California, federally designated as a Hispanic Serving Institution. All three sections were taught at the same time by three different instructors. Student demographic data were separated based on ethnicity, financial aid status, and first-generation status (see Table 1). Ethnicity and first-generation status were determined by self-reporting in admission files. First-generation status was defined as neither parent completing a 4-year degree (i.e., the highest level of education being "some college", "high school", or "some high school"). Student data were separated into two groups based on ethnicity. Students who self-reported as white or Asian were classified as not belonging to an under-represented minority group (not URM). Students belonging to all other ethnic backgrounds were classified as URMs.

Table 1. Descriptive statistics for the performance on the common final exam across the three groups. The control (N = 239), mastery (N = 242), and mastery with OLI (N = 244) sections had similar total enrollment. YES corresponds to students belonging to the corresponding demographic population listed in the left-hand column.

	Control		**Mastery**		**Mastery+OLI**	
	YES	**NO**	**YES**	**NO**	**YES**	**NO**
URM:						
Mean Final Exam %	60.2	67.8	59.3	70.7	67.3	75.0
Stnd. Dev.	21.5	19.2	20.1	17.4	18.5	14.3
% of Population	38.5%	61.5%	28.5%	71.5%	39.4%	60.6%
First-Generation:						
Mean Final Exam %	54.6	69.5	58.9	71.0	67.0	74.6
Stnd. Dev.	20.7	18.6	17.8	18.3	19.4	14.1
% of Population	31.0%	69.0%	28.9%	71.1%	34.4%	65.6%
Financial Aid Status:						
Mean Final Exam %	64.2	67.5	66.4	72.9	71.6	73.4
Stnd. Dev.	20.7	19.2	19.2	16.7	16.8	15.3
% of Population	79.9%	20.1%	83.9%	16.1%	80.9%	19.1%

The control group used traditional publisher textbook resources, while the Mastery group used the Atoms First General Chemistry text available through OpenStax [26]. To ensure equitable access, students enrolled in the Mastery+OLI group were given free access to the Open Learning Initiative general chemistry resources. Care was taken to ensure content coverage was the same throughout the three sections. The instructor for the control group wrote a set of 20 common questions to be administered on the final exam for the Control, Mastery, and Mastery+OLI courses. The common questions covered various topics throughout the course and were given to the instructors running the mastery grading sections after the last day of instruction to minimize bias. The common test questions were evaluated for content validity by the Mastery and Mastery+OLI instructors, and though these test items were not evaluated for internal reliability prior to being administered on the final exam in the three courses, post hoc item analyses suggest these items were reliable measures of content knowledge (see questions, item means, and item discrimination indexes in Supplementary Materials (Section S4)). The two mastery grading sections were coordinated in terms of structure, content, and instructional approach. The instructor running the Mastery grading section is a Distinguished Professor of Teaching, and the instructor running the Mastery+OLI section is an Assistant Professor of Teaching.

2.2. Design of the Mastery Grading System

The instructors for the Mastery and Mastery+OLI sections coordinated designing the mastery test retake system. The mastery grading system was based on a second-chance testing system in which the course content was divided into six units. Mastery unit assessments aligned with each unit's learning objectives were developed and administered in bi-weekly proctored assessment sessions. New versions of the unit mastery assessment were released two times a week, and students had the opportunity to demonstrate mastery of both new and previous content according to a predetermined testing schedule outlined in the syllabus (see Supporting Information). The testing schedule was designed to provide three attempts for each unit, with the highest score counting toward the final course grade. If students were satisfied with their score on the first mastery exam, retakes were optional. All mastery assessments were administered during the discussion sections and proctored by graduate teaching assistants (discussion sections are mandatory one-hour weekly meetings in which the class size is approximately 30–40 students).

The unit mastery exams were composed of ten multiple-choice and numerical response questions administered through the learning management system. The questions were selected from extensive question banks separated by course learning objectives. A mastery grading scheme was employed when scoring the mastery assessment. According to the mastery grading scheme, a score of 9/10 was assigned full points, scores at or below 5/10 were assigned zero, and scores in between were assigned the corresponding percentage. The mastery grading scheme was applied to the highest score of their multiple attempts, and each unit mastery exam accounted for 10% of the final course grade. The unit mastery assessments accounted for 60% of the final course grade. The threshold for mastery was defined as a score of 60% on the unit mastery assessments (historically, 60% has been defined as a 'C-' grade in the fixed grading scale used in the department). See Table 2 for a complete breakdown of the grading for each section and the syllabus language, along with a tentative course schedule in Section S1 in the Supporting Information.

Table 2. Breakdown of the grading scheme for the three groups. [a] Homework for the control was administered through a standard publisher online homework system, the mastery course assigned instructor-created assignments in Canvas, and homework was assigned within the OLI platform for the Mastery+OLI section. [b] For all three courses, participation and discussion points were assigned based on completing in-class poll questions during the main lecture and attendance at weekly recitation/discussion sections.

	Control	Mastery	Mastery+OLI
Midterm 1	12%	10%	10%
Midterm 2	12%	–	–
Final Exam	24%	10%	10%
Homework [a]	24%	10%	15%
Discussion/Participation [b]	28%	10%	5%
Mastery Exams	–	60%	60%

Traditional comprehensive midterm and final exams were administered at the midpoint and end of the quarter, respectively—the midterm and final exams each account for 10% of the final course grade. However, the midterm exam covered content from the first two units, and if the score on the midterm was higher, it was used to replace the mastery assessment score for either or both of the first two units. Similarly, the final exam covered content from all six units, and the final exam score could replace the mastery exam scores and midterm scores. The remaining 20% of the final course grade was allocated to homework and participation. The mastery-only section assigned homework using the pre-recorded lecture videos with embedded questions through Canvas Studio. Engagement with the online courseware was incentivized for the mastery+OLI section, with 15% of the final course grade allocated to homework assigned within the OLI platform. We considered this a pure mastery (second-chance) testing system as opposed to a specifications model. We were not building groups of tasks or student products. Instead, the mastery grading model assessed blocks of content through multiple attempts on individual assessments. Furthermore, there was no token economy; all students were provided multiple attempts on every mastery assessment [7].

2.3. Design of the Interactive Courseware Platform

A combination of Canvas practice exams, recorded lecture videos, and the OLI interactive courseware formed the foundation for asynchronous content delivery in the Mastery+OLI group. The interactive online chemistry courseware was developed through the Open Learning Initiative (OLI) at Carnegie Mellon University. OLI General Chemistry is a comprehensive, data-driven, and evidence-based general chemistry curriculum. It should be noted that the OLI General Chemistry textbook is based on the atoms-first OpenStax Chemistry text, which covers the same curriculum as the textbook used by the control sections [26]. The OLI platform is designed to promote asynchronous active learning and provides rich student user data [29]. The chemistry content consists of units equivalent to a textbook chapter. Each unit is separated into modules, which include 5 to 10 content pages. Each content page includes didactic instruction in the form of text, images, and videos. In addition, "Learn by Doing" and "Did I Get This" activities are interspersed throughout each content page to promote active engagement with the content. The "Learn by Doing" activities typically break the problem-solving process into steps and provide extensive hints and feedback to guide the student to the correct answer. "Did I Get This" activities typically provide less extensive feedback and scaffolding. Each module concludes with a checkpoint activity, which can serve as homework, where feedback is provided only at the end of the quiz. The checkpoint activities were assigned as homework, accounting for 15% of the final course grade. Interaction with the OLI courseware was incentivized through participation points, accounting for 2.5% of the final course grade. A detailed analysis of student interaction within the OLI courseware and mastery grading system was provided in previous work [22].

2.4. Statistical Analyses

The final exam data was analyzed as part of a post hoc observational study approved by the UCR Institutional Review Board (IRB) under protocol number 30202. The approved protocol included obtaining student demographic information from the UCR Office of Institutional Research, and because this was carried out as a post hoc observational study, students were not required to complete an informed consent. Analysis of variance (ANOVA) was used to compare final exam scores between the three study groups; these were carried out for the entire class populations and for specific demographic groups (see Figures 1 and 2). The assumptions for ANOVA were evaluated (e.g., the dependent variable was normally distributed, and homogeneity of variance was observed across the three study groups), and post hoc pair-wise tests were used to determine where significant differences in final exam scores were observed for cases when the omnibus ANOVA was found to be statistically significant. Effect sizes were calculated using the partial eta-squared statistic from the omnibus ANOVA. The omnibus ANOVA and post hoc pair-wise tests were carried out using the IBM SPSS software program version 28.0.0.0 [30].

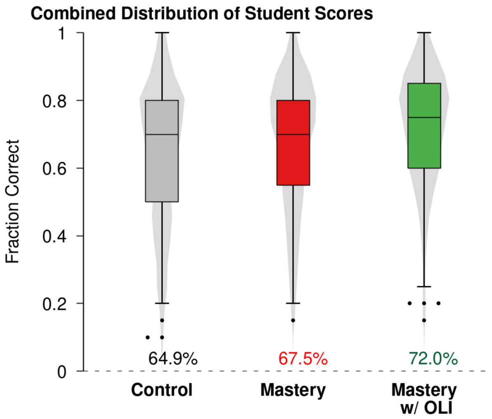


Figure 1. The distribution of student scores on the set of 20 common final exam questions is represented by the fraction of questions answered correctly. A box plot within each violin shows the median error (black line), middle 50th percentile (colored box), the range of errors (black lines), and outliers—represented by dots. The corresponding average scores, expressed as a percent correct, are provided below each distribution. Omnibus ANOVA comparing mean score on common final exam items ($F = 8.847$; $p < 0.001$); post hoc pair-wise comparisons with Bonferroni correction (Mastery+OLI vs. Control mean difference = 7.08, $p < 0.001$; Mastery+OLI vs. Mastery mean difference = 4.48, $p = 0.026$; see Supplementary Tables S2 and S3).

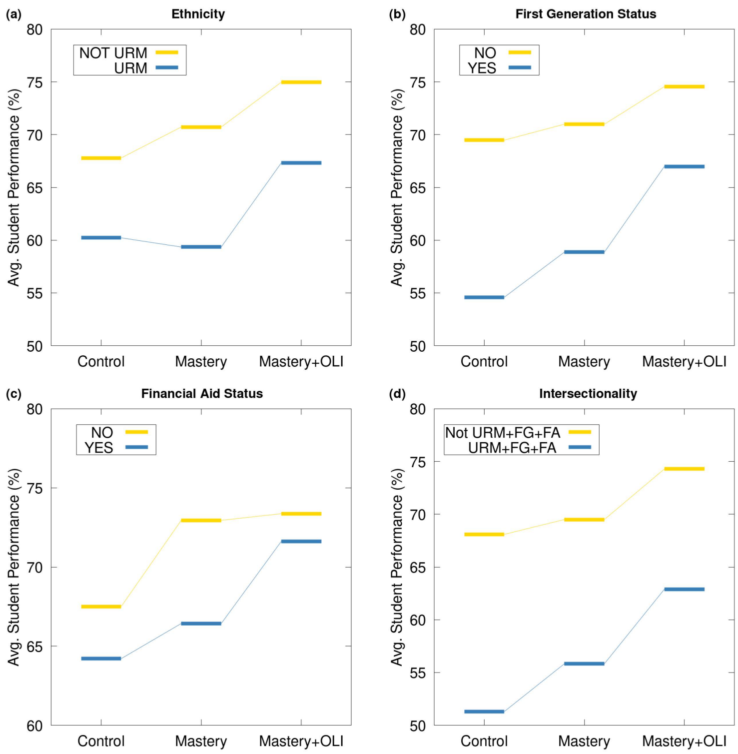


Figure 2. Average performance on common final exam questions disaggregated by (**a**) ethnicity: see Supplementary Tables S4 and S5, (**b**) first-generation status: see Supplementary Tables S6 and S7, and (**c**) financial aid status: see Supplementary Tables S8 and S9. Panel (**d**) compares the average student performance of all students to that of students belonging to all three groups: first-generation (FG), URM, and receiving financial aid (FA); see Supplementary Tables S10 and S11. (URM = underrepresented minority students; non-URM = white/Asian students).

3. Results

We begin by comparing student performance on the common final assessment questions. Figure 1 plots the distribution of student scores for the control (gray), Mastery (red), and Mastery+OLI (green) courses. The Mastery and Mastery+OLI courses led to higher scores on the common final exam questions than the control. Specifically, the mean performance expressed as the percent of correct responses on the common final exam questions for the Mastery+OLI group was 7.1 percentage points greater than the control ($p < 0.001$). Similarly, the mean performance for the Mastery group was 2.6 percentage points higher relative to the control; however, this improvement was not statistically significant at the 0.05 significance level. Figure 1 shows a narrower distribution in scores for the Mastery+OLI section, and the standard deviation for the Mastery+OLI group (16.4) is less than both the Mastery (18.9) and Control (20.4) groups.

Despite the fact the Mastery+OLI course had the highest percentage of URM, first-generation, and students receiving financial aid (see Table 1), the mean common exam scores for the Mastery+OLI course were significantly higher when the entire class population was analyzed (Figure 1). Disaggregating student performance data by ethnicity, first-generation status, and financial aid status reveals striking trends. In particular, Figure 2a shows that mastery grading alone did not significantly impact the average student performance for URM students (blue) relative to the control. However, when the mastery grading model was coupled with the OLI interactive courseware tools, the average performance for URM students improved by 7.1 percentage points relative to the control ($p < 0.001$), with an effect size of 0.181 (see Table 3). Even more pronounced improvements are observed for students with first-generation status, as seen in Figure 2b). Relative to the control, first-generation students enrolled in the Mastery+OLI course demonstrated a mean difference of 12.5 percentage points, with a moderate effect size of 0.272 ($p < 0.001$) compared to only a 5.1 percentage point improvement for students who did not identify as first-generation (see Table 3). These findings suggest mastery grading provides improved student learning for students with sufficient scaffolding for addressing gaps in content knowledge. The rich interactive tools provided through OLI in the Mastery+OLI course are likely crucial for assisting URM and first-generation college students in addressing gaps in their content knowledge. These findings are consistent with our recent work examining the link between engagement with the OLI courseware and performance on mastery assessments [22].

Table 3. Omnibus analysis of variance (ANOVA) comparing mean score on common final exam items. Results are reported for URM students (Table S4), first-generation college students (Tables S6 and S7), students who are receiving financial aid (Tables S8 and S9), and students belonging to all three categories (Tables S10 and S11).

	Omnibus ANOVA			Mastery+OLI vs. Control		Mastery+OLI vs. Mastery	
	F	***p***	***Cohen's f***	**Mean Diff.**	***p***	**Mean Diff.**	***p***
URM	4.19	0.016	0.181	7.10	0.048	7.97	0.038
First-Generation	8.36	<0.001	0.272	12.40	<0.001	8.13	0.031
Financial Aid	7.80	<0.001	0.163	7.40	<0.001	5.19	0.019
Intersectionality	3.97	0.021	0.248	11.60	0.018	7.07	0.343

Financial aid status was used as a proxy for identifying students more likely to have experienced financial hardship. Similar to the results found for ethnicity and first-generation status, students who received financial aid disproportionately benefited from the Mastery+OLI implementation relative to the control group. In particular, the Mastery+OLI group showed a 7.4 percentage point improvement in the mean relative to the control. Interestingly, the mean performance of the students who had not received financial aid in the Mastery and Mastery+OLI courses was the same. These results further support the hypothesis that students from more affluent backgrounds are more likely to have family members who have attended college and are, therefore, more likely to have developed habits and practices conducive to mastering chemistry content.

Finally, we considered intersectionality in the data by comparing the average performance of all students with those who simultaneously identify as first-generation college students, members of a URM group, and receive financial aid. Figure 2d and Table 3 show that mastery grading alone yielded a mean difference of 4.5 relative to the control ($p = 0.956$), and Mastery+OLI yielded a mean difference of 11.6 percentage points relative to the control ($p < 0.018$). The Mastery+OLI design showed a moderate effect, with a Cohen's f value of 0.248 (see Table 3) [31]. It should be noted that financial aid status was highly correlated with URM and first-generation status. Specifically, there were only two students who were both URM and first-generation status and not receiving financial aid.

Student Feedback Results

Recent work suggests a poor correlation between student evaluations and student learning, and the efficacy of student evaluations has been questioned [32]. However, this lack of correlation has been observed for numerical ranking systems, and evidence suggests student evaluations remain a crucial tool for providing instructors with feedback [33]. In particular, the comments section can provide valuable insights from the student's perspective, and such insights are particularly useful when evaluating novel instructional tools and course design. Nevertheless, analyzing hundreds of student responses for sentiment and relevance to a specific intervention while minimizing the introduction of bias is a challenging task. Recently, Hoar et al. suggested using natural language processing tools to facilitate this process [34]. Here, we apply natural language processing tools to student course evaluation data to analyze student feedback on the mastery grading system.

Student feedback was collected as anonymous course evaluation data for both mastery grading sections. The comments section in the course evaluation data was separated by sentence, providing 397 responses for the mastery with OLI section and 95 responses for the mastery grading section that did not use OLI. Each sentence was processed using Google's Natural Language API to identify keywords with a corresponding salience ranking. We selected keywords related to mastery grading and collected the associated sentiment scores (see the Supporting Information for details). This analysis resulted in 103 student comments related to the mastery grading model across both sections. Finally, each relevant student comment was subject to sentiment analysis using Google's Natural Language Processing tools. Google Cloud Sentiment Analysis (GCSA) provides a sentiment score between -1 and $+1$, with larger scores corresponding to more positive sentiment. Sentiment analysis was carried out using two additional algorithms and the results were similar to those reported in Figure 3 (see Supporting Information for details). Readers interested in the details of implementation and the comparative performance of alternative sentiment analysis algorithms are directed to the following review [35].

The histogram in Figure 3 illustrates the sentiment analysis results. For ease of interpretation, scores below -0.25 are classified as negative (red), between -0.25 and $+0.25$ as neutral (yellow), and above $+0.25$ as positive (green). This analysis was replicated with similar results using the VADER and TextBlob algorithms (see the Supporting Information for details) [36,37]. The mastery grading model resulted in generally positive or neutral student feedback. In particular, numerous students noted decreased stress and anxiety from the second-chance testing. On the other hand, numerous students expressed concern about sacrificing small-group discussion time in favor of repeated mastery exam attempts. Additionally, several students remarked on the lack of flexibility regarding testing times.

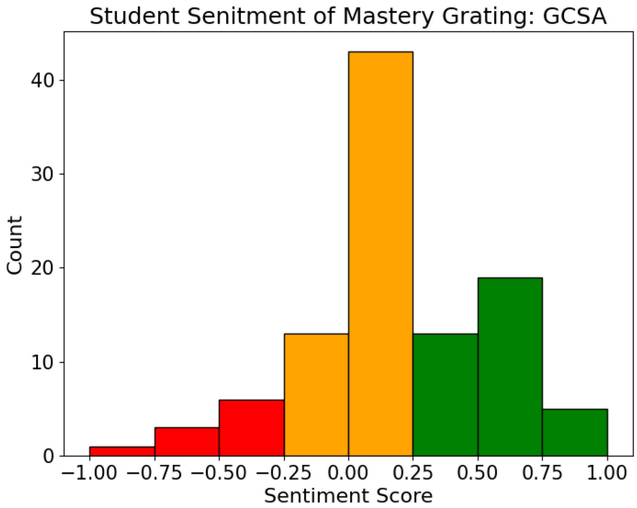


Figure 3. Student sentiment toward mastery grading based on 103 student comments from sections employing mastery grading. Each comment is classified based on the sentiment score, with scores below −0.25 classified as negative (red), between −0.25 and +0.25 as neutral (yellow), and above +0.25 as positive (green).

4. Discussion

The preliminary implementation of mastery grading in large-enrollment chemistry courses appears to have improved overall student learning outcomes and reduced equity gaps. This study contributes to the growing body of literature on alternative grading systems by demonstrating the efficacy of a mastery-focused approach, particularly when supplemented with interactive courseware like the General Chemistry curriculum provided through the Open Learning Initiative (OLI). The overall performance improvements we observe with the Mastery+OLI course are comparable to those seen in a pharmacokinetics (PK) and pharmacodynamics (PD) course that used a weekly quizzing model. Specifically, Henning et al. reported a 7.93 percentage point improvement in the average scores on the PK/PD component of the final exam [21].

Although several studies have explored alternative grading systems in chemistry, these studies do not directly compare performance across different demographics [6,11,12,14]. However, one study involving high school chemistry students found that a mastery grading model improves learning outcomes for students having difficulty with the content [6]. These previous results reported for high school chemistry are broadly corroborated by our finding that URM, first-generation, and students receiving financial aid showed the greatest improvement when using the Mastery+OLI model (Figure 2). Another study implementing a mastery grading model in an undergraduate engineering course found that women and URM students benefited from the alternative grading model to the same extent as the general population [3]. The present work did not consider gender; however, Figure 2a does show similar improvement for both URM and non-URM students when using the Mastery+OLI model.

Examining the disaggregated statistics for learning outcomes highlights the importance of incorporating metacognitive tools within the mastery grading model. Mastery grading alone did not improve student learning relative to the control for URM students (Figure 2a). These findings are not necessarily surprising when considered in the context of the recent literature. The role of retrieval practice in consolidating learning is well established [18,24]. However, consolidating learning through repeated testing may be of little value for students who are struggling to grasp the challenging and complex concepts in general chemistry [20]. Casselman et al. have shown that providing responsive online content designed to promote the development of metacognitive skills improved ACS exam performance by 4% relative to the control [19]. The OLI interactive courseware provides an accessible platform where students can regularly assess their abilities, receive detailed feedback regarding progress toward learning goals, and create a future study plan. Developing metacognitive skills is expected to be particularly impactful for students with minimal previous training of this nature (e.g., URM and first-generation students in Figure 2).

Chemsitry-specific growth mindset interventions in first-year general chemistry have been shown to improve the student learning experience and even eliminate the ethnicity achievement gap [38]. The mastery grading system reinforces the belief that chemistry content knowledge and problem-solving skills can be developed and improved over time. A closer examination of the sentiment analysis data presented in Figure 3 shows that the majority of the positive student feedback references a reduction in stress surrounding testing and a shift toward viewing mistakes as opportunities for growth, resulting in improved performance on subsequent exams. Though this was not directly explored in the study, we speculate that the absence of a curve in the mastery grading model promoted peer-to-peer engagement because the course grade was no longer tied to a student's performance relative to their peers.

Limitations and Future Work

Although we do not have incoming knowledge data, all students were placed into the course on the same track based on math placement. The distribution of students is likely equal across all three sections with respect to those math placement scores. However, as with any observational study, we ultimately could not account for the various confounding variables that might have negatively impacted student performance (prior chemistry knowledge, differences in co-curricular demands among the class populations, etc.) Additionally, though instructor bias could not be accounted for, it is noted the Mastery course was taught by a Distinguished professor of teaching and the control instructor provided the common questions. These factors suggest there was no bias favoring higher student exam performance in the Mastery+OLI group. Further studies are currently being designed to include initial knowledge assessment and the deployment of the Mastery+OLI model at scale across multiple institutions.

The common assessment items were generally emphasizing more traditional skills and knowledge. There is an emerging emphasis in the chemistry education community to move beyond a procedural and skill-based focus and promote learning objectives associated with conceptual understanding and more meaningful learning [39,40]. Therefore, future work will investigate how a mastery grading approach can improve this type of higher-order learning. However, multiple-choice questions administered through an online testing system better accommodate the volume of testing inherent to the Mastery+OLI model, and building a test retake system with more open-ended conceptual assessment items will be a challenge that needs to be overcome. Finally, the end of the term presents a logistical limitation, wherein students must complete multiple mastery attempts in rapid succession. This did lead to some student anxiety, and future implementations will focus on strategies for providing a more flexible testing schedule.

We have demonstrated that this test retake system can be implemented in large enrollment intro courses at an R1 institution with TA support and built-in recitations/discussions. The absence of either TA support or recitations/discussion sections would make it difficult to replicate this approach. Furthermore, sacrificing small-group discussion time in favor of testing presents a limitation, evidenced by student comments that expressed concern over the loss of small-group interaction time in the discussion/recitation sections. Future work will include exploring alternative second-chance testing models using a testing center, collecting more data on the student experience, and carrying out a more detailed study on the affective outcomes and/or outcomes related to student mindset.

5. Conclusions

Despite these limitations, the results described herein provide compelling evidence that a mastery outcomes/test retake system can significantly improve equity gaps in gateway STEM courses. There has been an ongoing long-term effort to improve retention of underserved students in higher education STEM [19,38], yet these equity gaps persist. It is proposed here that a contributing factor to this problem is the fact that many recruitment and retention programs are implemented in parallel to introductory STEM courses that continue to employ infrequent high-stakes exams. It is argued here that traditional assessment structures will ultimately limit the impact of these co-curricular programs and likely negatively impact outcomes for underserved students. It is hoped this current study provides a template for a new way forward.

Supplementary Materials: The following Supporting Information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci14111224/s1, Section S1. Syllabus Language; Section S2. Pair-wise Comparison Statistics; Section S3. Student Feedback; Section S4. Common Test Questions.

Author Contributions: Conceptualization, J.D.H. and J.F.E.; methodology, J.D.H. and J.F.E.; validation, J.D.H. and J.F.E.; formal analysis, J.D.H. and J.F.E.; writing—original draft preparation, J.D.H.; writing—review and editing, J.D.H. and J.F.E.; funding acquisition, J.D.H. and J.F.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the NSF grant number DUE-2421279. This project was also funded in part by the California Education Learning Lab, a state-funded grantmaking organization with a mission to improve learning outcomes and close equity gaps for students attending public higher institutions in California.

Institutional Review Board Statement: The study was approved by the Institutional Review Board of the University of California, Riverside (protocol code 30202 approved 6 May 2024).

Informed Consent Statement: Because all data collection and data analysis were carried out as part of a post hoc observational study subsequent to the course delivery and final grade submissions, no informed consent was administered to students.

Data Availability Statement: The raw datasets used to carry out the statistical analyses in this article are not available due to restrictions in the IRB human subjects research protocol.

Acknowledgments: We wish to thank our colleagues in the Chemistry Department at the University of California, Riverside for their continued support. In particular, we thank Christopher Bardeen and Linlin Zhao for numerous helpful discussions. Access to the interactive courseware was provided free of charge through the Open Learning Initiative at Carnegie Mellon University. We wish to thank the developers and support staff at OLI. In particular, we thank Mark Blaser, Sandra Rasor, David Yaron, Norman Bier, and Ariel Anbar.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

OLI	Open Learning Initiative
GCSA	Google Cloud Sentiment Analysis
URM	Under Represented Minority
FG	First-Generation
DFW	Drop, Fail, and Withdrawal

References

1. Bloom, B.S. Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Eval. Comment* **1968**, *1*, n2.
2. Zimmerman, B.J.; Dibenedetto, M.K. Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychol. Sch.* **2008**, *45*, 206–216. [CrossRef]
3. Morphew, J.W.; Silva, M.; Herman, G.; West, M. Frequent mastery testing with second-chance exams leads to enhanced student learning in undergraduate engineering. *Appl. Cogn. Psychol.* **2020**, *34*, 168–181. [CrossRef]
4. Kulik, C.L.C.; Kulik, J.A.; Bangert-Drowns, R.L. Effectiveness of mastery learning programs: A meta-analysis. *Rev. Educ. Res.* **1990**, *60*, 265–299. [CrossRef]

5. Toledo, S.; Dubas, J.M. A learner-centered grading method focused on reaching proficiency with course learning outcomes. *J. Chem. Educ.* **2017**, *94*, 1043–1050. [CrossRef]
6. Damavandi, M.E.; Kashani, Z.S. Effect of mastery learning method on performance, attitude of the weak students in chemistry. *Procedia-Soc. Behav. Sci.* **2010**, *5*, 1574–1579. [CrossRef]
7. Howitz, W.J.; McKnelly, K.J.; Link, R.D. Developing and implementing a specifications grading system in an organic chemistry laboratory course. *J. Chem. Educ.* **2020**, *98*, 385–394. [CrossRef]
8. McKnelly, K.J.; Howitz, W.J.; Thane, T.A.; Link, R.D. Specifications grading at scale: Improved letter grades and grading-related interactions in a course with over 1000 students. *J. Chem. Educ.* **2023**, *100*, 3179–3193. [CrossRef]
9. Nilson, L.B.; Stanny, C.J. *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time*; Routledge: London, UK, 2015.
10. Saucier, D.; Schiffer, A.; Renken, N. Five Reasons to Stop Giving Exams in Class. *Fac. Focus*. 2022. Available online: https://www.facultyfocus.com/articles/educational-assessment/five-reasons-to-stop-giving-exams-in-class/ (accessed on 1 January 2023).
11. Hollinsed, W.C. Applying innovations in teaching to general chemistry. In *Increasing Retention of Under-Represented Students in STEM Through Affective and Cognitive Interventions*; ACS Publications: Washington, DC, USA, 2018; pp. 145–152.
12. Ring, J. ConfChem conference on select 2016 BCCE presentations: Specifications grading in the flipped organic classroom. *J. Chem. Educ.* **2017**, *94*, 2005–2006. [CrossRef]
13. Houseknecht, J.B.; Bates, L.K. Transition to remote instruction using hybrid just-in-time teaching, collaborative learning, and specifications grading for organic chemistry 2. *J. Chem. Educ.* **2020**, *97*, 3230–3234. [CrossRef]
14. Ahlberg, L. Organic chemistry core competencies: Helping students engage using specifications. In *Engaging Students in Organic Chemistry*; ACS Publications: Washington, DC, USA, 2021; pp. 25–36.
15. Hunter, R.A.; Pompano, R.R.; Tuchler, M.F. Alternative assessment of active learning. In *Active Learning in the Analytical Chemistry Curriculum*; ACS Publications: Washington, DC, USA, 2022; pp. 269–295.
16. Kelz, J.I.; Uribe, J.L.; Rasekh, M.; Link, R.D.; McKnelly, K.J.; Martin, R.W. Implementation of specifications grading in an upper-division chemical biology course. *Biophys. J.* **2023**, *122*, 298a. [CrossRef]
17. Noell, S.L.; Rios Buza, M.; Roth, E.B.; Young, J.L.; Drummond, M.J. A bridge to specifications grading in second semester general chemistry. *J. Chem. Educ.* **2023**, *100*, 2159–2165. [CrossRef]
18. Karpicke, J.D.; Blunt, J.R. Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* **2011**, *331*, 772–775. [CrossRef] [PubMed]
19. Casselman, B.L.; Atwood, C.H. Improving general chemistry course performance through online homework-based metacognitive training. *J. Chem. Educ.* **2017**, *94*, 1811–1821. [CrossRef]
20. Van Gog, T.; Sweller, J. Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educ. Psychol. Rev.* **2015**, *27*, 247–264. [CrossRef]
21. Hennig, S.; Staatz, C.E.; Bond, J.A.; Leung, D.; Singleton, J. Quizzing for success: Evaluation of the impact of feedback quizzes on the experiences and academic performance of undergraduate students in two clinical pharmacokinetics courses. *Curr. Pharm. Teach. Learn.* **2019**, *11*, 742–749. [CrossRef]
22. Asher, M.W.; Hartman, J.D.; Blaser, M.; Eichler, J.F.; Carvalho, P.F. Test, Review, Repeat: Mastery-Based Testing and its Benefits for Student Engagement and Performance in a General Chemistry Course. *OSFPreprints* **2024**. *preprint*. [CrossRef]
23. Kapasi, A.; Pei, J. Mindset theory and school psychology. *Can. J. Sch. Psychol.* **2022**, *37*, 57–74. [CrossRef]
24. Karpicke, J.D.; Roediger, H.L., III. The critical importance of retrieval for learning. *Science* **2008**, *319*, 966–968. [CrossRef]
25. Canning, E.A.; Muenks, K.; Green, D.J.; Murphy, M.C. STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Sci. Adv.* **2019**, *5*, eaau4734. [CrossRef]
26. Flowers, P.; Theopold, K.; Langley, R.; Robinson, W.R. Chemistry (OpenStax), 2015. Available online: https://openstax.org/details/books/chemistry-2e/ (accessed on 1 January 2023).
27. Bier, N.; Moore, S.; Van Velsen, M. Instrumenting courseware and leveraging data with the Open Learning Initiative (OLI). In Proceedings of the Companion Proceedings 9th International Learning Analytics & Knowledge Conference, Tempe, AZ, USA, 4–8 March 2019.
28. Koedinger, K.R.; Kim, J.; Jia, J.Z.; McLaughlin, E.A.; Bier, N.L. Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In Proceedings of the Second (2015) ACM Conference on Learning@ Scale, Vancouver, BC, Canada, 14–18 March 2015; pp. 111–120.
29. Lovett, M.; Meyer, O.; Thille, C. JIME-The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *J. Interact. Media Educ.* **2008**, *2008*, 13. [CrossRef]
30. IBM Corp. *IBM SPSS Statistics for Windows*; Version 28.0; IBM Corp.: Armonk, NY, USA, 2021.
31. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Routledge: London, UK, 2013.
32. Uttl, B.; White, C.A.; Gonzalez, D.W. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Stud. Educ. Eval.* **2017**, *54*, 22–42. [CrossRef]
33. Benton, S.L.; Ryalls, K.R. *Challenging Misconceptions About Student Ratings of Instruction*; IDEA Paper# 58; IDEA Center, Inc.: Buffalo, NY, USA, 2016.
34. Hoar, B.B.; Ramachandran, R.; Levis-Fitzgerald, M.; Sparck, E.M.; Wu, K.; Liu, C. Enhancing the Value of Large-Enrollment Course Evaluation Data Using Sentiment Analysis. *J. Chem. Educ.* **2023**, *100*, 4085–4091. [CrossRef]

35. Al-Otaibi, S.T.; Al-Rasheed, A.A. A review and comparative analysis of sentiment analysis techniques. *Informatica* **2022**, *46*. [CrossRef]
36. Hutto, C.; Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; Volume 8, pp. 216–225.
37. Loria, S. textblob Documentation. *Release 0.15* **2018**, *2*, 269.
38. Fink, A.; Cahill, M.J.; McDaniel, M.A.; Hoffman, A.; Frey, R.F. Improving general chemistry performance through a growth mindset intervention: Selective effects on underrepresented minorities. *Chem. Educ. Res. Pract.* **2018**, *19*, 783–806. [CrossRef]
39. Stowe, R.L.; Scharlott, L.J.; Ralph, V.R.; Becker, N.M.; Cooper, M.M. You are what you assess: The case for emphasizing chemistry on chemistry assessments. *J. Chem. Educ.* **2021**, *98*, 2490–2495. [CrossRef]
40. Holloway, L.R.; Miller, T.F.; da Camara, B.; Bogie, P.M.; Hickey, B.L.; Lopez, A.L.; Ahn, J.; Dao, E.; Naibert, N.; Barbera, J.; et al. Using Flipped Classroom Modules to Facilitate Higher Order Learning in Undergraduate Organic Chemistry. *J. Chem. Educ.* **2024**, *101*, 490–500. [CrossRef]

education sciences

MDPI

Article

Lifting the Gate: Evaluation of Supplemental Instruction Program in Chemistry

Duyen Pham [1] and Li Ye [2,*]

[1] Department of Science, College of the Desert, Palm Desert, CA 92260, USA; dpham@collegeofthedesert.edu
[2] Department of Chemistry and Biochemistry, California State University, Northridge, CA 91330, USA
* Correspondence: li.ye@csun.edu

Abstract: College-level introductory chemistry has a high impact for predicting students' early success and long-term academic outcomes. Students from traditionally underserved communities are disproportionately held back in this course. To improve student success, the Supplemental Instruction (SI) program at a public four-year Hispanic-serving institution was revamped from a voluntary option to a co-requisite for the introductory chemistry course. The study evaluates the effectiveness of the new format of SI program and explores associated factors contributing to the impacts of the program. Students with or without SI were compared through multiple quantitative metrics, including course GPA, final exam score, DFW rate, and pass rate. Data analysis showed that students who completed SI with credit obtained 0.9 higher average course GPA than their counterparts and performed significantly better on all the other metrics. More importantly, the equity gaps between underserved and better-served students were narrowed down. Furthermore, students who took introductory chemistry with SI still obtained 0.3 higher average course GPA in their subsequent General Chemistry I course than those without it. The findings suggested that incorporating the SI into introductory chemistry as co-requisite is necessary and effective to improve students' success and narrow down the equity gaps in gateway chemistry courses.

Keywords: postsecondary chemistry; supplemental instruction; peer-led cooperative learning; student success; equity gaps; evaluation research

1. Introduction

1.1. High Impacts of Gateway Chemistry Courses

Introductory chemistry college-level courses had traditionally an average high failure rate of 30%, according to transcript data of 20,987 students from thirty-one institutions in the United States [1]. These introductory courses are considered "gateway" courses because they usually serve as prerequisites and student performance in these core courses is strongly associated with Science, Technology, Engineering, and Mathematics (STEM) degree completion [2]. Student success in gateway chemistry courses is essential for retaining STEM majors in the STEM fields. Therefore, there has been a critical need to improve the student success rates in those fundamental credit-bearing college-level introductory chemistry courses. Various strategies have been reported to improve student learning outcomes in these gateway courses, including bridging and corequisite support courses targeting at-risk students [3,4], innovative pedagogical practices emphasizing conceptual understanding and collaborative learning [5,6], and adaptive online preparatory modules or intelligent tutoring systems [7,8].

Research data also showed that students from traditionally underserved communities are disproportionately held back in these gateway courses [1,2]. In this study, underserved students refer to those student groups who are associated with disadvantaged backgrounds and factors that have shown to be related to college completion in the STEM disciplines. Those factors include gender and ethnic communities, first generation status, and those facing economic and financial challenges [9]. More specifically, Underrepresented Minority (URM) students who self-identify as Black/African American, Hispanic/Latino, American Indian, Native Hawaiian, and Other Pacific Islander; First-generation (FG) college students, or those whose parents did not obtain a degree from a four-year institution; Pell-grant recipients, or students who have demonstrated financial need and received the United States federal Pell grants. However, a recent study showed that when underserved students completed their first college chemistry course with a minimum grade of C or better, they are more likely to continue the general chemistry series and complete their degrees in STEM than their better-served peers [10]. This "hyperpersistent zone" phenomenon suggests that equity issues in STEM fields could be addressed by reducing the achievement gaps between underserved and better-served students early on in their learning.

1.2. Supplemental Instruction Program

Teaching and learning in chemistry have been dominated by lecture-based classes where professors present the materials during class time and students are in class to take notes and follow along [11,12]. With this type of instruction, there is not much interaction between students and instructors or students and fellow students. Instead, active learning is a more effective pedagogical approach because it engages students actively in the learning and teaching process. Active learning has been shown to be linked with higher student performance in STEM courses as compared to passive pedagogy [13–15]. In chemistry education research, representative active learning pedagogies that have been reported to have positive impacts on students' learning outcomes are Process-Oriented Guided Inquiry Learning (POGIL) [16–19], Peer-led Team Learning (PLTL) [20–23], problem-based learning [24,25], and the flipped classroom model [26–28]. A common feature of these pedagogies is that they often incorporate cooperative learning into instruction to promote teamwork and students' problem-solving skills in chemistry [29]. Recent research studies also indicated that these evidence-based instructional practices are promising to reduce the achievement gaps between underserved and better-served students [30–32] and improve students' long-term success in more advanced chemistry courses [33,34].

In a similar fashion, the Supplemental Instruction (SI) program also set up structured time for students to work together in small groups to solve chemistry problems with more advanced peers who lead the group discussion. The idea of SI originated from Professor Deanna Martin in the School of Dentistry at the University of Missouri at Kansas City and it was piloted in a human anatomy course in the early 1970s at the university [35]. Given the positive results, the SI pilot program was expanded into several other colleges within the university and then recognized nationwide as an exemplary educational program by the U.S. Department of Education. The SI program is designed to provide spaces for students to work on structured problem-solving sessions, giving them time to master the challenging topics in the associated lecture courses that have low passing rates at the university. The advantages of SI are twofold: first, they are usually independent of the lecture classes, so instructors can work with the SI leaders to coordinate content materials used in SI while still having the full lecture class time to cover the broad range of course objectives at the depth required without having to give up content or speed up their teaching pace [36]; second, the SI sessions are led by student peers who have mastered the course content more recently and therefore understand the struggles of the students, sharing their learning experiences with them. Students feel more comfortable asking questions from the SI leaders and can build a sense of belonging by interacting with other fellow students in the same lecture course. The SI leaders could also benefit from teaching others to deepen their scientific knowledge and develop skills such as leadership, communication, and time management [37,38].

Having an additional academic support program where active and cooperative learning can be implemented outside of the classroom is beneficial to support both instructors' instruction and engage students actively with a deeper understanding of the target learning objectives with peer support. The SI program has been reported to improve students' course grades, lower failure rates, and has been correlated with higher retention and graduation rates in college-level courses [39–43]. The SI has been implemented in a variety of subjects including English writing [44], mathematics [45], biology [46], anatomy [47], and chemistry [48–50]. Traditionally, students are encouraged to attend these sessions, but they are not required. Prior research studies have shown that there is a positive and significant correlation between the number of SI sessions attended and course performance. For instance, Yue and colleagues studied the impacts of SI programs in twenty-two college-level courses; among those, eighteen were STEM courses. They divided the number of SI sessions attended into four groups: 0, 1–7, 8–15, and 16 or more sessions. Students who attended SI regularly (16 or more sessions) received a higher average course grade than others. This review study indicates that regular attendance at SI plays a critical role in improving student success in the lecture courses [42]. Additionally, recent studies showed evidence that incorporating psychological and emotional support into SI programs and asset-based interventions could reduce achievement gaps and improve affect for underserved students [51,52].

In our study, we aimed to revamp and scale up the existing SI program at the research institution and evaluate the effectiveness of the new SI program. Originally, participating in SI was not required like most cases reported in the literature and only two optional SI sessions were offered. As such, students enrolled in any lecture classes could enroll in these two SI sessions and only a very small portion of the students (3–6%) are enrolled in the SI program. Although all classes of the introductory chemistry course use a common textbook and final exam, the instructors might not have the same pace and pedagogical approaches for teaching the lectures. The nature of this makes it challenging for the SI leaders to plan and facilitate their sessions. Students with different experiences in the lectures might not follow along well with each other in the SI sessions. Another observation by the instructors is that students who chose to take the SI were often more motivated and high-achieving students in the lecture classes, while those who really needed it did not participate in the SI. To address these issues, the SI program was redesigned to be a co-requisite course of the introductory chemistry lectures; each lecture class has three corresponding SI sessions that are led by SI leaders who facilitate the problem-solving sessions. Students are required to enroll in one of the three SI sessions that usually take place one hour before or after their lecture course. The new design of the program addresses the aforementioned issues of only more motivated students enrolled in SI and the lack of coordination between the lectures and SI sessions.

This study investigated the effectiveness of the revamped SI program to student academic outcomes and equity gaps in the introductory chemistry course in the context of a Hispanic-serving institution. Academic outcomes including student exam performance, letter grade, and course GPA. Course GPA number was calculated from student letter grade to a scale of 0 to 4, where 4 is the highest GPA corresponding to an A and 0 is the lowest grade corresponding to an F. Equity gaps are calculated by the differences in course GPA between student groups. Additionally, the effectiveness of longitudinal impacts of the SI program in the subsequent chemistry (General Chemistry I) were also investigated to see whether the impact of the SI has lasted for the next level of chemistry courses.

The findings suggested that incorporating the SI into introductory chemistry as a co-requisite is necessary and students' success rates measured by multiple quantitative metrics improved significantly. The equity gaps between student groups who are from underserved and better-served groups, especially between URM and non-URM, FG and non-FG, were narrowed down. The positive effects of the SI also carried over to the next chemistry course, General Chemistry I. The historically high failure rates in the introductory chemistry course were reduced dramatically by this new format of the SI program thus far and the support from the university is long-lasting given the effectiveness of the program.

1.3. Guided Research Questions

The central research questions that guided our study are listed as follows:

(1) What are the impacts of the revamped SI program on the introductory chemistry students' academic success (measured by course GPA, common final exam, DFW rate, and pass rate)?
(2) To what extent does the SI program close the equity gaps between student groups (i.e., Non-URM vs. URM; Non-FG vs. FG; Non-Pell vs. Pell; Male vs. Female)?
(3) What other factors contributed to the SI program's effectiveness?
(4) What is the longitudinal impact of the new SI program on students' academic performance in the subsequent course (i.e., General Chemistry I)?

2. Methods

2.1. Research Design

The research study employed a quasi-experimental research design that the SI program was considered an intervention applied to the experimental group while the control group did not have SI. It is worth noting that the students who enrolled into SI sessions were not randomly assigned into the program, students chosen to stay might be more motivated or they believe that they need more support outside of the lecture classes. Although the participants were not randomly assigned to the two groups, certain confounding variables were matching, including participants who were enrolled in the same semester and used common textbooks, online homework assignments, assessments, and grading schemes. Student demographics and backgrounds between the two study groups were also compared to ensure they were comparable samples. Multiple student performance outcome variables such as course GPA, final exam scores, and pass rates were chosen to evaluate the effectiveness of the SI program. In addition to the matching-only design, factorial designs were also used to study the interactions between the independent variables (e.g., URM status, FG status, gender) with the outcome variable of course GPA [53].

2.2. Study Setting and Sample

Introductory chemistry lecture course: At the research institution, the introductory chemistry course entitled Principles of Chemistry is usually the first science college-level course taken by the majority of the students who major in STEM. It is also required by several health-related majors and serves as a prerequisite course for students who are required to take the General Chemistry I course if they do not pass or take the Chemistry Placement Test. The introductory chemistry lecture is taught twice a week, 75 min each. There were nine classes taught by four instructors in Spring 2022 with the revamped SI program. The average class size of the lecture was 57 students. All lecture classes used a common textbook, learning objectives, online homework assignments, and a common final exam written by all the instructors collaboratively who taught the course. A list of sample final exam questions can be found in the Supplementary Material.

SI program: SI sessions are listed as a co-requisite course for the introductory chemistry lecture course. Students obtained an independent grade (Credit/No Credit) from the lecture based on attendance in the SI class. Each SI class is taught twice a week, 50 min each in the form of peer-facilitated problem-solving sessions. The SI sessions are usually scheduled an hour before or after the corresponding lecture classes. As planned, students in the introductory chemistry lecture course needed to take one of the three SI classes corresponding to their lecture in Spring 2022. However, several sessions were canceled at the beginning of the semester due to staffing issues and some SI sessions were not correctly linked as a co-requisite course due to system errors. These problems provided a good comparison group for this research study because it ended up that about half of the students enrolled in SI and the other half did not. There were twelve SI sessions taught by eleven SI leaders. To evaluate the effectiveness of the implementation of the SI program, the study sample is all the students who enrolled in the introductory chemistry course in the Spring 2022 semester. We compared the students who enrolled in the introductory chemistry lecture with and without the SI. The average class size of the SI class was 20 students. All lectures and SI sessions were taught virtually only for the first three weeks due to the

COVID-19 pandemic and campus restrictions. All the rest of the semester, all chemistry lectures and SI sessions took place in person.

Coordination between introductory chemistry lecture and SI: Instructors were encouraged to emphasize the benefits of the SI classes in their lectures and had SI leaders introduce themselves during the first week of their lecture classes. The instructors added the SI leaders to their Learning Management (i.e., Canvas course sites) and shared course materials with leaders in advance. The SI leaders were required to attend all the lecture classes taught by associated introductory chemistry instructors to ensure the content and pace taught in the lectures and SI sessions were well-aligned and consistent in explanations of the chemistry concepts. SI leaders also served as facilitators for the in-class group activities during the lectures if needed.

SI leaders and training: SI leaders were selected by a series of criteria including their course GPA, academic level, work ethic, sense of responsibility, and tutoring or related teaching experience. Each eligible applicant was interviewed for an hour by the SI coordinator and the introductory chemistry course coordinator before being offered the SI leader position. SI leaders were provided two days of professional training before the semester started and they met as a group weekly with a facilitator to discuss challenges faced during the week of teaching and to brainstorm pedagogical strategies to handle those challenges.

2.3. Data Collection

Student demographics and course outcome data were collected from the Institutional Research office and student exam performance data were collected from the instructors who taught the introductory chemistry lectures. This study was reviewed and approved by the Institutional Review Board at the research institution (#IRB-FY20-411).

2.4. Data Analysis

The demographics of the SI and non-SI participants in the introductory chemistry course in Spring 2022 are listed in Table 1. Among 515 students who enrolled in the course, 233 students (45%) of the students enrolled in SI sessions, and the remaining 282 students (55%) were not due to staffing issues and system errors. Of the 233 students who participated in SI, 197 (85%) of the students passed SI with Credit, and the rest, 36 (15%), did not earn Credit in SI (See Table 1). Since the research institution is a four-year public university that is designated as Hispanic-serving institution, the majority of the student population is from under-served communities. The percentages of the female, URM, FG, and Pell grant recipients were comparable between students who participated in SI and those without SI. One-way Analysis of Variance (ANOVA), independent *t*-tests, and Chi-square tests were used to compare students' academic success by multiple metrics, including course GPA, the common final exam, DFW rate (students who received D and F grades and Withdraw from the course), and pass rate (C and above, the minimum grade students need to advance to General Chemistry I). Two-way between-groups ANOVAs were used to investigate the interaction effect between student groups and the impact of the SI program. One-way ANOVA was also used to examine the impact of the chemistry instructors and SI leaders. All the statistical tests were performed by the SPSS statistics software version 28.

Table 1. Comparison of demographics of SI and non-SI participants in introductory chemistry.

Semester Spring 2022	N of Students	Female	URM	FG	Pell
a. SI (Credit)	197	47%	79%	77%	61%
b. SI (No credit)	36	50%	80%	83%	53%
c. Without SI	282	51%	76%	77%	67%

a. 197 students who passed the SI session with Credit. b. 36 students who did not pass SI session and received No credit. c. 282 students who did not attend SI session.

3. Results

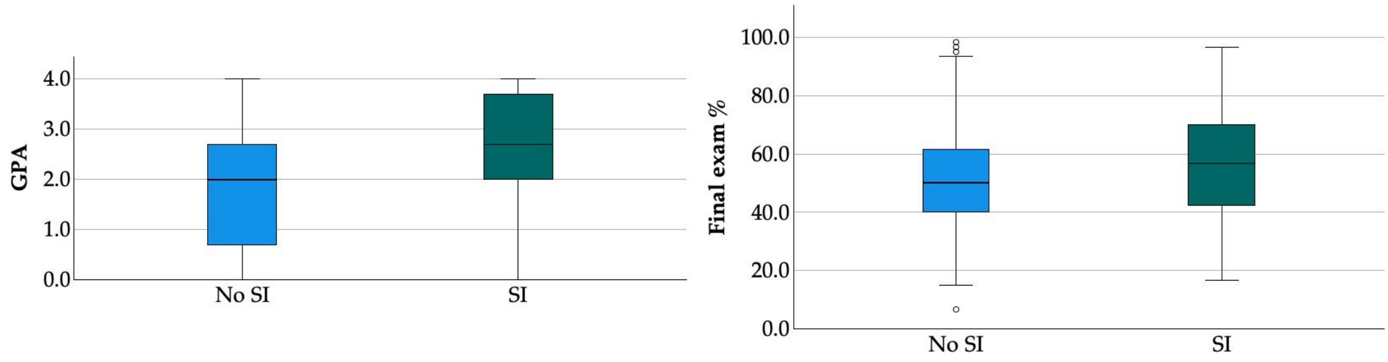
3.1. SI Impact on Students' Academic Outcomes in the Introductory Chemistry

To evaluate the impact of the SI program on introductory chemistry students' academic success in Spring 2022, multiple metrics were utilized and compared. First, the SI program significantly improved all student success metrics (See Table 2 and Figure 1). The average course GPA of students in introductory chemistry who passed the SI with credit was 0.89 (about a full letter grade) higher than those without SI. As expected, the students who participated in the SI but did not pass with the credit performed much worse than those without the SI (students with SI who received no credit course GPA = 1.07 versus students without SI course GPA = 1.71). One-way ANOVA showed a statistically significant difference in the introductory chemistry course GPA between the three groups of students, F (2, 514) = 46.37, $p < 0.001$. The effect size was 0.15, indicating that there is a large effect size [54]. Post hoc comparisons between each pair of the three groups using Tukey's test indicated that the average course GPA was significantly different between all pairs of the three groups (students with SI who received credit, students with SI who received no credit, and students without SI). Furthermore, similar trends have been found in other metrics, including the common final exam, DFW rate, and pass rate (C and above). Compared to the students without SI, passing the SI with credit improved students' final exams by 5.4%, decreased the student DFW rate in introductory chemistry by 24%, and increased the pass rate with C and above by 25%. The 1% difference was due to the students who obtained a C- grade still not meeting the minimum criteria to advance to the next level of chemistry courses. Independent *t*-test results showed that the difference in final exam scores between the students passing SI with credit and without SI was statistically significant, t (1152) = 2.87, $p < 0.001$. A Chi-square test of independence indicated a negative and significant association between the SI program and introductory chemistry DFW rate, χ^2 (1, 479) = −31.20, $p < 0.001$. The association between the SI program and introductory chemistry pass rate with C and above was positive and significant, χ^2 (1, 479) = 29.97, $p < 0.001$. Additionally, Pearson product coefficient correlations showed positive and significant relationships at the $p = 0.01$ level between the implementation of the SI program and introductory chemistry course GPA, the final exam, and pass rate (see Table 3). Figure 1 shows box plots of the course GPA and final exam of the students with SI credit and without SI. It is clear that the whole class distributions of the course GPA and final exam scores shift upwards due to SI. More specifically, Figure 1a indicated that the students who did not participate in SI had a lower median course GPA (around 2.0) and a wider range with more variability in their course GPAs. The 75% quartile reached close to 3.0 while the 25% quartile extended to nearly 1.0. Instead, the students participated in SI obtained a higher median GPA close to 3.0 with a more compact range from 2.0 to almost 4.0 course GPA; Figure 2b showed that students with SI had median final exam scores around 60% in the range between 17% and 97% while those without SI were about 50% in similar range of scores. These box plots suggested a positive association between SI participation and higher course GPAs and final exam scores.

Table 2. Comparison of introductory chemistry course GPA, final exam, DFW rate, and pass rate.

Semester Spring 2022	Course GPA	Final Exam *	DFW Rate	Pass Rate (C and Above)
a. SI (Credit)	2.60	57.1%	16%	76%
b. SI (No credit)	1.07	49.6%	67%	30%
c. Without SI	1.71	51.7%	40%	51%
Difference by SI (between a and c)	+0.89	+5.4%	−24%	+25%

* 192 students (97%) in group a took the final exam, 19 students (53%) in group b took the final exam, 214 students (76%) in group c took the final exam.



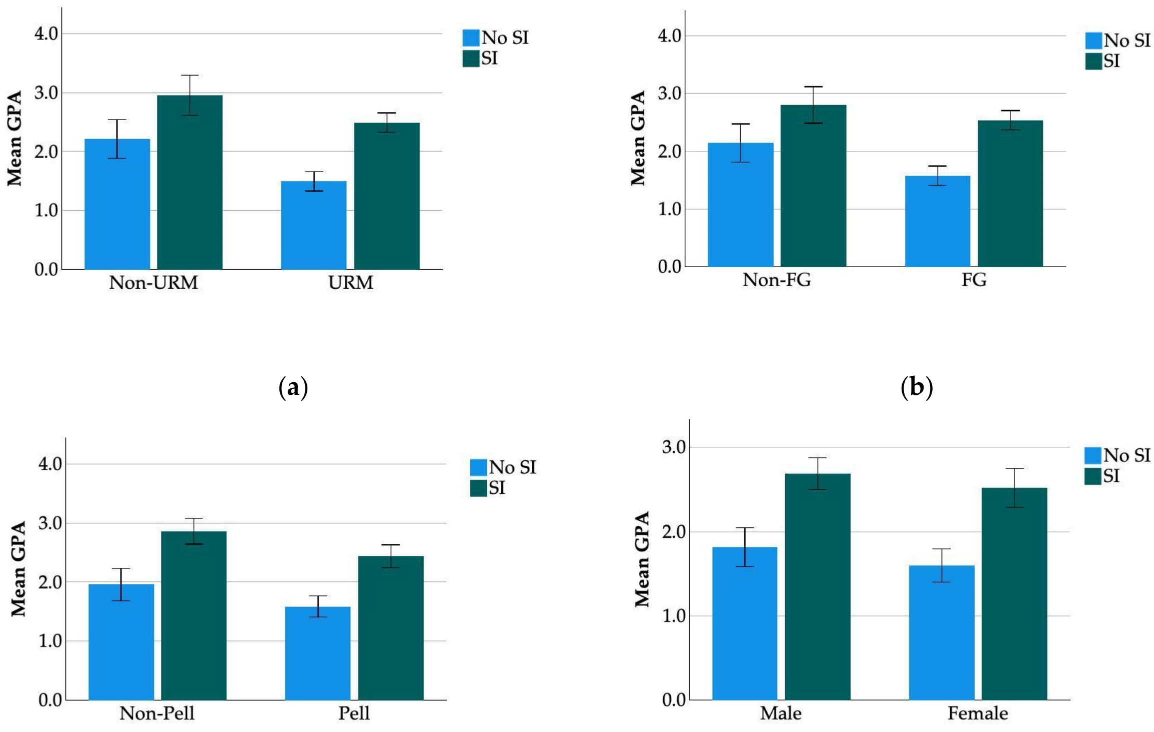
(**a**) (**b**)

Figure 1. Comparison of introductory chemistry course GPA and the final exam between students without and with SI credit: (**a**) box plots of course GPA; (**b**) box plots of final exam scores.

Table 3. Pearson product coefficient correlations between SI and student success metrics.

Correlations	Course GPA	Final Exam	DFW Rate	Pass Rate (C and Above)
SI	0.348 **	0.155 **	−0.255 **	0.250 **

** Correlation is significant at the 0.01 level.



(**a**) (**b**) (**c**) (**d**)

Figure 2. Comparison of introductory chemistry course GPA between student groups: (**a**) Non-URM vs. URM; (**b**) Non-FG vs. FG; (**c**) Non-Pell vs. Pell; (**d**) Male vs. Female. Error bars: 95% confidence interval.

3.2. SI Impact on Students' Equity Gaps in the Introductory Chemistry

There were thirty-six students who did not earn credit in the SI and therefore were removed from future analysis because the SI grades were solely based on class attendance.

No impact could be made if a student did not attend the SI sessions. As shown in Table 4 and Figure 2, the SI program benefited all student groups, including all the traditionally underserved student groups such as URM students, FG students, students from low-income backgrounds, and female students. With passing grades in SI, the introductory chemistry course GPA of URM students increased by 0.99 (~a full letter grade), FG students increased by 0.96, students from low-income backgrounds increased by 0.85, and female students increased by 0.92 as compared to those students without SI. Equity gaps are measured by the differences in the introductory chemistry course GPA between underserved student groups and their counterparts. By passing SI, the equity gaps between URM and non-URM, FG and non-FG, and female and male students decreased by 0.25, 0.30, and 0.05, respectively, compared to students without SI. Two-way between-groups analyses of variance (ANOVAs) were conducted to understand whether the impact of the SI program was statistically significant between student groups. The results indicated no significant interaction effects between SI and URM status F (1, 475) = 0.904, p = 0.342, FG status F (1, 475) = 1.340, p = 0.248, Pell grant status F (1, 475) = 0.050, p = 0.823, or gender F (1, 475) = 0.052, p = 0.819. To further explore which ethnic groups were impacted differently by the SI program. Student data were disaggregated by ethnic group and two-way ANOVA did not show significant difference, F (4, 445) = 0.834, p = 0.504. However, the exciting finding from the disaggregated data is that among all students, the SI program increased the Black/African American students' course GPAs the most (+1.38 ~a full grade and half), then followed by White, Hispanic/Latino, and Asian students (see Figure 3 and Table 5).

Table 4. Comparison of equity gaps between student groups.

Spring 2022 Group	SI (Credit)			Without SI			Mean Course GPA Difference
	N	Mean Course GPA	SD	N	Mean Course GPA	SD	
URM	148	2.49	1.01	205	1.50	1.20	+0.99
Non-URM	39	2.95	1.04	63	2.21	1.31	+0.74
Equity Gap		0.46			0.71		+0.25
FG	151	2.54	1.03	218	1.58	1.24	+0.96
Non-FG	46	2.80	1.06	64	2.14	1.33	+0.66
Equity Gap		0.26			0.56		+0.30
Pell	121	2.44	1.06	190	1.59	1.25	+0.85
Non-Pell	76	2.86	0.95	92	1.96	1.32	+0.90
Equity Gap		0.42			0.37		−0.05
Female	92	2.52	1.10	144	1.60	1.21	+0.92
Male	104	2.69	0.97	136	1.82	1.36	+0.87
Equity Gap		0.17			0.22		+0.05

Table 5. Comparison of the impact of SI on introductory chemistry course GPA by student ethnic group.

Spring 2022	SI (Credit)			Without SI			
	N	Mean Course GPA	SD	N	Mean Course GPA	SD	
Black/African American	10	2.66	1.05	20	1.28	1.15	+1.38
White	21	3.15	0.96	44	2.16	1.34	+0.99
Hispanic/Latino	137	2.49	1.01	184	1.53	1.21	+0.96
Asian	18	2.73	1.11	19	2.35	1.27	+0.38

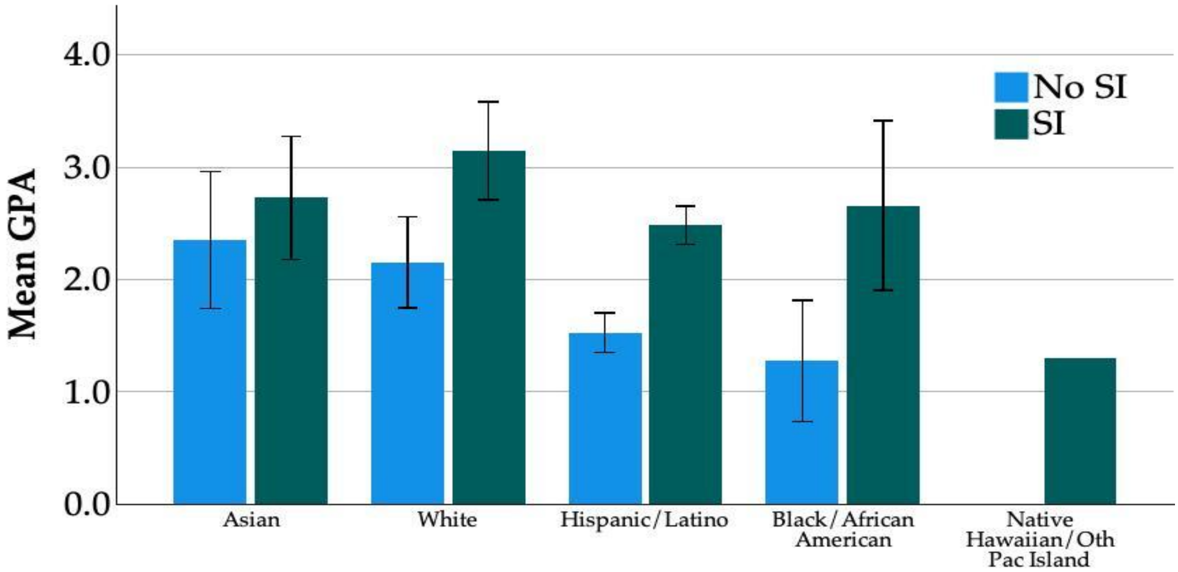


Figure 3. Comparison of introductory chemistry course GPA separated by ethnic groups; error bars: 95% confidence interval.

3.3. Other Factors Associated with the Effectiveness of the SI Program

Furthermore, other associated factors that might influence the effectiveness of the SI program were explored. Student data were separated by academic level, student achievement (separated by introductory chemistry course letter grade), and whether the introductory chemistry instructors and SI leaders made a difference. As shown in Figure 4, the SI program benefited the freshmen most among students with all academic levels and students who obtained the lowest grades of D, F, W, or WU among all letter grades. The results of the two-way ANOVAs showed that the impact of the SI program was statistically significant among different student achievement levels, F (3, 471) = 17.87, $p < 0.001$. However, the impact of the SI program was not statistically significant for students with different academic levels. One-way ANOVA results indicated that having SI classes significantly improved students' course GPA regardless of who the introductory chemistry instructor was. Additionally, another one-way ANOVA was conducted to explore the impact of SI leaders on introductory chemistry course GPA. The results indicated no statistically significant differences in introductory chemistry course GPA between students taught by different SI leaders.

3.4. Longitudinal Impacts of SI on the Subsequent General Chemistry I Course

Thus far, the effectiveness of the SI program towards students' academic outcomes and equity gaps in the target introductory chemistry course has been shown. It is worthwhile to examine whether the impacts of SI carry over to the subsequent chemistry course. To examine the longitudinal impact of the SI program, students' academic performance in the subsequent course (i.e., General Chemistry I) were compared. Of the 301 students from introductory chemistry in the Spring 2022 who were eligible to take General Chemistry I, 165 (55%) of them took the course in the Fall 2022 semester; 4 of those students received no SI credit and were excluded from the analysis. As listed in Table 6, students who took SI in introductory chemistry on average received approximately 0.36 higher course GPA (~half letter grade) in the General Chemistry I course compared to their counterparts who did not take SI. The DFW rate decreased by 15% for those with SI in introductory chemistry, suggesting that students with SI were more likely to be successful in the subsequent chemistry course.

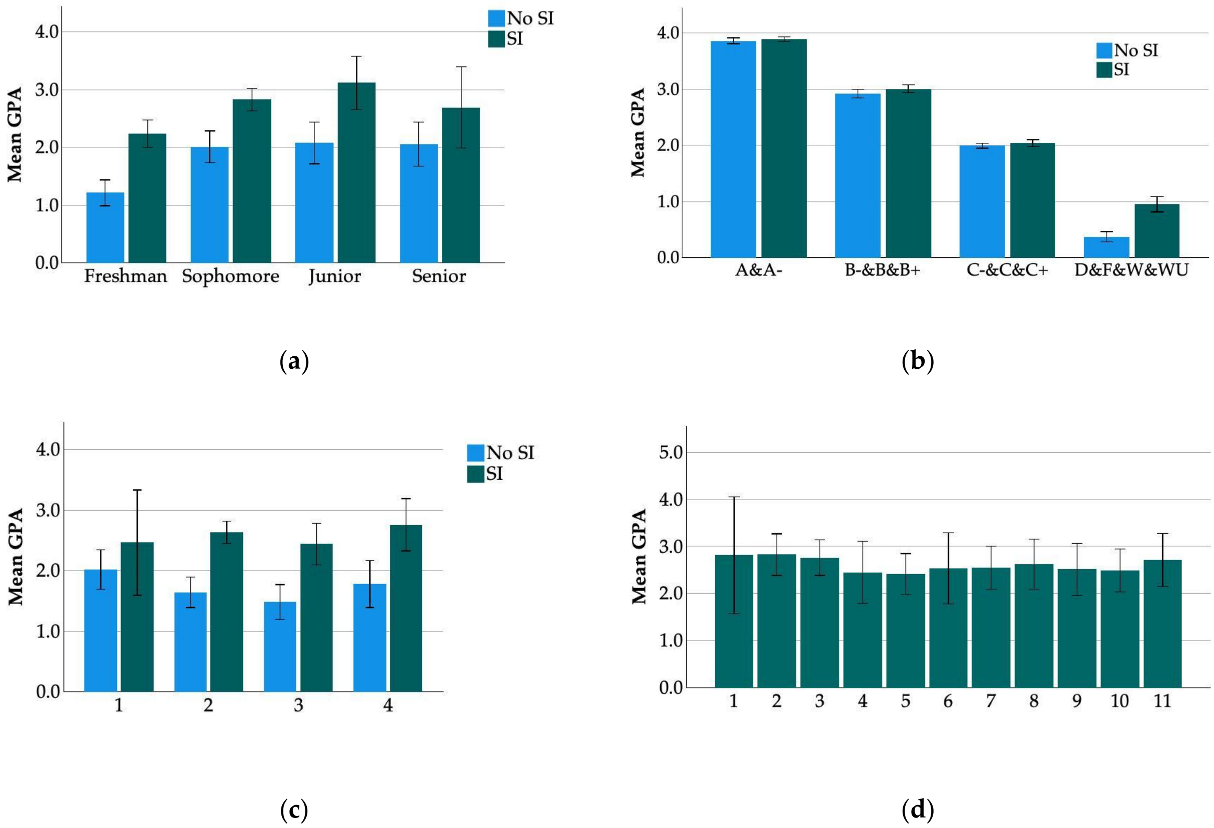


Figure 4. Comparison of course GPA separated by different factors: (**a**) academic level; (**b**) student achievement; (**c**) lecture instructor; (**d**) SI leader; error bars: 95% confidence interval.

Table 6. Comparison of the impact of SI on General Chemistry I course GPA and letter grades.

General Chemistry I Course GPA and Letter Grade in Fall 2022					
	Course GPA	**A**	**B**	**C**	**D, F, W**
Introductory Chemistry with SI Spring 2022 (n = 104)	2.03	9.6%	27.8%	33.6%	28.9%
Introductory Chemistry without SI Spring 2022 (n = 57)	1.67	7.0%	15.8%	33.3%	43.9%
Difference	+0.36	+2.2%	+12.0%	+0.3%	−15.0%

4. Discussion

Peterfreund and colleagues studied the impact of SI over six years on students' learning outcomes in twenty-two STEM courses from San Francisco State University [33]. They found that students performed better in all of the eight chemistry courses with SI than those without SI. About 16% of students were enrolled in optional SI sessions in four introductory chemistry courses (i.e., general chemistry I and II), and the SI improved average student course GPA by 0.24 and increased the average pass rate of the courses by 10%; URM students with SI had 0.28 higher course GPA and were 10% more likely to pass the courses than those URM students without SI.

Our studies showed more prominent results than the aforementioned study and this could be due to the nature of the SI, the structure of the chemistry lecture course, and its coordination with SI sessions. First, high-quality professional training for the SI leaders plays a critical role in the success of the program. In our study, the SI leaders are selected carefully and trained by an SI coordinator from the Learning Resource Center, who is dedicated to equipping SI leaders with pedagogical strategies, communication tips, role-playing activities, example materials, and monthly meetings to cultivate teaching and leadership skills and forming a learning community for leaders. Second, the introductory chemistry course coordinator ensures that the chemistry lecture instructors employ common textbooks, learning objectives, online homework, and final exams. The chemistry instructors are provided with and recommended to adopt a set of evidence-based instructional strategies, such as low-stakes frequent in-class poll questions, structured collaborative learning worksheets, and personalized messages. Third, the chemistry lecture coordinator works closely with the SI coordinator. Chemistry instructors are communicated with best practices for supporting SI leaders. The SI leaders attend every chemistry lecture class corresponding to the SI session they teach to ensure alignments between lecture classes and SI sessions. Such support structures reinforce the coordination across the stakeholders and ensure that students in the SI sessions are engaged actively in learning and these factors certainly contribute to the success of the SI program.

One of the limitations of the study is that the SI program was designed to be mandatory as a co-requisite for the introductory course but then only about half of the students were enrolled in SI in the Spring 2022 semester due to staffing issues and system errors. Students who chose to stay in the SI could be potentially more motivated than those who decided to drop SI. This session is written to provide a complete story of the redesign of the SI program at the research institution. Due to the historical high failure rates and the high enrollment of the introductory chemistry course, the Chemistry and Biochemistry Department started to collaborate with the Office of Student Success and Learning Resource Center on campus to redesign the existing SI program and planned to implement the new mandatory SI program in Fall 2021. Due to the COVID-19 disruption, the first round of the mandatory SI program was implemented virtually with all students in the virtual introductory chemistry course at our research institution. The impacts of the virtual SI program in Fall 2021 were compared to the student data in Spring 2021 when the students took a virtual introductory chemistry lecture without SI. The results show that the full-scale virtual mandatory SI program increased students' course GPA by 0.30 (~half a letter grade), final exam by 3%, pass rate by 8%, and decreased the DFW rate by 7% [55]. Although our virtual SI program was found to be beneficial for students' learning and enhance academic performance overall like some other peer-led cooperative learning programs implemented in chemistry [56–58], our results indicated that the equity gaps between underserved students and their counterparts were larger. These results might not be very significant because of the sudden change in the instruction from in-person to virtual for both lecture and SI classes and also could be due to the fact that those underserved students might be negatively impacted more by lack of resources such as a quiet learning environment and stable internet access at home to meet the needs of virtual learning [59,60]. As compared to the impacts of the virtual SI, the in-person SI program implemented in Spring 2022 improved student success metrics significantly more and made positive impacts on narrowing equity gaps between student groups. In Spring 2022, the main findings of this study showed the students with SI improved students' course GPA by 0.89 (~half a letter grade), the final exam by 5.4%, pass rate by 25%, and decreased the DFW rate by 24% compared to those without SI in the same semester. Because this is the first semester that students came back to in-person instruction from the pandemic disruption, the adjustments were certainly needed, and the overall improvement would have to take into account the transition back to in-person instruction.

Up to the present, the evaluation research of the revamped SI program is still ongoing. From Fall 2022 to Spring 2024, the revamped SI program has been implemented in full-scale for all students enrolled in the introductory course for two consecutive academic years. Figure 5 shows the trends of the students' DFW rates in the past ten years and how incorporating the SI program interrupted those trends. Due to the COVID-19 pandemic disruption, data from Spring 2020 to Fall 2021 were not included because instructions and assessments were virtual and considered as abnormal; therefore, student data prior to or after these times were used as "normal" with in-person teaching. These trends from Figure 5 indicated that the mandatory SI program successfully decreased the course DFW rates by almost half from the average DFW rate between Fall 2012 to Fall 2019 Fall (~39%) to the average DFW rate between Fall 2022 to Fall 2023 (~22%). This 17% drop suggests that the full scale of in-person SI significantly improved student success in the high impact introductory chemistry course and impact is substantial given that the average enrollment of the course is approximately 1600 students per year. The improvement of the student success in the gateway chemistry courses lifts the gate for students to advance to higher-level courses and move forward to their degrees and career pathways.

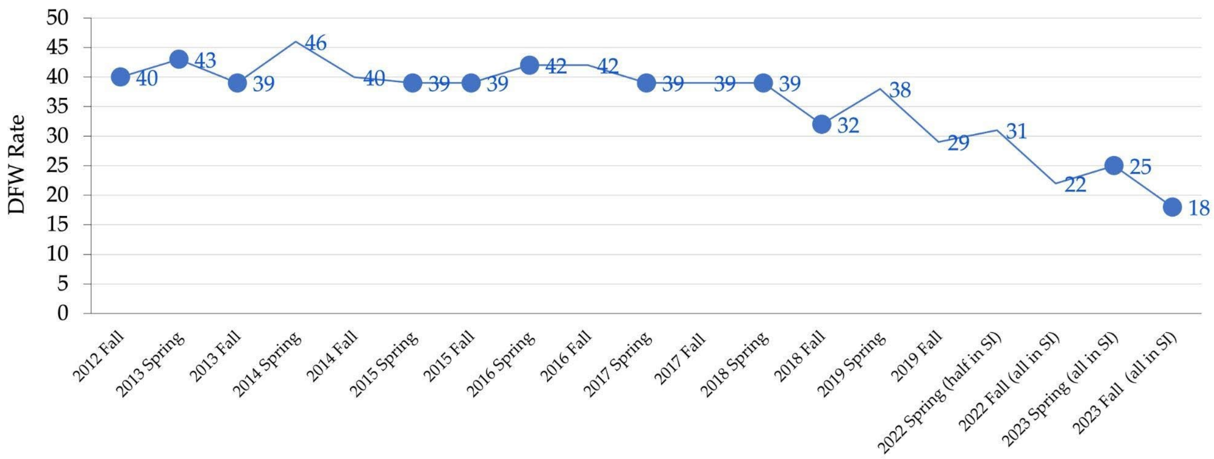


Figure 5. Trends of introductory chemistry DFW rates in the past ten years by semester.

Last but not least, longitudinal studies on how the SI program impacts students' graduation rates and choices of career pathways would also be interesting for future research. Such research will help the chemistry education research community to learn how the SI program addresses the STEM pipeline issues beyond the chemistry classrooms. Qualitative studies investigating stakeholders' perspectives are also needed to gain more insights and comprehensive understandings about the holistic impacts of similar supplemental academic programs. These might include but are not limited to how the SI program benefits the students for their sense of belonging in the science community, SI leaders' leadership and employable skills, and chemistry instructors' instructional approaches, and the mechanism and consistency of coordination between instructors and SI leaders.

5. Conclusions

The redesigned SI program made SI a co-requisite of the introductory chemistry lecture course and significantly improved student academic success in the target chemistry course at our research institution. Students with SI credit increased course GPA by 0.89 (~a full letter grade), final exam scores by 5.4%, pass rate by 25%, and decreased the DFW rate by 24% as compared to the students without SI. More exciting and worth noting, with this study sample, the SI program also showed promising impacts on closing the equity gaps between underserved and better served students, especially between URM and non-URM, FG and non-FG students. The breakdowns between more fine-grained ethnic groups showed that the SI program benefited all underserved student groups as compared to their counterparts without SI, including URM (+0.99), FG (+96), Pell grant (+0.85), and female students (+0.92) and benefited the Black/African American students (+1.38, ~one and half letter grade change) the most among all student groups.

Our study also indicated that other factors associated with the program's effectiveness are students' academic level and student achievement. The SI program impacted the freshmen most and the students with lowest achievement the most. Interestingly, data indicated that the effectiveness of the SI program does not depend on the chemistry course instructor, suggesting that the effectiveness of the program is consistent regardless of the instructional approaches in the associated chemistry lecture classes. Additionally, the benefits of the SI program did not stop at the introductory chemistry course and the data suggested the impact still continues in the subsequent General Chemistry I course by improving the course GPA by 0.36 (~half a letter grade).

These promising results indicated the need to continue the SI program as an academic intervention for addressing the STEM pipeline issues.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci14111196/s1, Introductory chemistry sample exam questions can be found in the Supplementary Material.

Author Contributions: Conceptualization, L.Y.; methodology, L.Y. and D.P.; data analysis, L.Y. and D.P.; writing—original draft preparation, L.Y. and D.P.; writing—review and editing, L.Y. and D.P.; visualization, L.Y. and D.P.; supervision, L.Y.; project administration, L.Y.; funding acquisition, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Provost and the Office of Student Success at California State University, Northridge.

Institutional Review Board Statement: The study has met the requirements of federal regulations governing human subject research of United Sates and approved by the Institutional Review Board of California State University, Northridge (IRB number is IRB-FY20-411 and date of approval was 10 July 2020).

Informed Consent Statement: Informed consent was waived due to the reason that only secondary data after the completion of the courses were used and the data were analyzed retrospectively.

Data Availability Statement: Data are unavailable due to privacy or ethical restrictions from the institution.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Koch, A.K.; Drake, B.M. *Digging into the Disciplines II: Failure in Historical Context—The Impact of Introductory U.S. History Courses on Student Success and Equitable Outcomes*; John N. Gardner Institute for Excellence in Undergraduate Education: Brevard, NC, USA, 2018; pp. 1–4.
2. Hatfield, N.; Brown, N.; Topaz, C.M. Do Introductory Courses Disproportionately Drive Minoritized Students out of STEM Pathways? *PANS Nexus* **2022**, *1*, pgac167. [CrossRef] [PubMed]
3. Reid, S.A. Design, Evolution, and Evaluation of a General Chemistry-Bridging Course. *Educ. Sci.* **2023**, *13*, 891. [CrossRef]
4. Shah, L.; Butler Basner, E.; Ferraro, K.; Sajan, A.; Fatima, A.; Rushton, G.T. Diversifying Undergraduate Chemistry Course Pathways to Improve Outcomes for At-Risk Students. *J. Chem. Educ.* **2020**, *97*, 1822–1831. [CrossRef]
5. Talbert, L.E.; Bonner, J.; Mortezaei, K.; Guregyan, C.; Henbest, G.; Eichler, J.F. Revisiting the Use of Concept Maps in a Large Enrollment General Chemistry Course: Implementation and Assessment. *Chem. Educ. Res. Pract.* **2019**, *21*, 37–50. [CrossRef]
6. Ye, L.; Eichler, J.F.; Gilewski, A.; Talbert, L.E.; Mallory, E.; Litvak, M.; Rigsby, E.M.; Henbest, G.; Mortezaei, K.; Guregyan, C. The Impact of Coupling Assessments on Conceptual Understanding and Connection-Making in Chemical Equilibrium and Acid–Base Chemistry. *Chem. Educ. Res. Pract.* **2020**, *21*, 1000–1012. [CrossRef]
7. Reid, S.A.; MacBride, L.; Nobile, L.; Fiedler, A.T.; Gardinier, J.R. Implementation and Evaluation of an Adaptive Online Summer Preparatory Course for General Chemistry: Whom Does It Benefit? *Chem. Educ. Res. Pract.* **2021**, *22*, 303–311. [CrossRef]
8. Hickey, D.T.; Robinson, J.; Fiorini, S.; Feng, Y. Internet-Based Alternatives for Equitable Preparation, Access, and Success in Gateway Courses. *Internet High. Educ.* **2020**, *44*, 100693. [CrossRef]
9. Bevly, D.; Maguire, J. Diversity/Inclusivity Style Guide. Available online: https://www.calstate.edu/csu-system/csu-branding-standards/editorial-style-guide/Pages/diversity-style-guide.aspx#low-income (accessed on 11 October 2024).
10. Harris, R.B.; Mack, M.R.; Bryant, J.; Theobald, E.J.; Freeman, S. Reducing Achievement Gaps in Undergraduate General Chemistry Could Lift Underrepresented Students into a "Hyperpersistent Zone". *Sci. Adv.* **2020**, *6*, eaaz5687. [CrossRef]
11. Stains, M.; Harshman, J.; Barker, M.K.; Chasteen, S.V.; Cole, R.; DeChenne-Peters, S.E.; Eagan, M.K.; Esson, J.M.; Knight, J.K.; Laski, F.A.; et al. Anatomy of STEM Teaching in North American Universities. *Science* **2018**, *359*, 1468–1470. [CrossRef]
12. Wang, Y.; Apkarian, N.; Dancy, M.H.; Henderson, C.; Johnson, E.; Raker, J.R.; Stains, M. A National Snapshot of Introductory Chemistry Instructors and Their Instructional Practices. *J. Chem. Educ.* **2024**, *101*, 1457–1468. [CrossRef]
13. Prince, M. Does Active Learning Work? A Review of the Research. *J. Eng. Educ.* **2004**, *93*, 223–231. [CrossRef]

14. Freeman, S.; Eddy, S.L.; McDonough, M.; Smith, M.K.; Okoroafor, N.; Jordt, H.; Wenderoth, M.P. Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8410–8415. [CrossRef] [PubMed]
15. Rahman, T.; Lewis, S.E. Evaluating the Evidence Base for Evidence-based Instructional Practices in Chemistry through Meta-analysis. *J. Res. Sci. Teach.* **2020**, *57*, 765–793. [CrossRef]
16. Moog, R.S. *Process Oriented Guided Inquiry Learning (POGIL)*; American Chemical Society: Washington, DC, USA, 2008.
17. Vincent-Ruz, P.; Meyer, T.; Roe, S.G.; Schunn, C.D. Short-Term and Long-Term Effects of POGIL in a Large-Enrollment General Chemistry Course. *J. Chem. Educ.* **2020**, *97*, 1228–1238. [CrossRef]
18. Hein, S.M. Positive Impacts Using POGIL in Organic Chemistry. *J. Chem. Educ.* **2012**, *89*, 860–864. [CrossRef]
19. Rodriguez, J.-M.G.; Hunter, K.H.; Scharlott, L.J.; Becker, N.M. A Review of Research on Process Oriented Guided Inquiry Learning: Implications for Research and Practice. *J. Chem. Educ.* **2020**, *97*, 3506–3520. [CrossRef]
20. Varma-Nelson, P.; Cracolice, M.S. *Peer-Led Team Learning: General, Organic, and Biological Chemistry*; Prentice Hall PTR: Hoboken, NJ, USA, 2000.
21. Wilson, S.B.; Varma-Nelson, P. Small Groups, Significant Impact: A Review of Peer-Led Team Learning Research with Implications for STEM Education Researchers and Faculty. *J. Chem. Educ.* **2016**, *93*, 1686–1702. [CrossRef]
22. Wamser, C.C. Peer-Led Team Learning in Organic Chemistry: Effects on Student Performance, Success, and Persistence in the Course. *J. Chem. Educ.* **2006**, *83*, 1562. [CrossRef]
23. Hockings, S.C.; DeAngelis, K.J.; Frey, R.F. Peer-Led Team Learning in General Chemistry: Implementation and Evaluation. *J. Chem. Educ.* **2008**, *85*, 990. [CrossRef]
24. Jansson, S.; Söderström, H.; Andersson, P.L.; Nording, M.L. Implementation of Problem-Based Learning in Environmental Chemistry. *J. Chem. Educ.* **2015**, *92*, 2080–2086. [CrossRef]
25. Aidoo, B.; Boateng, S.K.; Kissi, P.S.; Ofori, I. Effect of Problem-Based Learning on Students' Achievement in Chemistry. *J. Educ. Pract.* **2016**, *7*, 103–108.
26. Eichler, J.F.; Peeples, J. Flipped Classroom Modules for Large Enrollment General Chemistry Courses: A Low Barrier Approach to Increase Active Learning and Improve Student Grades. *Chem. Educ. Res. Pract.* **2016**, *17*, 197–208. [CrossRef]
27. Bokosmaty, R.; Bridgeman, A.; Muir, M. Using a Partially Flipped Learning Model to Teach First Year Undergraduate Chemistry. *J. Chem. Educ.* **2019**, *96*, 629–639. [CrossRef]
28. Ryan, M.D.; Reid, S.A. Impact of the Flipped Classroom on Student Performance and Retention: A Parallel Controlled Study in General Chemistry. *J. Chem. Educ.* **2016**, *93*, 13–23. [CrossRef]
29. Eberlein, T.; Kampmeier, J.; Minderhout, V.; Moog, R.S.; Platt, T.; Varma-Nelson, P.; White, H.B. Pedagogies of Engagement in Science: A Comparison of PBL, POGIL, and PLTL. *Biochem. Mol. Biol. Educ.* **2008**, *36*, 262–273. [CrossRef]
30. Clark, T.M. Narrowing Achievement Gaps in General Chemistry Courses with and without In-Class Active Learning. *J. Chem. Educ.* **2023**, *100*, 1494–1504. [CrossRef]
31. Frey, R.F.; Fink, A.; Cahill, M.J.; McDaniel, M.A.; Solomon, E.D. Peer-Led Team Learning in General Chemistry I: Interactions with Identity, Academic Preparation, and a Course-Based Intervention. *J. Chem. Educ.* **2018**, *95*, 2103–2113. [CrossRef]
32. Snyder, J.J.; Sloane, J.D.; Dunk, R.D.P.; Wiles, J.R. Peer-Led Team Learning Helps Minority Students Succeed. *PLoS Biol.* **2016**, *14*, e1002398. [CrossRef]
33. Mitchell, Y.D.; Ippolito, J.; Lewis, S.E. Evaluating Peer-Led Team Learning across the Two Semester General Chemistry Sequence. *Chem. Educ. Res. Pract.* **2012**, *13*, 378–383. [CrossRef]
34. Lewis, S.E. Investigating the Longitudinal Impact of a Successful Reform in General Chemistry on Student Enrollment and Academic Performance. *J. Chem. Educ.* **2014**, *91*, 2037–2044. [CrossRef]
35. Arendale, D.R. *Understanding the Supplemental Instruction Model*; Josey-Bass: Hoboken, NJ, USA, 1994.
36. Widmar, G.E. Supplemental Instruction: From Small Beginnings to a National Program. *New Dir. Teach. Learn.* **1994**, *60*, 3–10. [CrossRef]
37. Helde, R. Supplemental Instruction (SI)-Learning Leadership and Leadership Development. In *Supplemental Instruction: Volume 3: Organisation and Leadership*; Waxmann: Munster, NY, USA, 2021; pp. 53–74.
38. Ellison, I.; Haynes, G.; Toglia, J. The Perceived Impact of Leading Supplemental Instruction on Student Leaders at a Hispanic-Serving Institution. *Educ. Sci.* **2023**, *13*, 1144. [CrossRef]
39. Dawson, P.; van der Meer, J.; Skalicky, J.; Cowley, K. On the Effectiveness of Supplemental Instruction: A Systematic Review of Supplemental Instruction and Peer-Assisted Study Sessions Literature Between 2001 and 2010. *Rev. Educ. Res.* **2014**, *84*, 609–639. [CrossRef]
40. Peterfreund, A.R.; Rath, K.A.; Xenos, S.P.; Bayliss, F. The Impact of Supplemental Instruction on Students in Stem Courses: Results from San Francisco State University. *J. Coll. Stud. Ret.* **2008**, *9*, 487–503. [CrossRef]
41. Ogden, P.A.D.; Thompson, D.A.D.; Russell, A.A.D.; Simons, C.A.D. Supplemental Instruction: Short- and Long-Term Impact. *J. Dev. Educ.* **2003**, *26*, 2–4.
42. Yue, H.; Rico, R.S.; Vang, M.K.; Giuffrida, T.A. Supplemental Instruction: Helping Disadvantaged Students Reduce Performance Gap. *J. Dev. Educ.* **2018**, *41*, 18–25.

43. Achat-Mendes, C.; Anfuso, C.; Johnson, C.; Shepler, B. Learning, Leaders, and STEM Skills: Adaptation of the Supplemental Instruction Model to Improve STEM Education and Build Transferable Skills in Undergraduate Courses and beyond. *J. STEM Educ. Innov. Res.* **2020**, *2*, 14–22.
44. Ochse, R. Writing Partners: Improving Writing and Learning through Supplemental Instruction in Freshman Writing Classrooms. In Proceedings of the Annual Meeting of the National Council of Teachers of English, San Diego, CA, USA, 16–21 November 1995.
45. Fayowski, V.; MacMillan, P.D. An Evaluation of the Supplemental Instruction Programme in a First Year Calculus Course. *Internat. J. Math. Ed. Sci. Tech.* **2008**, *39*, 843–855. [CrossRef]
46. Moore, R.; LeDee, O. Supplemental Instruction and the Performance of Developmental Education Students in an Introductory Biology Course. *J. Coll. Read. Learn.* **2006**, *36*, 9–20. [CrossRef]
47. Ellison, I.; Esser, F.; Walsh, J.; Lucio, C.; Toglia, J. Improving Student Success through Supplemental Instruction in an Anatomy and Physiology Laboratory. *Educ. Sci.* **2023**, *13*, 1153. [CrossRef]
48. Rath, K.A.; Peterfreund, A.; Bayliss, F.; Runquist, E.; Simonis, U. Impact of Supplemental Instruction in Entry-Level Chemistry Courses at a Midsized Public University. *J. Chem. Educ.* **2012**, *89*, 449–455. [CrossRef]
49. Salame, I.I. The Impact of Supplemental Instruction on the Learning Achievements and Attitudes of Organic Chemistry Students. *Interdiscip. J. Environ. Sci. Educ.* **2020**, *17*, e2232. [CrossRef]
50. Guyot, M.; Hsu, S.; Nizomov, J.; Antonenko, P.; Habenicht, S. Examining How Supplemental Instruction Impacts Students' Motivation in Organic Chemistry. *Int. J. Chem. Educ. Res.* **2024**, *8*, 7–16. [CrossRef]
51. Stanich, C.A.; Pelch, M.A.; Theobald, E.J.; Freeman, S. A New Approach to Supplementary Instruction Narrows Achievement and Affect Gaps for Underrepresented Minorities, First-Generation Students, and Women. *Chem. Educ. Res. Pract.* **2018**, *19*, 846–866. [CrossRef]
52. Sevian, H.; King-Meadows, T.D.; Caushi, K.; Kakhoidze, T.; Karch, J.M. Addressing Equity Asymmetries in General Chemistry Outcomes Through an Asset-Based Supplemental Course. *JACS Au* **2023**, *3*, 2715–2735. [CrossRef]
53. Fraenkel, J.; Wallen, N.; Hyun, H. *How to Design and Evaluate Research in Education*, 10th ed.; McGraw Hill Education: New York, NY, USA, 2019.
54. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1998.
55. Pham, D. Lifting Gate: Peer-Led Cooperative Learning Supplemental Instruction (SI) Improved Student Success Postsecondary Introductory Chemistry. Master's Thesis, California State University, Northridge, CA, USA, 2023.
56. Young, J.D.; Lewis, S.E. Evaluating Peer-Led Team Learning Integrated into Online Instruction in Promoting General Chemistry Student Success. *J. Chem. Educ.* **2022**, *99*, 1392–1399. [CrossRef]
57. Smith, J.; Wilson, S.B.; Banks, J.; Zhu, L.; Varma-Nelson, P. Replicating Peer-Led Team Learning in Cyberspace: Research, Opportunities, and Challenges. *J. Res. Sci. Teach.* **2014**, *51*, 714–740. [CrossRef]
58. Wilson, S.B.; Varma-Nelson, P. Characterization of first-semester organic chemistry peer-led team learning and cyber peer-led team learning students' use and explanation of electron-pushing formalism. *J. Chem. Educ.* **2018**, *96*, 25–34. [CrossRef]
59. Wu, F.; Teets, T.S. Effects of the COVID-19 pandemic on student engagement in a general chemistry course. *J. Chem. Educ.* **2021**, *98*, 36333642. [CrossRef]
60. Bahrami, P.; Blanco, D.F.; Thetford, H.R.; Ye, L.; Chan, J.Y.K. Capturing Student and Instructor Experiences, Perceptions, and Reflections on Remote Learning and Teaching in Introductory Chemistry Courses During COVID-19. *J. Coll. Sci. Teach.* **2023**, *52*, 6–9. [CrossRef]

education sciences

MDPI

Review

Delving into the Design and Implementation of Specifications Grading Systems in Higher Education

William J. Howitz [1], Kate J. McKnelly [2] and Renée D. Link [3,*]

[1] Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA; whowitz@umn.edu
[2] Department of Chemistry, Emory University, Atlanta, GA 30322, USA; kate.joy.mcknelly@emory.edu
[3] Department of Chemistry, University of California—Irvine, Irvine, CA 92697, USA
* Correspondence: rlink@uci.edu

Abstract: Specifications grading is an alternative grading system that has been used with increasing frequency in higher education. Since first introduced by Linda Nilson in 2014, more than 90 publications on the design and implementation of specifications grading systems have been published. This work presents a systematic review of the current literature to analyze the variety of ways specifications grading systems are executed, including the diverse design and implementation considerations, as well as to present and discuss emergent themes. We analyzed 90 publications and present their relevant findings in the results. The following databases were last searched on 5 October 2024 for publications: IEEE Xplore, ACS Publications, ASEE PEER, PER, Scopus, ERIC, ACM, ScienceDirect, and Web of Science. All peer-reviewed journal articles, conference proceedings, and book chapters that implemented at least two structural features of specifications grading in an undergraduate or graduate course were included in this review. Theses, dissertations, conference abstracts, posters, workshops, blogs, opinion pieces, social media exchanges, and content provided on websites were not included. Additionally, reports of specifications grading systems in K-12 courses or those that only presented the design and/or implementation of less than two structural features of the grading system were similarly excluded. Our findings from the literature reveal that the following themes emerge from educators who use specifications grading: instructor commentary on time investment, academic performance, and student reactions to specifications grading. This review provides a resource for those interested in exploring this alternative grading system, and the emergent themes indicate that there are ripe opportunities for future study.

Keywords: specifications grading; alternative grading; assessment

1. Introduction

What is a course grade meant to represent? In U.S. higher education, course letter grades are determined frequently by the number of overall points a student accumulates throughout a term culminating with a final letter grade assigned on an A-F scale (Schinske & Tanner, 2014; Brookhart et al., 2016). The A–F scale has been the dominant grading scheme in the U.S. since the mid-1900s (Durm, 1993). The first documented use of the A–F scale defined each letter grade as follows: A is excellent, B is good, C is fair, D is passed, and F is failed. However, what each letter grade means at each institution, course, etc. is not standardized.

The importance of letter grades to earn degrees, maintain scholarships, and gain access to graduate and professional programs incentivizes students to focus on accumulating points (an extrinsic motivation) rather than on learning (an intrinsic motivation) (Schinske

& Tanner, 2014). Furthermore, when we assign points and partial credit to student work and sum the points, there is often no direct connection between the grade students earn and what course learning outcomes (LOs) they have achieved (Brookhart et al., 2016; Nilson & Stanny, 2014). Grades become a ranking of students against each other rather than an indication of the achievement of the LOs by each individual student (Brookhart et al., 2016). This ranking issue is exacerbated in courses that employ curves to determine how many points correspond to a letter grade (Seymour, 1997; Bowen & Cooper, 2022).

The problematic use of points to assign grades causes further challenges. The focus on accumulating points sets up an antagonistic relationship between the instructor and students, instills competition between students, and amplifies student and faculty stress, anxiety, and mental health issues (Nilson & Stanny, 2014; Eyler, 2024; Hammoudi Halat et al., 2023). A feature of points-based grading systems is the inclusion of partial credit allotment. This inclusion increases the time-consuming faculty activity of meeting with students who argue for partial credit. Another feature of points-based grading systems is often including high-stakes assignments with no opportunities for showing proficiency or competency with feedback and opportunities to try again. This feature benefits students who come from more privileged backgrounds and penalizes students from minoritized groups (Smeding et al., 2013).

Alternative grading practices, including mastery grading, standards-based grading (SBG), contract grading, and specifications grading, have been developed to address these challenges with traditional, points-based grading systems (Clark & Talbert, 2023).

In a mastery grading system, a student's work on an assessment of a topic or learning outcome must meet the instructor's established performance threshold before the student can move on to a new topic or learning outcome (Bloom, 1968; Kulik et al., 1990). Because the assessment does not count toward the student's final course grade until the work meets the defined threshold, students must be provided multiple attempts without penalty. These reassessment opportunities lower the stakes of assignments and provide an incentive for students to use instructor-provided feedback to demonstrate learning over time.

In a standards-based grading (SBG) system, a student's performance on an assessment is evaluated with respect to one or more of the course learning outcomes, and the number and/or type of learning outcomes a student meets determines their final letter grade (Knight & Cooper, 2019). Partial credit and points are removed entirely in an SBG system. The incentive for students to focus on accumulating points is eliminated, minimizing antagonistic interactions between the instructor and students over the points and partial credit allocated to an assignment. Similarly to mastery grading, SBG also permits multiple attempts without penalty, encouraging students to leverage instructor feedback provided on assignments to address the gaps in their knowledge to meet the course learning outcomes.

Contract grading establishes a collaborative relationship between the student and the instructor, returning agency to students by giving them more control over the way in which their course grade is determined. In a contract grading system, each student negotiates a contract with the instructor that specifies what assessments they need to complete and how the assessments must be completed to earn a predetermined final letter grade chosen by the student (Hassencahl, 1979; Taylor, 1980; Beare, 1986). The transparency inherent in contract grading provides students with clear expectations and a roadmap to achieve their desired final letter grade.

The alternative grading systems described above employ a variety of frameworks, including mindset theory (Dweck, 2013; Harsy et al., 2021; Lewis, 2022), achievement goal orientation theory (Elliot & McGregor, 2001; Elsinger & Lewis, 2020), and self-efficacy theory (Bandura, 1986; Carberry et al., 2012). However, the extent to which these frameworks and others are discussed in the literature on alternative grading systems in U.S. higher education is limited (Hackerson et al., 2024). Clark and Talbert describe the "four pillars of alternative grading" as a practical framework that can support instructors in designing effective alternative grading systems (Clark & Talbert, 2023). Their framework identifies the features of alternative grading systems that focus on student learning over time: (1) clearly defined standards describing acceptable evidence of learning, (2) actionable feedback used by students to improve their learning, (3) marks on student work that indicate progress toward meeting standards, and (4) reassessment of student work without penalty to provide students with opportunities to use the actionable feedback.

Specifications grading, first reported by Nilson in 2014 (Nilson & Stanny, 2014), combines aspects of the previous grading systems and incorporates the four pillars described by Clark and Talbert (Elkins, 2016; Howitz et al., 2021). Leslie and Lundblom provide the following summary of the principles underlying specifications grading (Leslie & Lundblom, 2020).

The core principles of specifications grading are:

- Course assignments are aligned with course learning objectives.
- Expectations ("specifications") are clear.
- Students decide what grade they aim for (self-imposing learning demands consistent with the grade).
- Feedback relates expectations to performance.
- Defined (and limited) options are provided for revisions.
- Assignments are completed at a clearly defined level of performance (e.g., corresponding to a grade of B or C) to demonstrate competency.
- Advanced learning options in breadth and/or depth are offered for self-motivated students.

While the principles delineated above describe the underpinnings of specifications grading, these principles must be operationalized. In practice, the following structural components comprise specifications grading systems (Nilson & Stanny, 2014):

- At the course level, students are provided with defined grade bundles that clearly delineate what assignments they need to complete and at what level to earn their chosen letter grade.
- At the assignment level, where assignments encompass all work submitted by a student, including homework, quizzes, papers, exams, etc., students are provided clear rubrics that contain the specifications required for the assignments and the set threshold they must achieve to demonstrate competency. Partial credit is not available.
- Students are provided with one or more mechanisms by which they are able to revise work that does not meet the required specifications.
- Optionally, a token system that can be used to limit opportunities for revisions and provide flexibility to students in how they navigate the course is provided.

Numerous examples of specifications grading systems have been published in the primary teaching and education literature of many individual disciplines. A recent scoping review by Hackerson et al. highlights alternative grading systems in science, technology, engineering, and mathematics (STEM) courses (Hackerson et al., 2024). Harrington et al. examined the body of literature on contract grading and specifications grading in computer science courses (Harrington et al., 2024). However, reviews of the existing literature focused solely on describing implementations of specifications grading across all disciplines do not exist. Here we provide a review of publications describing the design and implementation of specifications grading systems in all disciplines across higher education. Our aims in this review are (1) to provide a resource for instructors designing their own specifications grading systems and (2) to build a roadmap for education researchers to facilitate collaboration with practitioners to study outcomes and impacts of specifications grading in higher education based on emergent themes in the current literature.

2. Methods

2.1. Research Questions

Our research was guided by the following questions regarding specifications grading systems described in the literature:

1. What is the current landscape of peer-reviewed literature describing implementations of specifications grading in higher education?
2. What are the structures of the specifications grading system implementations currently described in the literature?
3. What themes have emerged in the literature on specifications grading that have not yet been studied systematically?

2.2. Article Selection and Analysis

We conducted a literature search following the PRISMA guidelines using the keywords "specifications grading," "specs grading," "specifications-based grading," and "alternative grading" to find relevant publications ranging from October 2014 through September 2024 (Figure 1). October 2014 was chosen as the start date because it was when Linda Nilson's book on the subject was published. The databases that were searched included the following: Web of Science, Scopus, the Education Resources Information Center (ERIC), Institute of Electrical and Electronics Engineers (IEEE) Xplore, ScienceDirect, the American Chemical Society (ACS) Publications, the Association for Computing Machinery (ACM) Digital Library, the American Society for Engineering Education (ASEE) Papers on Engineering Education Repository (PEER), and Physics Education Research (PER) Central. The number of publications identified using the search terms from each database is specified in Table 1. Several publications were identified through more than one database. Citation searching was used to locate any publications not found through the aforementioned databases. A total of 21 additional publications that met the inclusion criteria described below were found by this method.

Table 1. Total number of publications found and included in this review, separated by the search database in which the publications were found. The total number of included publications below is greater than the number of publications in the corpus because multiple publications were found in more than one database.

Database	Publications Identified Using Search Terms After Removing Duplicates	Publications Included in Review Corpus After Applying Inclusion and Exclusion Criteria
IEEE Xplore	1	1
ACS Publications	67	17
ASEE PEER	123	17
PER	2	0
Scopus	96	56
ERIC	22	14
ACM	18	3
ScienceDirect	3	1
Web of Science	40	27

Because this review was intended to characterize the landscape of empirical research on specifications grading in the higher education (undergraduate- or graduate-level) setting, we focused on peer-reviewed publications (Table 2). Theses, dissertations, and informal means of communication such as blogs, opinion pieces, social media exchanges, and websites were excluded from the corpus as they are not peer-reviewed. We limited the scope of this review to publications that describe the design and implementation of specifications grading systems. Publications that discussed the design of a specifications grading system but did not implement it or that referenced the implementation of specifications grading without additional detail or expansion were excluded from the analysis. Specifically, each publication was only included if commentary on the implementation of at least two of the four structural features of specifications grading—grade bundles, rubrics with specifications and defined passing thresholds, opportunities to revise and resubmit work, and a token system—was included. As such, journal articles, conference papers, and book chapters were included, but conference abstracts, posters, and workshops were excluded as they lacked sufficient detail to meet the inclusion criteria. The total number of publications included in the corpus was 90. The full corpus of manuscripts included in this review is available in Table S1.

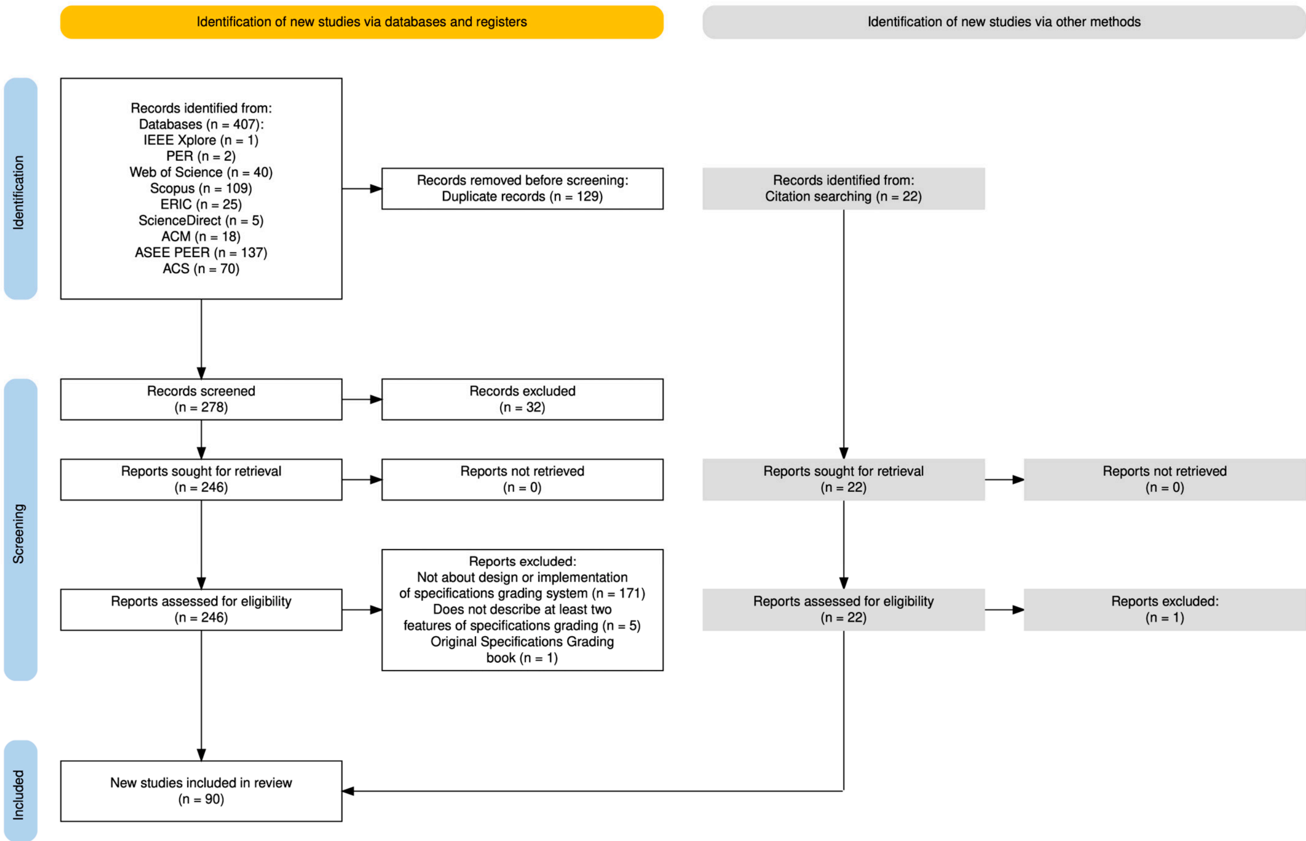


Figure 1. Diagram of this study corpus selection process generated with the PRISMA flow diagram Shiny app (Haddaway et al., 2022).

Table 2. Inclusion and exclusion criteria used in the analysis conducted for this review.

Inclusion Criteria	Exclusion Criteria
Peer-reviewed empirical research, including journal articles, conference proceedings, and book chapters	Theses, dissertations, conference abstracts, posters, workshops, blogs, opinion pieces, social media exchanges, websites
Implementation of at least two structural features of specification grading	Design only or implementation of less than two structural features of specifications grading
Undergraduate and graduate populations	K-12 populations

Two reviewers screened each publication for inclusion independently. Data from each publication were extracted by one reviewer and verified by a second reviewer. Because most publications included in the review were descriptive studies conducted by instructors on their own courses, the risk of bias in individual studies was not assessed. Characteristics of courses were taken directly from the descriptions included in the publications, and disciplines were assigned based on the name or description of the course(s).

3. Results

3.1. Publication Trends

Following Nilson's book in 2014, the first publications on the design and implementation of specifications grading appeared in 2016. In every year after 2016, there have been at least five publications, with 2023 as the year with the most publications at 22 (Figure 2). From January to 30 September 2024, a total of 12 publications were released.

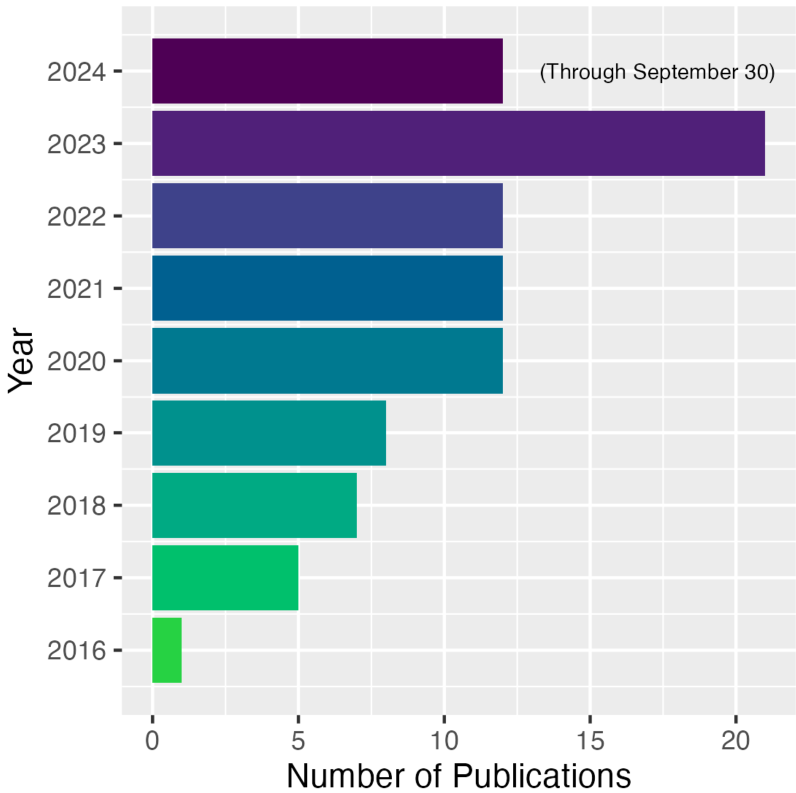


Figure 2. Number of peer-reviewed publications describing specifications grading systems published annually, January 2016–September 2024. There were no publications in 2014 or 2015 aside from Nilson's book.

STEM disciplines are most represented in publications describing specifications grading systems (Figure 3). Chemistry represents the greatest number of publications with 20, followed by engineering with 19, computer science with 10, and mathematics with 8. All other disciplines have four or fewer publications, and disciplines within the humanities are least represented. The type of publications also varies by discipline. While journal articles are most common amongst nearly all disciplines, conference papers are the only type of publication from engineering. Book chapters are almost exclusive to chemistry, with the only exceptions being one publication from information literacy, one from French studies, and one from computer science. These discrepancies in types of dissemination are tied to disciplinary norms. In engineering and computer science, peer-reviewed conference papers are more common than journal articles, and the American Chemical Society specifically publishes peer-reviewed books as an alternative form of dissemination.

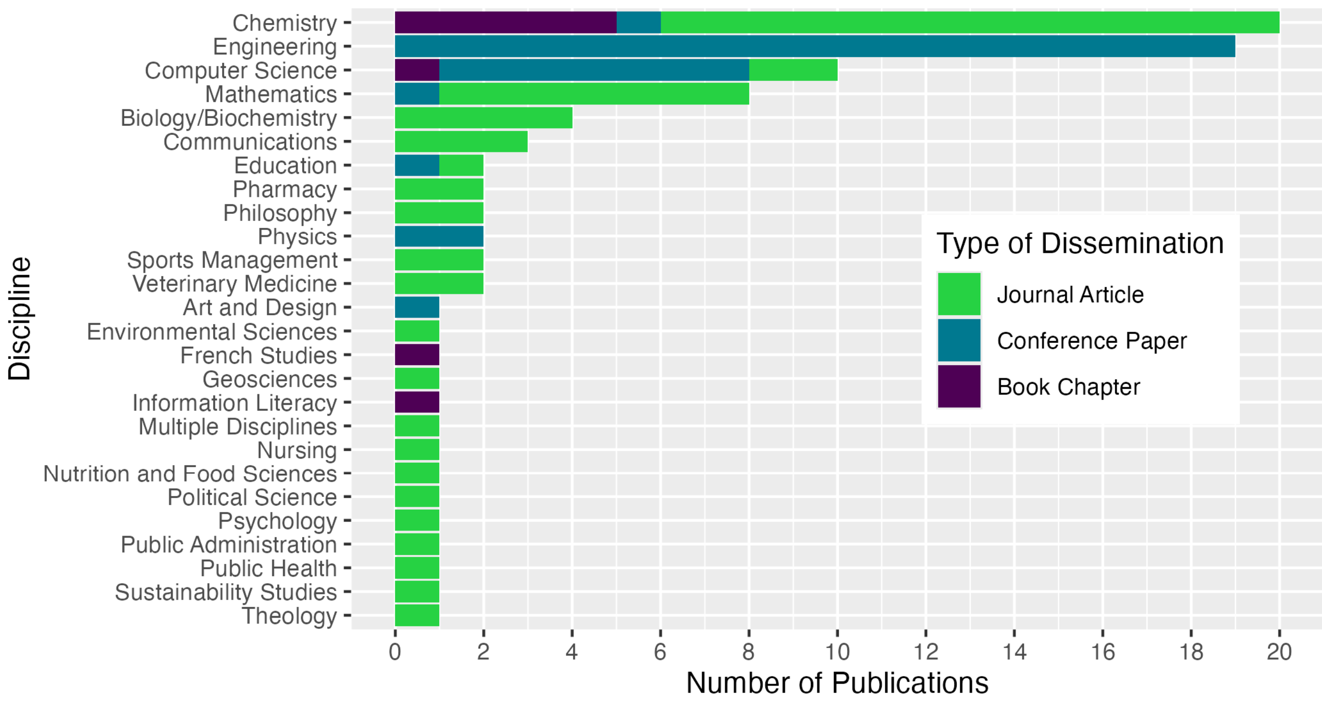


Figure 3. Peer-reviewed publications describing specifications grading systems, organized by discipline and type of dissemination.

Descriptions of course sizes vary from publication to publication. Some include descriptions of a single course or multiple courses with no information about size, a single course with information about size, or multiple courses with information about size. Course sizes were extracted from the descriptions provided in the publications and were grouped into five categories: 35 or fewer, 36–60, 61–100, 101–999, and 1000 or greater. Of the courses described that include information about course size, 63% (56 courses) are small courses with total enrollments of 35 or fewer students (Figure 4). The commonality of this course size is unsurprising, as the majority of these publications describe courses taught at smaller colleges and universities. The majority of published implementations of specifications grading in courses with enrollments of 100 students or greater describe introductory-level courses, mostly at large, four-year institutions. Although most published examples of specifications grading describe courses at the undergraduate level, including independent study or capstone courses (Earl, 2021; Mendez, 2024; Martin et al., 2021; Gargac, 2022a; Fernandez et al., 2020), a small number of graduate and professional courses do appear (Blodgett, 2017; Dennen & Bagdy, 2020; Gay & Poproski, 2023; Hofmeister et al., 2022a, 2022b; Jones, 2020; Quintana & Quintana, 2020; Walden, 2022; Joseph et al., 2023; Joshi, 2023; Dupree et al., 2024; Moster & Zingales, 2024; Santucci & Golas, 2023). Of the 90 publications included in the review corpus, 20 publications describe implementations of specifications grading in online or hybrid modality courses (Elkins, 2016; Dennen & Bagdy, 2020; Gay & Poproski, 2023; Quintana & Quintana, 2020; Moster & Zingales, 2024; Santucci & Golas, 2023; Mendez, 2018a, 2019; Houseknecht & Bates, 2020; Shields et al., 2019; Gestwicki, 2021; Gratwick et al., 2020; Wasniewski et al., 2021; Evensen, 2022; Gargac, 2022b; Dabney & VanDerWoude, 2023; Suresh, 2023; Closser et al., 2024; Kinnear et al., 2022; Johanesen et al., 2024).

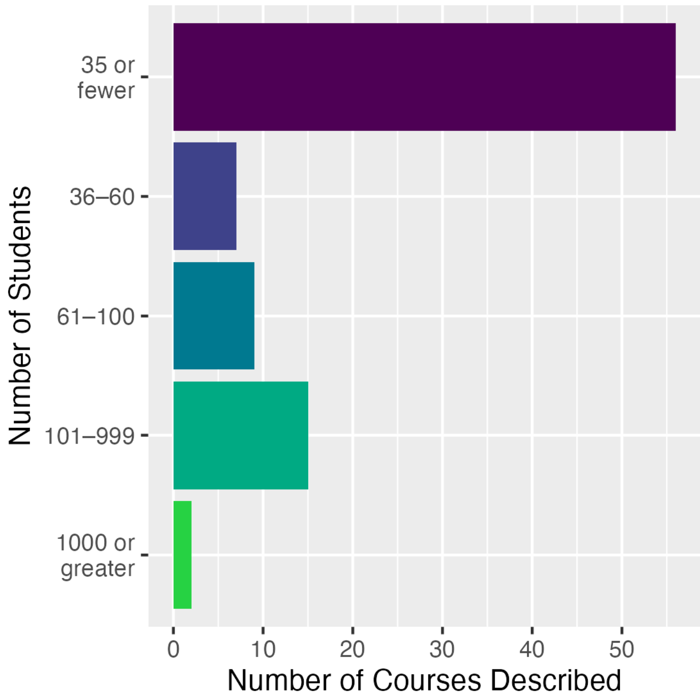


Figure 4. Peer-reviewed publications describing specifications grading systems organized by course size. The total number of courses described does not match the total number of publications included in the corpus, as 11 publications describe the design and implementation of specifications grading in more than one course.

3.2. Structures of Specifications Grading Systems

The four structural components of specifications grading—grade bundles, rubrics with specifications and defined passing thresholds, opportunities to revise and resubmit work, and a token system—are incorporated into courses in a variety of ways and to different degrees. In this section we discuss three of the four structural components independently, giving context for how they have been implemented. The extent to which retakes and revisions are allowed varies, depending on how the other structural components were designed. As such, a conversation about this component will be threaded throughout the next three subsections.

3.2.1. Grade Bundles

A core principle of specifications grading systems is to align assignments with course LOs, and a key structure of these systems is bundling these assignments together to determine course letter grades (Nilson & Stanny, 2014). Four main methods of bundling, which we call configurations, emerged in the analysis of peer-reviewed descriptions of specifications grading. Tsoi et al. originally described three main configurations (called "implementations") in the context of lecture courses: core and additional LOs, all equal LOs, and modules (Tsoi et al., 2019). A fourth configuration best described as all equal LOs with repetition and/or complexity (ELORC) arose in the context of laboratory and writing courses.

In the core and additional LOs configuration, all course LOs are sorted into "core" and "additional" categories. "Core" LOs are those deemed by the instructor to be essential for earning a grade of C or above. All other LOs are categorized as "additional," and meeting a larger subset of these LOs is required to earn B or A grades. Frequently this configuration is used in introductory lecture courses that serve as prerequisites, where having fundamental knowledge and skills is essential for success in subsequent courses.

While the terms “core” and “additional” arose from Tsoi et al., a number of publications use different terminology, such as core and advanced LOs or essential and general LOs, to refer to this same configuration (Carlisle, 2020; Ring, 2017). An example of a course with the core and additional LOs configuration is Carlisle’s discrete and combinatorial algebra course, in which each of the six major subject areas is divided into a core topic and an advanced topic (Carlisle, 2020). To earn at least a C grade, students must pass all six core topics and one advanced topic. If a student wishes to earn a B grade, they need to pass an additional two advanced topics, and if a student wishes to earn an A grade, they need to pass an additional four advanced topics. Similarly, LOs in Biers’s first-year French course are classified as basic or advanced and are grouped into three main categories: proficiency, cultural knowledge, and metalinguistic awareness (Biers, 2022). To earn at least a C grade, students must pass all of the basic LOs in all three categories. If a student wishes to earn a B grade, they also need to pass advanced LOs from one of the three categories (or two of the categories if they wish to earn an A).

In an all equal LOs configuration, all course LOs are given equal priority. Students earn their letter grade based on the total number of LOs met. Frequently this configuration is used in advanced or elective courses. For example, in Carlisle’s differential equations course, the content is split into 20 “problem topics.” Students must pass 14 of the topics to earn a C, 17 to earn a B, and all 20 to earn an A (Carlisle, 2020). Similarly, in Mendez’s sophomore-level thermodynamics course, there are 15 LOs, and each is assessed by one quiz (Mendez, 2018a). Students must pass 11 quizzes to earn a C, 13 for a B, and all 15 for an A.

In the modules configuration, thematically related course LOs are binned together into modules. To pass a module, students must meet a predetermined number of the LOs in that module. According to Tsoi et al., the module configuration is used when the “. . .skills and knowledge central to the course. . .(cannot). . .be distilled into discrete objective statements without negatively impacting the student learning in the course (Tsoi et al., 2019).” While Tsoi et al. list modules as a third, standalone configuration, the modules configuration could be considered a subcategory of either the “core and additional LOs” or the “all equal LOs” configurations, depending on how the instructor chooses to bundle the module for letter grades. Tsoi et al. indicate that modules may be classified as “essential” (consistent with the core and additional LOs configuration), while in other cases, modules may be ranked equally (consistent with the all equal LOs configuration).

The three configurations described by Tsoi et al. all emerged in the context of lecture courses, whereas the fourth ELORC configuration emerged in the context of laboratory and writing courses. Lecture courses generally have a large number of LOs to cover all of the content knowledge students must learn, especially at the introductory level. In contrast to lecture courses, laboratory and writing courses tend to be more focused on the practical application of knowledge and the development of technical skills. In these types of courses, there are often fewer course LOs, and they are often bundled in ways that require students to meet the same LO(s) multiple times and potentially at varying levels of complexity to earn higher grades. In the context of a laboratory course, LOs align with students both applying knowledge they have learned and developing practical technical skills to obtain and analyze data. In the context of a writing course, LOs focus on students practicing the processes of brainstorming, outlining, drafting, editing, and polishing various pieces of writing. An example of a course with this configuration is McKnelly et al.’s Writing for Chemists course (McKnelly et al., 2021). In this course, there are four large writing assignments that all assess the same course LO, “Students will be able to create professional papers, proposals, reports, and other forms of scientific writing.” These four assignments also assess different amounts of additional LOs. Students must earn a low pass on three of the assignments to earn a C, a low pass on all four assignments to earn a B, and a high pass on three and a low pass on one to earn an A. (Passing threshold levels will be discussed in the next section).

Although Nilson's original introduction of specifications grading specifically called for designing assignments and grade bundles that are aligned with course LOs, this alignment is not always clearly demonstrated in publications describing specifications grading systems. Some instructors described grade bundles (and thus implied configurations) based on assignments but did not clearly describe how these assignments were aligned with course LOs. Tsoi et al. proposed the configurations with the assumption that individual assignments are mapped to individual course LOs. However, in some of the courses described in the specifications grading literature, one assignment may correspond to one LO, one assignment may correspond to multiple LOs or multiple assignments may correspond to the same LO. These discrepancies are reflected in Yik et al.'s analysis of specifications grading in chemistry courses in which the grade bundles observed for some courses reflect a focus on specific assignments rather than specific LOs (Yik et al., 2024). Without clear descriptions of alignment between assignments and LOs in some publications included in our review, it was not always possible to determine what type of LO-focused configuration was being employed.

3.2.2. Rubrics with Specifications and Passing Thresholds

Within a specifications grading system, the specifications are embedded as the rubric criteria for assignments. In general, student work is evaluated against each rubric criterion or specification, and then the assignment outcome is determined by performance across all of the specifications set for the assignment. The most common way to determine if students have met a specification is using a binary system, which in practice generally appears as a student earning credit for a rubric criterion (specification) or not. In contrast, a student's overall assignment may be evaluated using a 2-level, 3-level, or 4-level system. Other variations for both specification and assignment evaluations exist, but these are more complex (Gargac, 2022a, 2022b; Gestwicki, 2021; Toledo & Dubas, 2017; Mirsky, 2018; Henriksen et al., 2020; Cosoroaba, 2020; Donato & Marsh, 2023; Rupakheti et al., 2018). It is not uncommon for different assignments in a course to have different evaluation outcome types as needed. Descriptions and examples of binary specifications rubrics with 2-level, 3-level, and 4-level assignment outcomes follow.

The majority of publications included in this review use assignments with binary specifications and 2-level assignment outcomes. In this approach, a student's work either does or does not meet an individual specification. The instructor sets a threshold, i.e., a number of specifications that must be met for the overall assignment to earn credit. Instructors may also set some specifications as "required" so that the assignment does not earn credit if those "required" specifications are not met, regardless of how many others are met. Wording for the assignment outcomes varies, but some of the common phrases that have been used are pass/fail, satisfactory/unsatisfactory, satisfactory/needs revision, meets specifications/does not meet specifications, accept/revise, or complete/incomplete (Howitz et al., 2021; Blodgett, 2017; Jones, 2020; Santucci & Golas, 2023; Blackstone & Oldmixon, 2019; Lillard & Taggart, 2022).

A closely related, but less common approach is to use binary specifications with 3-level assignment outcomes, in which two different thresholds are set. Depending on how the instructor sets up their specifications grading system, students may only earn credit for an assignment if they meet the higher of the two thresholds (Williams, 2018), or they may earn credit as long as they meet at least one of the two thresholds (McKnelly et al., 2021). McKnelly et al.'s Writing for Chemists course, described previously, provides an example of the latter case. Grades are bundled based not only on the number of times a student is able to pass a large writing assignment but also at what passing threshold (a high pass, low pass, or needs revision). For example, students must pass all four writing assignments to earn an A or a B, but for the B grade, students must earn low passes or better on all four assignments, whereas for the A grade, students need to earn at least three high passes and may earn only one low pass.

Binary specifications rubrics may also be combined with 4-level assignment outcomes based on the EMRF rubric that allows for two levels of passing work (E: excellent, M: meets expectations) and two levels of work that does not pass (R: needs revision, F: fragmentary) (Stutzman & Race, 2004). An advantage of this 4-level system is the ability to differentiate the quality of student work within the passing and not passing categories. In most cases, the F designation has been replaced with N (not assessable) to overcome students' association of the F with failing, resulting in the more frequently used acronym, EMRN (Talbert (n.d.). In Mendez's sophomore-level thermodynamics course, described previously, each quiz is assessed using the EMRN rubric (Mendez, 2018a). While students need to pass 11 quizzes to earn a C, 13 for a B, and 15 for an A, the threshold at which students pass the quiz also matters. For example, to earn a C, none of the quizzes need to be assessed as excellent (E), but to earn a B, 7 of the 13 quizzes need to be assessed as excellent (E), and to earn an A, 11 of the 15 quizzes must be assessed as excellent (E).

3.2.3. Token Systems

Of the 90 publications selected for this review, 46 specifically commented on the inclusion of a token system, and three commented specifically on choosing not to include a token system. The remaining 41 publications do not include any specific information about token systems. Instructors chose to provide tokens in their specifications grading systems by providing a set number of tokens to each student at the beginning of the course (seven examples) (Fernandez et al., 2020; Joshi, 2023; Dabney & VanDerWoude, 2023; Lillard & Taggart, 2022; Vitale & Concepción, 2021; Johnson, 2023; Tamés, 2021), by providing opportunities for students to earn tokens throughout the term of a course (fourteen examples) (Howitz et al., 2021; Joseph et al., 2023; Evensen, 2022; Hunter et al., 2022; Ludwigsen, 2017; Helmke, 2019; Martin, 2019; Prasad, 2020; McKnelly et al., 2023; Kelz et al., 2023; Rojas & Quan, 2023; Saluga et al., 2023; Howitz et al., 2023; Cerkez, 2024), or by combining both aforementioned approaches (seventeen examples) (Fernandez et al., 2020; Dennen & Bagdy, 2020; Moster & Zingales, 2024; Suresh, 2023; Closser et al., 2024; Johanesen et al., 2024; Tsoi et al., 2019; Carlisle, 2020; McKnelly et al., 2021; Cosoroaba, 2020; Blackstone & Oldmixon, 2019; Williams, 2018; Brown & Kennedy, 2022; Fierke, 2024; Yang & Korsnack, 2024; Mio, 2024; Copp, 2024). In the earn-only or combination approaches, students were given opportunities to earn tokens by a variety of means, such as completing metacognitive reflection assignments or completing low-stakes course activities such as readings and homework.

In all token systems described, students were given the option to use tokens for additional attempts at assignments (either through revision or attempting a new version of an assignment such as a quiz), for flexibility on assignment deadlines, or both. In six examples, students could also use tokens to earn back credit on a low-stakes assignment that they missed originally or to replace attendance credit lost for missing class when attendance was required (Dennen & Bagdy, 2020; McKnelly et al., 2021; Vitale & Concepción, 2021; McKnelly et al., 2023; Mio, 2024; Kiefer & Earle, 2023). In all cases, the choice of how to use tokens was left with the students, although instructors did provide encouragement to use tokens as needed and reminders of how to do so.

Few authors provided details on how many tokens they chose to make available, the rates at which students used tokens or the specific methods by which instructors tracked tokens. However, the few examples provided do give useful guidance for instructors considering implementing a token system in their courses. Hunter et al. suggest determining the number of tokens to provide by counting the number of high-stakes assignments and adding one (Hunter et al., 2022). Vitale and Concepción suggest a similar approach—providing tokens that correspond to the number of high-stakes assignments plus or minus one (Vitale & Concepción, 2021).

Based on the limited examples provided, students do not appear to run out of tokens. In a first-year engineering course, 69% of students used at least one token in the course, and a grand total of 41% of available tokens in the course were used by the end (Fernandez et al., 2020). On any given assignment in this engineering course, at least one student chose to use a token and the number of students who used a token on an assignment increased as the term progressed. Kelz et al. found that only four of ninety-nine students used all of their tokens (Kelz et al., 2023). Dennen and Bagdy indicated that few students used all of the provided tokens, and those who did chose specifically to do so (Dennen & Bagdy, 2020). With one exception in which students can trade tokens for candy (Williams, 2018), authors did not report rewarding students for tokens left over at the end of a course. Despite this lack of reward, two publications reported that students display token-hoarding behaviors (Tsoi et al., 2019; Martin, 2019). One instance of "gaming the system" was described, in which a team of students working on a group project opted to submit work that did not display a good-faith effort at completion and then replace the missing credit for that work with a token (Fernandez et al., 2020). The authors stated that this "gaming the system" was performed to gain more time to work on another aspect of the course.

Current learning management system (LMS) options are not designed to support specifications grading and do not provide tools to support a token system. However, a placeholder assignment (Martin, 2019) or ungraded quizzes in the course LMS (Dennen & Bagdy, 2020) can be used to track students' tokens. Alternatively, token usage can be tracked using an online form with a spreadsheet alone (Fernandez et al., 2020; Blackstone & Oldmixon, 2019; Kelz et al., 2023) or in combination with an LMS placeholder assignment (Howitz et al., 2021; McKnelly et al., 2023).

3.3. Themes and Opportunities

Many publications on the design and implementation of specifications grading discuss outcomes, impacts, and lessons learned. A review of the current literature reveals there are common themes that emerge from these discussions. These themes include instructor time investment, comparisons of academic performance, and student reactions to the grading scheme.

3.3.1. Instructor Commentary on Time Investment

One concern about adopting specifications grading is an instructor time commitment. While some publications do indicate an increase in the time spent grading (Hofmeister et al., 2022b; Joseph et al., 2023; Closser et al., 2024; Ring, 2017; Henriksen et al., 2020; Tamés, 2021; Hunter et al., 2022; Martin, 2019; Rojas & Quan, 2023; Cerkez, 2024; Kiefer & Earle, 2023; Hollinsed, 2018), the majority of adopters of specifications grading reported that they spent about the same amount of time (Elkins, 2016; Earl, 2021; Moster & Zingales, 2024; Suresh, 2023; Carlisle, 2020; McKnelly et al., 2021; Blackstone & Oldmixon, 2019; Lillard & Taggart, 2022; Vitale & Concepción, 2021; McKnelly et al., 2023; Lovell, 2018; Largent, 2024), or less time (Howitz et al., 2021; Jones, 2020; Walden, 2022; Dupree et al., 2024; Toledo & Dubas, 2017; Mirsky, 2018; Williams, 2018; Kelz et al., 2023; Fierke, 2024; Copp, 2024; Tuson &

Hickey, 2022; Sanft et al., 2021; Mendez, 2018b, 2023; Pascal et al., 2020; Trachsler et al., 2023; Noell et al., 2023; Tuson & Hickey, 2023) grading under the new system than they spent using a traditional points-based grading system (Table 3). The most common reason cited for saving time was the removal of partial credit (Earl, 2021; Fernandez et al., 2020; Walden, 2022; Dabney & VanDerWoude, 2023; Toledo & Dubas, 2017; Williams, 2018; Rojas & Quan, 2023; Copp, 2024; Largent, 2024; Tuson & Hickey, 2022; Tuson & Hickey, 2023). Without partial credit, the cognitive load associated with choosing the appropriate allocation of points is reduced. Instructors who did not experience time savings noted that the time they spent on grading was allocated differently; they could spend more time providing feedback because they spent less time deciding how much partial credit to award. Aside from removing partial credit, reducing the number of LOs being assessed (Mendez, 2018a; Toledo & Dubas, 2017) and removing assignments from the course that did not map to the LOs (Jones, 2020) were cited as contributing to the time saved when grading. Although not explicitly stated, it can be inferred that time may have also been saved from students choosing to not submit work for assignments that were not required for their target letter grade (Jones, 2020; Blackstone & Oldmixon, 2019).

Table 3. Number of publications that reported the time commitment required to design and implement a specifications grading system in a course.

Reported Time Commitments for Designing and Implementing Specifications Grading	Number of References
Increased time grading	12
No change in time grading	12
Decreased time grading	18
Removing partial credit saved time when grading regardless of changes in overall time spent grading	11
Significant time investment to design specifications grading system	20
Significant time investment to generate multiple versions of quizzes or exams	11

Although most publications on specifications grading indicate no change in, or a reduction in, grading time, several comments on the time investment required to design and build the system. The consensus is that the time investment associated with designing the specifications grading system is substantial (Elkins, 2016; Earl, 2021; Blodgett, 2017; Jones, 2020; Joseph et al., 2023; Dupree et al., 2024; Moster & Zingales, 2024; Shields et al., 2019; Henriksen et al., 2020; Blackstone & Oldmixon, 2019; Vitale & Concepción, 2021; McKnelly et al., 2023; Rojas & Quan, 2023; Mio, 2024; Copp, 2024; Trachsler et al., 2023; Noell et al., 2023; Mendez, 2023; LeHew, 2019; Anzovino et al., 2023). Time-consuming aspects of designing a specifications grading system that were mentioned include developing quiz questions and building question banks (Suresh, 2023; Henriksen et al., 2020; Rojas & Quan, 2023; Copp, 2024; Kiefer & Earle, 2023; Lovell, 2018; Tuson & Hickey, 2022; Noell et al., 2023; Mendez, 2023; Anzovino et al., 2023; Mattfeld, 2023), constructing new rubrics and assignment guidelines (Elkins, 2016; Earl, 2021; Dupree et al., 2024), and crafting grade bundles in which assignments are appropriately mapped to final letter grades (Elkins, 2016; Dupree et al., 2024; Rojas & Quan, 2023; Anzovino et al., 2023).

Additionally, some publications indicated that extra time was required for the specifications grading implementation that did not involve grading student work. One report of additional time focused on the need to meet with students outside of class because time for additional assignment attempts was not built into the course schedule (Lovell, 2018). It was also reported that additional time was required to address student concerns around the removal of partial credit, to normalize the experience of needing to try again, and to achieve student buy-in for an unfamiliar grading system (LeHew, 2019). While the majority of publications do not describe the time commitment required to implement a token system, only two described the time commitment as onerous (Hofmeister et al., 2022b; Martin, 2019) and five describe it as minimal (Howitz et al., 2021; Gestwicki, 2021; McKnelly et al., 2021; Kelz et al., 2023; Copp, 2024). Management of the token system through the course LMS was cited as a way to keep the workload manageable (Copp, 2024).

3.3.2. Impacts on Academic Performance

One of the ways improvement in student learning is evaluated following the implementation of novel pedagogy is to examine whether the distribution of final letter grades in the course changed or not. For the majority of those who have published their implementations of specifications grading, the distribution of final letter grades either shifts in the direction of a larger percentage of collective A and B grades (Howitz et al., 2021; Hofmeister et al., 2022b; Jones, 2020; Moster & Zingales, 2024; Houseknecht & Bates, 2020; Evensen, 2022; Carlisle, 2020; Toledo & Dubas, 2017; Lillard & Taggart, 2022; Vitale & Concepción, 2021; Helmke, 2019; McKnelly et al., 2023; Kelz et al., 2023; Rojas & Quan, 2023; Mio, 2024; Kiefer & Earle, 2023; Hollinsed, 2018; Lovell, 2018; Noell et al., 2023; Anzovino et al., 2023; Katzman et al., 2021; Bunnell et al., 2023) or remains the same (Earl, 2021; Dennen & Bagdy, 2020; Closser et al., 2024; Blackstone & Oldmixon, 2019; Copp, 2024; Largent, 2024; Ahlberg, 2021) (Table 4). The reported impacts on course drop, fail, and withdrawal (DFW) rates are mixed. Some adopters of specifications grading report a decrease in their overall DFW rate (Moster & Zingales, 2024; Evensen, 2022; Toledo & Dubas, 2017; McKnelly et al., 2023; Kiefer & Earle, 2023; Anzovino et al., 2023), while others report no change (Earl, 2021; Evensen, 2022; Hollinsed, 2018; Bunnell et al., 2023) or an increase (Gargac, 2022a; Lillard & Taggart, 2022; Noell et al., 2023; Anzovino et al., 2023). Noell et al. observed an increase in the DFW rate, noting that the Ds and Fs decreased, but the Ws increased (Noell et al., 2023). The increase in withdrawals was attributed to the transparency of the specifications grading system, allowing students to know if they would be able to pass the course before the withdrawal deadline.

Table 4. Number of publications that reference how implementing a specifications grading system impacted course letter grades and DFW rates.

Specifications Grading Impact on Course Letter Grades and DFW Rates	Number of References
Larger percentage of collective A and B grades	22
No change in course letter grade distribution	7
Course GPA decreased	4
DFW rate increased	4
DFW rate decreased	6
No change in DFW rate	4

Only four publications did not report a positive shift in the final letter grade distribution (Gargac, 2022a, 2022b; Jones, 2020; Wasniewski et al., 2021). One of these publications specifically reported that more students chose to complete the assignments associated with a B grade than with an A grade, which likely accounted for the decrease in the course GPA (Jones, 2020). No commentary was provided in the other three publications to explain what may account for the decrease in the final letter grade distribution (Gargac, 2022a, 2022b; Wasniewski et al., 2021).

It is reasonable to expect that the positive shift in final letter grade distributions observed by many adopters of specifications grading could be the result of an improvement in the quality of student work. However, the majority of publications only address comparisons of student work anecdotally between versions of the course taught using a points-based grading system and the version taught using a specifications grading system. In these cases, the perception reported was that the quality of student work submitted under the specifications grading system was similar (Quintana & Quintana, 2020; Walden, 2022; Donato & Marsh, 2023; Blackstone & Oldmixon, 2019; Noell et al., 2023) or better (Earl, 2021; Fernandez et al., 2020; Dennen & Bagdy, 2020; Jones, 2020; Joshi, 2023; Wasniewski et al., 2021; McKnelly et al., 2021; Cosoroaba, 2020; Kelz et al., 2023; Fierke, 2024; Sanft et al., 2021; Mendez, 2018b) than work submitted in points-based systems.

Five publications included a quantitative comparison of the quality of work between the two grading systems. Two of these publications compared final exams between the points-based and specifications grading versions of their courses. Ring compared student work by grading final exams from their previous points-based version of their course with the same rubric as the specifications grading course (i.e., without partial credit) and found that the students from the specifications grading version of the course passed questions at a higher rate (Ring, 2017). However, Ring did not comment on whether there was a change in the final letter grade distribution between the two courses. In Martin's courses, the same final exam composed of 60 multiple-choice questions was used for both a points-based and a specifications grading version of their general chemistry course (Martin, 2019). The average score in four previous course iterations that used points-based grading ranged between 30 and 38 (50–63%), whereas in the three terms following the adoption of a specifications grading system, the averages ranged between 38 and 42 (63–70%). Changes to the final letter grade distribution by adopting specifications grading were not discussed.

In addition to the two studies analyzing student performance on final exams, three other publications reported quantitative comparisons of the quality of work unrelated to final exams in different grading versions of the courses. Helmke set the passing threshold for assignments and exams in their specifications grading course to a B (85%) and compared the number of students in the specifications grading course that met the threshold to the number of students that earned at least an 85% on assignments and exams in the points-based course (Helmke, 2019). Helmke found that fewer students from the specifications grading course passed on the first attempts of both the homework assignments and unit exams; however, more students passed on the final attempts for homework assignments, and a similar number passed on the final attempts for the unit exams. Under Helmke's specifications grading system, the percentage of students earning As remained the same, but the percentage of Bs increased and the percentage of Cs decreased. Katzman took a different approach by administering an end-of-semester survey to students who completed a version of their course under a points-based grading system and to students who completed a version of the course under a specifications grading system (Katzman et al., 2021). Students who completed the course using a specifications grading system earned higher median and maximum scores on the content assignment questions in the survey than those who completed the course using a points-based grading system. Katzman observed no difference in the percentage of students who did pass and who did not pass the course between the two grading systems. However, amongst the passing students, the percentage of As increased, and the percentage of Bs and Cs decreased. Amongst the non-passing students, the percentage of Ds and Ws increased, but the percentage of Fs decreased. Finally, McKnelly et al. compared student performance on a laboratory report (McKnelly et al., 2023). They graded the same number of reports for a single assignment from the points-based grading and specifications grading versions of the course using the original points-based rubric and found no statistically significant difference in the average scores on the reports. Despite the similar scores, McKnelly et al. found an increase in the percentage of students earning A and B grades, but a decrease in students earning C and D grades. One explanation they offer for the discrepancy is the limited sample size and comparison of quality on a single assignment. They propose that assignment quality may have increased for some assignments and not others.

One publication did not directly compare the quality of work between the two grading systems but did compare the impact of taking a course graded using a points-based system versus a specifications grading system on the passing rate in that course and the one following it. In both a general and an organic chemistry course sequence, Anzovino et al. found no statistically significant difference amongst students continuing to the second course in the sequence regardless of whether they took the first course in each sequence under a points-based system or a specifications grading system (Anzovino et al., 2023). Additionally, they found no statistically significant differences in the passing rates (C or better) in the second course in each sequence, regardless of whether students took the first course in each sequence graded using a points-based system or a specifications grading system. This may suggest that the quality of work submitted by students under a specifications grading system is not any lower than that submitted by students under a points-based grading system.

Finally, one publication directly addresses potential grade inflation in specifications grading—a common concern. In a specification grading graduate-level organic chemistry course using a modules configuration (six essential and seven general), Moster and Zingales observed higher final letter grades than in the prior points-based iteration (Moster & Zingales, 2024). Students earned more As and fewer Bs and Cs, with 10% more students passing the course overall. Students who earned As earned higher average scores on a 50-question final exam than students who earned Bs. Moster and Zingales connect this higher exam performance to the fact that students who earned A grades were required to complete an additional essential module and four general modules as compared to students who earned B grades. Because students had to complete 54% of the course content (seven weekly modules; five essential and two general) before taking the final exam, and the passing threshold on each module was set to 80%, it follows that students passing the course should be scoring at least 43% ($0.54 \times 0.80 = 0.43$) on the final exam. Because 90% of the students met this criterion, with many earning scores far greater than 43%, Moster and Zingales argue that the final exam outcomes indicate that students have met the course SLOs—and retained the knowledge—at levels that are clearly commensurate with the letter grades they earned.

3.3.3. Student Reactions to Specifications Grading

Within the theme of student reactions to specifications grading, three primary subthemes emerged as follows: focus on learning, the transparency of the grading system, and stress and anxiety. These subthemes are derived from 1) anecdotal data from instructor observations, conversations with students, and quotes pulled from final evaluations and 2) solicited feedback from students through surveys that were not validated instruments. These subthemes are tied to goals of specifications grading according to Nilson: to shift student focus from points to learning, to provide a transparent grading scheme, and to reduce student stress and anxiety.

Focus on Learning. Many educators who have implemented specifications grading observed a shift from students focusing on accumulating points to focusing on learning the course material (Howitz et al., 2021; Joseph et al., 2023; Santucci & Golas, 2023; Wasniewski et al., 2021; Suresh, 2023; Johanesen et al., 2024; Mirsky, 2018; Lillard & Taggart, 2022; Johnson, 2023; Tamés, 2021; Hunter et al., 2022; Ludwigsen, 2017; Helmke, 2019; McKnelly et al., 2023; Kelz et al., 2023; Cerkez, 2024; Mio, 2024; Copp, 2024; Mattfeld, 2023). For example, Hunter et al. state, "Students clearly grasp why their assignment earned the grade that it did, and I no longer field questions about point allocation. Instead, student questions after an exam focus on concepts and improvement, and the process of revision absolutely improves the students' understanding of the material (Hunter et al., 2022)". Similarly, Henriksen et al. and Lovell observed that students took greater advantage of office hours to ask questions about concepts covered earlier in the course because the incentive for students to seek and apply feedback was greater under the specifications grading system where revision is encouraged (Henriksen et al., 2020; Lovell, 2018). Part of the reduction in students' focus on points, aside from revision opportunities, was attributed to the greater transparency and clarity of the specifications grading system, specifically the expectations detailed in assignment rubrics (Fernandez et al., 2020; Vitale & Concepción, 2021; Pascal et al., 2020).

Despite increased student focus on learning, some students opposed the removal of partial credit. In four publications, authors noted that students in their specifications grading courses felt the passing thresholds were set too high (Gratwick et al., 2020; Kinnear et al., 2022; McKnelly et al., 2021; McKnelly et al., 2023). Kinnear et al., McKnelly et al., and Williams observed student frustration when they just missed meeting the passing threshold (Kinnear et al., 2022; Williams, 2018; McKnelly et al., 2023). In these cases, students felt the effort they put into the work they submitted was not being taken into account, as no partial credit was given for assignment submissions that did not pass. Reports from Toledo and Dubas, from Rojas, and from Blodgett also acknowledge student displeasure with not earning partial credit on work that did not meet the passing threshold (Blodgett, 2017; Toledo & Dubas, 2017; Rojas & Quan, 2023).

Grading Transparency. Many adopters of specifications grading report that students find the grading system to be more transparent than points-based courses. For some students, this transparency refers to the clarity of assignment expectations and the associated rubrics (Henriksen et al., 2020; Cosoroaba, 2020; Vitale & Concepción, 2021; Lovell, 2018). For others, transparency refers to knowing what assignments need to be completed to earn their desired final letter grade and what their standing is in the course at any point throughout the term (Howitz et al., 2021; Dennen & Bagdy, 2020; Hofmeister et al., 2022b; Kelz et al., 2023; Copp, 2024; Pascal et al., 2020). In two publications, authors commented that their students appreciated that the transparency of the specifications grading system allowed them to choose what assignments they needed to complete to achieve the grade they aimed for (LeHew, 2019; Mirth, 2017). Jones indicated that their students found that the choices afforded by the specifications grading system led to a greater ability to balance coursework in other classes and their life obligations (Jones, 2020). In three cases, authors noted that their students felt that the transparency allowed them to direct their efforts in the course because the grading system provided guidance about which topics they understood well and which they needed to continue working on (Toledo & Dubas, 2017; Hunter et al., 2022; Rojas & Quan, 2023).

Although many instructors reported that students found the grading system more transparent, there were several who reported student confusion about the new grading system (Earl, 2021; Martin et al., 2021; Hofmeister et al., 2022b; Joseph et al., 2023; Moster & Zingales, 2024; Santucci & Golas, 2023; Evensen, 2022; Closser et al., 2024; Biers, 2022; Lillard & Taggart, 2022; Hunter et al., 2022; Mio, 2024). Some students, at least initially, felt the new assignment expectations and rubrics were unclear (Howitz et al., 2021; Kelz et al., 2023; Cerkez, 2024). Others were unaware of how to determine their final letter grade according to the grade bundles (Howitz et al., 2021; Joseph et al., 2023; Toledo & Dubas, 2017). Some instructors adopted strategies in an attempt to achieve student buy-in and minimize confusion with this novel grading system. Early in the term, some instructors included in class or video explanations of the grading system (Elkins, 2016; Howitz et al., 2021; Hofmeister et al., 2022b; Jones, 2020; Quintana & Quintana, 2020; Dupree et al., 2024; Dabney & VanDerWoude, 2023; Tsoi et al., 2019; Carlisle, 2020; Blackstone & Oldmixon, 2019; Hunter et al., 2022; Prasad, 2020; McKnelly et al., 2023; Cerkez, 2024; Fierke, 2024; Mio, 2024; Pascal et al., 2020; LeHew, 2019), some compared specifications grading to traditional grading to help students understand the purpose behind why it was adopted (Martin et al., 2021; Prasad, 2020), and some developed activities or tools for students to learn how to track and determine their final letter grade (Howitz et al., 2021; Shields et al., 2019; Carlisle, 2020; Toledo & Dubas, 2017; Lillard & Taggart, 2022; McKnelly et al., 2023; Pascal et al., 2020; Reck, 2022). Other instructors provided regular reminders throughout the term (Jones, 2020; Dupree et al., 2024; Santucci & Golas, 2023; Evensen, 2022; Dabney & VanDerWoude, 2023; Carlisle, 2020; Johnson, 2023; McKnelly et al., 2023; Fierke, 2024; Reck, 2022) and/or implemented metacognitive reflections or goal-setting exercises (Howitz et al., 2021; Biers, 2022; Johnson, 2023; Prasad, 2020; McKnelly et al., 2023; Yang & Korsnack, 2024; Copp, 2024), which prompted students to think through how they could earn their desired final letter grade under the specifications grading system. The extent to which these interventions mitigated or resolved student confusion varied.

Stress and Anxiety. Student perceptions of the impact of specifications grading on their stress and anxiety compared to points-based grading vary. Twenty-four articles report a reduction in student stress and anxiety and comment on possible reasons for this change. In some cases this reduction is attributed to multiple revision opportunities being provided, lowering the stakes on assignments (Earl, 2021; Fernandez et al., 2020; Walden, 2022; Joseph et al., 2023; Moster & Zingales, 2024; Evensen, 2022; Closser et al., 2024; Carlisle, 2020; Biers, 2022; Henriksen et al., 2020; Hunter et al., 2022; McKnelly et al., 2023; Cerkez, 2024; Fierke, 2024; Tuson & Hickey, 2022; Mattfeld, 2023; Ahlberg, 2021; Pope et al., 2020). In other instances, students reported feeling that the grading was lower stakes, which allowed them to put more of their focus into learning the course material. However, they did not elaborate upon what made the grading feel lower stakes (Jones, 2020; Helmke, 2019). Trachsler et al. reported that students felt reduced stress and anxiety stemmed from the flexibility afforded to them by being able to choose which assignments to complete to earn their desired grade (Trachsler et al., 2023). Additionally, students perceived greater transparency in assignment expectations and how to earn their desired final letter grade, which accounted for a reduction in stress and anxiety (Howitz et al., 2021; Earl, 2021; Quintana & Quintana, 2020; Biers, 2022; Cerkez, 2024; LeHew, 2019; Pope et al., 2020).

In contrast to a decrease in stress and anxiety, other students reported an increase with the specifications grading system. Many of the concerns raised by students were related to the pass/fail nature of specifications grading, and some students equated not passing an assignment on the first attempt to failing despite the fact that they were able to try again without a grade penalty. Some students felt that the expectations to pass an assignment were too high (Joshi, 2023; Shields et al., 2019; Kinnear et al., 2022; Carlisle, 2020; Helmke, 2019; Kiefer & Earle, 2023; Lovell, 2018; Pope et al., 2020). Many students did not like the absence of partial credit and felt that they should receive some credit for the work they submitted rather than not earning any credit for an assignment that did not meet the passing threshold (Howitz et al., 2021; Shields et al., 2019; Prasad, 2020; McKnelly et al., 2023; Copp, 2024; Noell et al., 2023). Students expressed stress and anxiety about how one mistake could make the difference between passing and not passing an assignment and consequently affect their grades in the course (McKnelly et al., 2021; Cosoroaba, 2020; McKnelly et al., 2023; LeHew, 2019). Elkins acknowledges that not earning credit for the work they submitted can be a "harsh reality for students who are used to earning at least partial credit no matter how low their level of work (Elkins, 2016)". Other instructors identified similar perceptions in which students appeared to believe that needing to revise an assignment was equivalent to "receiving an F (Johanesen et al., 2024)," "a demoralizing '0' rather than an opportunity for improvement" (Rojas & Quan, 2023), or would "result in the assignment somehow being worth 'less' (LeHew, 2019)." Blodgett shared a student quote that highlights this demoralization: "Making a great deal of effort and then getting a 0 for an assignment made me wonder why bother at all (Blodgett, 2017)". Noell et al. and Hunter et al. found that students are not accustomed to having opportunities to revise work, so it is useful in these situations to give reminders that not being assessed as passing on the first attempt is not a sign of failure and that revision opportunities are built into the course (Hunter et al., 2022; Noell et al., 2023).

Other sources of stress and anxiety unrelated to the pass/fail nature of specifications grading included the frequency of testing (Ring, 2017; Noell et al., 2023), the responsibility placed on the students to self-track their grades (Toledo & Dubas, 2017), and the tendency for increased procrastination by some students because they knew they had retake opportunities (Henriksen et al., 2020; Tuson & Hickey, 2022). Students who expressed having good grades going into a final assignment did not like that poor performance on a final assignment could negatively impact their final letter grade (Carlisle, 2020; Prasad, 2020; McKnelly et al., 2023).

4. Discussion

As the number of publications describing the design and implementation of specifications grading has grown, so has the breadth of disciplines from which instructors have reported their experiences. Despite specifications grading systems having four common features—grade bundles, rubrics with specifications and defined passing thresholds, opportunities to revise and resubmit work, and a token system—the details of each design and implementation vary substantially from instructor to instructor. Design choices likely impacted the outcomes of each implementation. Following, we discuss the themes that arose from instructors' design choices and how they influenced implementation outcomes in specifications grading systems, focusing particularly on impacts on time, academic performance, and student reactions to the grading system.

4.1. Instructor Commentary on Time Investment

In general, switching to specifications grading required a substantial time commitment prior to the start of a course to design the system. One of the time-consuming aspects of the design was the construction of the grade bundles. As instructors worked to construct grade bundles, they were incentivized to reevaluate their course LOs and the assignments that mapped to those LOs. This backward course design in tandem with a specifications grading approach where the focus is put on the student achievement of LOs rather than the accumulation of points to earn a letter grade suggests that a grade a student earns in a course under specifications grading may more accurately reflect the knowledge and skills they have gained by completing the course (Earl, 2021; Blackstone & Oldmixon, 2019; McKnelly et al., 2023; Mio, 2024).

Although the time required to design a specifications grading system was substantial, most instructors who implemented specifications grading in their courses reported that the time required to implement the course was about the same or less than the amount of time spent implementing prior points-based courses. Instructors cite that the time they would have spent on allocating points on assignments in the points-based course was instead spent on giving students more feedback on assignments (Earl, 2021; Moster & Zingales, 2024; Suresh, 2023; Largent, 2024; Mendez, 2018b). Previous research has shown that when both scores and feedback are provided to students, students will not necessarily pay attention to the feedback, in part due to the spatial separation of scores and feedback within the LMS (Winstone et al., 2021). However, when students have opportunities to revise their work, they show higher cognitive engagement with the feedback (Fredricks et al., 2016; Espasa et al., 2022). The opportunity for students to revise or retake assignments under a specifications grading system may incentivize students to more carefully review the feedback that was given rather than only checking their scores. In this way, the additional feedback instructors leave on student assignments in courses using specifications grading may be more likely to lead to students producing higher quality work and achieving higher final letter grades compared to points-based courses.

Although the majority of instructors found the time to implement specifications grading was not greater than for a points-based course, a small number reported spending more time implementing the new system. In analyzing how these instructors structured their grading systems, it became apparent that students were given many opportunities and/or unlimited time to retake or resubmit assignments (Henriksen et al., 2020; Martin, 2019; Prasad, 2020; Hollinsed, 2018). Allowing many reattempts can result in students submitting low-quality work on early attempts because the incentive to prepare appropriately for assignments is not present (Prasad, 2020; Hollinsed, 2018; Mendez, 2018b). If deadlines for revising and resubmitting or retaking assignments are not provided, it may increase the likelihood of procrastination (Closser et al., 2024; Prasad, 2020). This can result in a high instructor workload at the end of the term when many assignments are submitted, and it may also result in students being unable to earn the grade they are aiming for because there is insufficient time to complete all of the unfinished assignment revisions (Closser et al., 2024; Prasad, 2020). It appears that token systems were not used in many of these cases, which could have mitigated some of the challenges with implementing opportunities for students to try again. Instructors who proactively included limited revision opportunities, deadlines for revising and resubmitting or retaking assignments, and/or a token system in their design and implementation of specifications grading reported manageable time investments (Howitz et al., 2021; Dabney & VanDerWoude, 2023; McKnelly et al., 2023; Kelz et al., 2023; Copp, 2024; Mendez, 2023).

Several publications included discussions of how the specifications grading system evolved over time, specifically focusing on changes that were made to streamline the course and reduce the instructor's time commitment. Changes that saved faculty time included a reduction in the number of LOs assessed and/or the number of assignments necessary to earn a letter grade (Earl, 2021; Dupree et al., 2024; Mendez, 2018a; Toledo & Dubas, 2017), consolidation of course content coverage (Closser et al., 2024), limiting the number of retakes per assignment (Closser et al., 2024; Prasad, 2020; Hollinsed, 2018), and/or imposing deadlines on the time available to students to revise or retake an assignment (Closser et al., 2024; Fierke, 2024). This suggests that while the adoption of specifications grading may require a significant time investment initially, the time savings become more apparent during implementation, especially if appropriate revisions are made to the system over multiple iterations of the course. While not necessarily implemented, other instructors suggest that changing assignments to be auto-graded (Hofmeister et al., 2022b), developing methods to auto-generate test questions (Mattfeld, 2023), and building on the work of others, such as through a faculty learning community (Anzovino et al., 2023), could also save time in future implementations of their specifications grading course.

4.2. Impacts on Academic Performance

The majority of publications report a shift in the final letter grade distribution toward more A and B grades after adopting specifications grading. In general, it would be expected that as the final letter grade distribution shifts toward more A and B grades, the quality of student work would show substantial improvement. However, this potential correlation has not been studied systematically. Most reported comparisons of the quality of student work between points-based and specifications grading systems are anecdotal. This lack of detailed comparisons of student work is unsurprising due to the drastic changes that are made to a course when converting from a points-based grading system to a specifications grading system. Assignments are typically modified to better align with the LOs. The removal of partial credit makes conducting a rigorous comparison of student work particularly challenging because the scales are not the same between the two grading systems (e.g., an assignment task may be graded out of five in a points-based grading system versus being graded using the binary pass/no pass rating in a specifications grading system).

Without rigorous studies indicating that the quality of student work is higher under specifications grading, an argument could be made that the positive shift in final letter grades is a form of grade inflation due to reduced rigor. However, this argument fails to take into account that students are held to a high standard to earn credit on assignments when no partial credit is included. For example, an instructor may set a threshold of B-level work or above for a student to earn credit on an assignment, as recommended by Nilson. Holding students to these high standards often results in students not passing all assignments on the first attempt. Without the ability to accumulate partial credit, students are held accountable for revisiting the material and taking advantage of revision or resubmission opportunities to demonstrate an improved understanding of LOs. Providing opportunities for students to try again to meet the high standard set by the instructor supports learning and results in grade elevation, not grade inflation (Streifer & Palmer, 2021). This sentiment is supported by work conducted by Moster and Zingales in their organic chemistry course described earlier in this review (Moster & Zingales, 2024). They suggest the higher final letter grades students earn under the specifications grading system are consistent with the achievement, of course, LOs based on their performance on a final exam. While their work presents some initial evidence that students' quality of work correlates with final letter grades, additional studies are necessary to quell the concerns of those who are hesitant to adopt alternative grading systems for fear of propagating grade inflation.

4.3. Student Reactions to Specifications Grading

Overall, student sentiment toward specifications grading systems tends to be positive and appears to improve as a course progresses. Negative reactions to specifications grading may be attributed to student unfamiliarity with the grading scheme—because many students are habituated to traditional points-based grading—and/or to specific design choices within courses. Reports of student confusion about the specifications grading system and final letter grade determination at the start of a course were common. The confusion was likely due to students encountering an alternative grading approach for the first time and having to adjust to changes in grading norms. Instructors observed improvements in students' understanding of the specifications grading system when the number of LOs was reduced, course content coverage was consolidated, grade bundles were streamlined, and/or when a tool was provided for students to track progress toward and determine their final letter grade. Many publications reported attempts to increase students' buy-in to the specifications grading system. However, these interventions seemed to have varied success in increasing students' comfort with the system. Future studies could investigate the efficacy of buy-in interventions, evaluate how student buy-in changes during a term, and explore whether student buy-in improves after multiple offerings of the course under the specifications grading system.

In addition to reports of student confusion, cases of increased stress and anxiety were also reported. This increase in stress and anxiety, much like the confusion experienced by students taking a course taught using a specifications grading system for the first time, may stem primarily from students' habituation to traditional points-based grading systems. For example, students did not like that they could not earn any credit when they submitted work that was not of passing quality. In these cases, students wanted credit for effort expended and/or felt that the passing threshold was set too high. This desire for partial credit and/or credit awarded for effort regardless of performance may reflect a performance-focused goal orientation adopted by these students. The fact that these complaints come from only a small percentage of students in a given course with a specifications grading system may indicate that the majority of the students adopted a mastery-focused goal orientation (Elliot & McGregor, 2001; Elsinger & Lewis, 2020).

Design choices made by instructors when creating their specifications grading systems may also contribute to some of the increased student stress and anxiety about grades. For example, setting the passing threshold of an assignment to 100% would send the message to students that perfection is required and could easily lead to students feeling that they will never pass an assignment or the course. Additionally, providing unlimited attempts or unlimited time with no clear due dates may remove the structure that students need to stay on track and avoid procrastination. These observations suggest that setting an appropriate passing threshold for an assignment and providing structure around assignment attempts are critical design choices. Sorensen-Unruh argued that while self-regulated learning theory (SRL) is often cited as the underlying theory of alternative grading systems (captured under the umbrella term "ungrading"), SRL is inherently deficit-framed (Sorensen-Unruh, 2024). Instructors may unintentionally be adopting a deficit framing when designing their specifications grading systems, and this underlying assumption may be leading to design choices that undermine the goals of alternative grading.

While negative sentiments arose from students experiencing specifications grading systems, often for the first time, most students reacted positively to the structural components of these grading systems that are not often, if ever, present in traditional points-based grading systems. For example, having opportunities to revise and resubmit reduced student stress and anxiety by reducing the stakes of any single assignment. Additionally, the combination of not being expected to complete or pass every assignment and the pass/fail approach of specifications grading systems encouraged students to focus on the learning process rather than the accumulation of points to earn their final letter grade. From a theoretical perspective, the shift away from a focus on points may indicate a change in students' achievement goal orientation toward a mastery focus (Elliot & McGregor, 2001; Elsinger & Lewis, 2020). From a practical perspective, this shift away from a focus on points is logical because allocating partial credit on assignments creates many borderline cases, which in turn incentivizes students to fight for every point. Without partial credit, the number of borderline cases drops drastically as each specification is frequently graded on a binary scale (Joshi, 2023). The structural components discussed above, combined with students having the ability to choose the type of assignment to complete, provided students with greater agency and, in most cases, a reduction in stress and anxiety. This increased agency also allowed students to better balance their obligations outside of academics and to better allocate time across all of their academic courses.

The features of specifications grading systems discussed above may not only be improving student academic performance but may also be developing students' professional and social identities in their chosen disciplines. Observed gains in students' math identity (Villalobos et al., 2024) suggest that specifications grading systems can provide an avenue to improve inclusion and equity in STEM courses, and similar studies could be expanded to non-STEM disciplines. The implications of specifications grading systems to help address equity concerns in higher education provide ripe opportunities for further research to investigate these avenues.

4.4. Limitations

In addition to those discussed above, there are other limitations present in our analysis. Less formal methods of dissemination, such as theses, posters, workshops, etc., were excluded, so there are some design and/or implementation considerations about how instructors are implementing specifications grading systems in their courses that we did not discuss. Additionally, there are instructors who are contingent faculty or who are not in positions where publication is necessary for their academic institution, so there are unpublished examples of specifications grading systems that we cannot analyze and include in this review. This limitation may be especially present in implementations of specifications grading in community colleges and other two-year institutions. As is discussed above, most of the instructors publishing in the specifications grading system literature have written descriptive papers, which do not necessarily include controlled studies, so conclusions about the effects of specifications grading systems on students are limited. Finally, in a finding that echoes work by Hackerson et al., we note that connections to theoretical frameworks were rarely discussed in the specifications grading literature, which limited our ability to address how theory is being incorporated into the design and implementation of specifications grading systems (Hackerson et al., 2024).

5. Conclusions

Alternative grading systems have emerged to address challenges associated with points-based grading systems. These challenges include but are not limited to, a misalignment between the LOs students achieve and their final course grade, a student's focus on achieving a grade due to the influence of external rather than internal motivators, and a rise in student stress, anxiety, and mental health challenges. Specifications grading systems have been gaining popularity as one of the alternative systems that have emerged in the hopes of mitigating some of the challenges associated with points-based grading systems. The findings in this review indicate that publications on the design and implementation of specifications grading continue to grow, with the majority of publications occurring in STEM fields (chemistry, engineering, computer science, and mathematics) and in small courses ($\leq$35 students), although examples across all course sizes and a wide variety of disciplines do exist (research question 1). Grade bundles, rubrics with specifications and defined passing thresholds, opportunities to revise and resubmit work, and a token system are the four structural components of specifications grading systems and are found to be incorporated in a variety of ways (research question 2).

Analysis of the implementations of the specifications grading systems and emergent themes of instructor commentary on time investment, academic performance, and student reactions to specifications grading reveal (1) important considerations for the design and development of specifications grading systems and (2) a roadmap for education researchers to collaborate with practitioners to study outcomes and impacts of specifications grading in higher education (research question 3). Many of the current publications on specifications grading are descriptive studies on individual designs and implementations. While additional descriptive publications will be valuable resources for practitioners seeking inspiration for their courses, especially from disciplines or types of courses that are not yet represented, enough evidence now exists to point to the fact that we need future research to include systematic studies on topics such as impacts on student academic performance, implications of design choices, effective practices for securing student buy-in, and potential for increased equity of specifications grading systems. Hackerson et al. have called for interdisciplinary studies on alternative grading systems across STEM fields (Hackerson et al., 2024), and we broaden that call to include the need for interdisciplinary studies on specifications grading systems across both STEM and non-STEM disciplines.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci15010083/s1, Table S1: Full Corpus of Papers Analyzed.

Author Contributions: Methodology, conceptualization, research, writing, review, and editing. W.J.H., K.J.M. and R.D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The full corpus of publications analyzed for this review is available in the supplementary files.

Acknowledgments: The authors thank Clarissa Sorensen-Unruh and Katy Hosbein for helpful conversations on review structures.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Ahlberg, L. (2021). Organic chemistry core competencies: Helping students engage using specifications. In *Engaging students in organic chemistry* (Vol. 1378, pp. 25–36). American Chemical Society.

Anzovino, M. E., Behmke, D., Villanueva, O., & Woodbridge, C. M. (2023). Specifications grading and COVID. In *ACS Symposium Series* (pp. 89–105). American Chemical Society.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.

Beare, P. G. (1986, October 9–11). *The contract--an individualized approach to competency-based learning and evaluation*. Thinking across the Disciplines. Annual Conference of the International Society for Individualized Instruction, Atlanta, GA, USA.

Biers, K. (2022). Decolonial and feminist course design and assessment in the first-year french curriculum. In S. Bouamer, & L. Bourdeau (Eds.), *Diversity and decolonization in French studies: New approaches to teaching* (pp. 245–257). Springer International Publishing.

Blackstone, B., & Oldmixon, E. (2019). Specifications grading in political science. *Journal of Political Science Education*, *15*, 191–205. [CrossRef]

Blodgett, B. J. (2017). Grading matters in theological education. *Teaching Theology & Religion*, *20*, 314–326. [CrossRef]

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*, n2.

Bowen, R. S., & Cooper, M. M. (2022). Grading on a curve as a systemic issue of equity in chemistry education. *Journal of Chemical Education*, *99*, 185–194. [CrossRef]

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, *86*, 803–848. [CrossRef]

Brown, R., & Kennedy, T. (2022, June 26–29). *Work in progress: Differentiated learning in a specifications grading framework*. 2022 ASEE Annual Conference & Exposition, ASEE Conferences, Minneapolis, MN, USA.

Bunnell, B., LeBourgeois, L., Doble, J., Gute, B., & Wainman, J. W. (2023). Specifications-based grading facilitates student–instructor interactions in a flipped-format general chemistry II course. *Journal of Chemical Education*, *100*, 4318–4326. [CrossRef]

Carberry, A. R., Siniawski, M. T., & Dionisio, J. D. N. (2012, October 3–6). *Standards-based grading: Preliminary studies to quantify changes in affective and cognitive student behaviors*. 2012 Frontiers in Education Conference Proceedings (pp. 1–5), Seattle, WA, USA.

Carlisle, S. (2020). Simple specifications grading. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *30*, 926–951. [CrossRef]

Cerkez, E. B. (2024). Application of specifications grading to an analytical chemistry lab. *Journal of Chemical Education*, *101*, 4268–4275. [CrossRef]

Clark, D., & Talbert, R. (2023). *Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education*. Taylor & Francis.

Closser, K. D., Hawker, M. J., & Muchalski, H. (2024). Quantized grading: An ab initio approach to using specifications grading in physical chemistry. *Journal of Chemical Education*, *101*, 474–482. [CrossRef]

Copp, D. A. (2024, June 23–26). *Specifications grading in an undergraduate engineering dynamics course*. 2024 ASEE Annual Conference & Exposition, Portland, OR, USA.

Cosoroaba, E. (2020, June 22–26). *Helping students write it right: Instilling good report-writing habits in a linear circuit lab course*. 2020 ASEE Virtual Annual Conference Content Access Proceedings, ASEE Conferences, Online.

Dabney, B. W., & VanDerWoude, C. J. (2023). Implementing a specifications grading system in a nursing course. *Nurse Educator*, *48*, 187–191. [CrossRef] [PubMed]

Dennen, V., & Bagdy, L. (2020, April 14–16). *Shifting to specifications grading: Two design cases*. TCC 2020 Conference Papers, TCC Worldwide Online Conference, Online.

Donato, J. J., & Marsh, T. C. (2023). Specifications grading is an effective approach to teaching biochemistry. *Journal of Microbiology & Biology Education*, *24*, e00236-22. [CrossRef]

Dupree, L. H., Augustine, J. M., & Miller, S. W. (2024). How did we get here? Evolution of specifications grading in a required skills-based course series. *Currents in Pharmacy Teaching and Learning*, *16*, 102124. [CrossRef]

Durm, M. W. (1993). An A Is Not an A Is Not an A: A History of Grading. *The Educational Forum*, *57*, 294–297. [CrossRef]

Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development* (1st ed.). Psychology Press.

Earl, D. (2021). Two years of specifications grading in philosophy. *Teaching Philosophy 45*, 23–64. [CrossRef]

Elkins, D. M. (2016). Grading to learn: An analysis of the importance and application of specifications grading in a communication course. *Kentucky Journal of Communication*, *35*, 2016.

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 Achievement goal framework. *Journal of Personality and Social Psychology*, *80*, 501–519. [CrossRef] [PubMed]

Elsinger, J., & Lewis, D. (2020). Applying a standards-based grading framework across lower level mathematics courses. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *30*, 885–907. [CrossRef]

Espasa, A., Guasch, T., Mayordomo, R. M., & Martinez-Melo, M. (2022). Prior experience with online feedback: Its influence on students' engagement. *Distance Education*, *43*, 444–465. [CrossRef]

Evensen, H. (2022, June 26–29). *Specifications grading in general physics and engineering physics courses*. 2022 ASEE Annual Conference & Exposition, ASEE Conferences, Minneapolis, MN, USA.

Eyler, J. R. (2024). *Failing our future*. Johns Hopkins University Press.

Fernandez, T., Martin, K., Mangum, R., & Bell-Huff, C. (2020, June 22–26). *Whose grade is it anyway?: Transitioning engineering courses to an evidence-based specifications grading system*. 2020 ASEE Virtual Annual Conference Content Access Proceedings, ASEE Conferences, Online.

Fierke, K. L. (2024). Utilizing specification grading in sport management classes. *Sport Management Education Journal*, *18*, 151–156. [CrossRef]

Fredricks, J. A., Wang, M. -T., Schall Linn, J., Hofkens, T. L., Sung, H., Parr, A., & Allerton, J. (2016). Using qualitative methods to develop a survey measure of math and science engagement. *Learning and Instruction*, *43*, 5–15. [CrossRef]

Gargac, J. (2022a, June 6–8). *Incorporating industry-based metrics into a specifications grading system for capstone assessment*. Capstone Design Conference 2022, Capstone Design Community, Dallas, TX, USA.

Gargac, J. (2022b, June 26–29). *Manufacturing Engagement: Improving Student Learning through Modifying Content Delivery and Assessment*. 2022 ASEE Annual Conference & Exposition, ASEE Conferences, Minneapolis, MN, USA.

Gay, J. L., & Poproski, R. (2023). Specifications grading for graduate students: Motivated, less stressed, and achieving in public health. *Pedagogy in Health Promotion*, *9*, 237–243. [CrossRef]

Gestwicki, P. (2021). Godot engine and checklist-based specifications: Revising a game programming class for asynchronous online teaching. *Journal of Computing Sciences in Colleges*, *37*, 30–40.

Gratwick, R., Kinnear, G., & Wood, A. K. (2020). An online course promoting wider access to university mathematics. In R. Marks (Ed.), *Proceedings of the British society for research into learning mathematics* (1st ed., Vol. 40). Available online: https://www.pure.ed.ac.uk/ws/portalfiles/portal/158931832/BSRLM_CP_40_1_04.pdf (accessed on 25 October 2023).

Hackerson, E. L., Slominski, T., Johnson, N., Buncher, J. B., Ismael, S., Singelmann, L., Leontyev, A., Knopps, A. G., McDarby, A., Nguyen, J. J., Condry, D. L. J., Nyachwaya, J. M., Wissman, K. T., Falkner, W., Grieger, K., Montplaisir, L., & Hodgson, A. (2024). Alternative grading practices in undergraduate STEM education: A scoping review. *Disciplinary and Interdisciplinary Science Education Research*, *6*, 15. [CrossRef]

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and shiny app for producing prisma 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell Systematic Reviews*, *18*, e1230. [CrossRef] [PubMed]

Hammoudi Halat, D., Soltani, A., Dalli, R., Alsarraj, L., & Malki, A. (2023). Understanding and fostering mental health and well-being among university faculty: A narrative review. *Journal of Clinical Medicine*, *12*, 4425. [CrossRef]

Harrington, B., Galal, A., Nalluri, R., Nasiha, F., & Vadarevu, A. (2024, March 20–23). *Specifications and contract grading in computer science education*. 55th ACM Technical Symposium on Computer Science Education V. 1, Portland, OR, USA.

Harsy, A., Carlson, C., & Klamerus, L. (2021). An analysis of the impact of mastery-based testing in mathematics courses. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *31*, 1071–1088. [CrossRef]

Hassencahl, F. (1979). Contract Grading in the Classroom. *Improving College and University Teaching*, *27*, 30–33. [CrossRef]

Helmke, B. P. (2019, June 15–19). *Specifications grading in an upper-level BME elective course*. ASEE Annual Conference & Exposition, ASEE Conferences, Tampa, FL, USA.

Henriksen, M., Kotas, J., & Wentworth, M. (2020). Specifications-based grading reduces anxiety for students of ordinary differential equations. *CODEE Journal*, *13*, 1. [CrossRef]

Hofmeister, E. H., Fogelberg, K., Conner, B. J., & Gibbons, P. (2022a). Specifications grading for veterinary medicine. *Journal of Veterinary Medical Education 50*, 15–18. [CrossRef] [PubMed]

Hofmeister, E. H., Fogelberg, K., Conner, B. J., & Gibbons, P. (2022b). Specifications grading in a cardiovascular systems course: Student and course coordinator perspectives on the impacts on student achievement. *Journal of Veterinary Medical Education*, *50*, 172–182. [CrossRef] [PubMed]

Hollinsed, W. C. (2018). Applying innovations in teaching to general chemistry. In *Increasing Retention of Under-Represented Students in STEM through Affective and Cognitive Interventions* (Vol. 1301, pp. 145–152). American Chemical Society.

Houseknecht, J. B., & Bates, L. K. (2020). Transition to Remote Instruction Using Hybrid Just-in-Time Teaching, Collaborative Learning, and Specifications Grading for Organic Chemistry 2. *Journal of Chemical Education*, *97*, 3230–3234. [CrossRef]

Howitz, W. J., Frey, T., Saluga, S. J., Nguyen, M., Denaro, K., & Edwards, K. D. (2023). A specifications-graded, sports drink-themed general chemistry laboratory course using an argument-driven inquiry approach. *Journal of Chemical Education*, *100*, 672–680. [CrossRef]

Howitz, W. J., McKnelly, K. J., & Link, R. D. (2021). Developing and implementing a specifications grading system in an organic chemistry laboratory course. *Journal of Chemical Education*, *98*, 385–394. [CrossRef]

Hunter, R. A., Pompano, R. R., & Tuchler, M. F. (2022). Alternative Assessment of Active Learning. In *Active Learning in the Analytical Chemistry Curriculum* (Vol. 1409, pp. 269–295). American Chemical Society.

Johanesen, K. E., Claiborne, L. L., Falk, E. S., Hubbard, K. P., Kohfeld, K. E., Nadin, E. S., & Schmidt, A. H. (2024). Common-Sense teaching for the 2020s: Ungrading in response to COVID-19 and beyond. *Journal of Geoscience Education*, *72*, 422–437. [CrossRef]

Johnson, S. (2023). An alternative grading approach in undergraduate anatomy & physiology courses. *HAPS Educator*, *27*, 99–108. [CrossRef]

Jones, P. A. (2020). Implementing specifications grading in MPA courses: A potential strategy for better work-life balance. *Journal of Public Affairs Education*, *26*, 531–547. [CrossRef]

Joseph, M. L., Miller, S. W., Diec, S., & Augustine, J. M. (2023). Successes and challenges in implementing specifications grading in skills-based laboratory courses: Experiences at two colleges of pharmacy. *Currents in Pharmacy Teaching and Learning*, *15*, 186–193. [CrossRef]

Joshi, A. (2023, October 18–21). *Assessing the impact of specifications grading on a data visualization course*. 2023 IEEE Frontiers in Education Conference (FIE) (Vol. 25, pp. 1–6), College Station, TX, USA.

Katzman, S. D., Hurst-Kennedy, J., Barrera, A., Talley, J., Javazon, E., Diaz, M., & Anzovino, M. E. (2021). The effect of specifications grading on students' learning and attitudes in an undergraduate-level cell biology course. *Journal of Microbiology & Biology Education*, *22*, e00200–e00221. [CrossRef]

Kelz, J. I., Uribe, J. L., Rasekh, M. F., Takahashi, G. R., Gibson, W. S., Link, R. D., McKnelly, K. J., & Martin, R. W. (2023). Implementation of specifications grading in an upper-division chemical biology lecture course. *The Biophysicist*, *4*, 11–29. [CrossRef]

Kiefer, S. F., & Earle, A. J. (2023, June 25–28). *Work in progress: Specifications grading in a system modeling course*. 2023 ASEE Annual Conference & Exposition, ASEE Conferences, Baltimore, MD, USA.

Kinnear, G., Wood, A. K., & Gratwick, R. (2022). Designing and evaluating an online course to support transition to university mathematics. *International Journal of Mathematical Education in Science and Technology*, *53*, 11–34. [CrossRef]

Knight, M., & Cooper, R. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment, and student behavior. *NASSP Bulletin*, *103*, 65–92. [CrossRef]

Kulik, C. -L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, *60*, 265. [CrossRef]

Largent, D. L. (2024). Specifications grading meets computer science: Implementations in skill-and project-based courses. In A. D. Tomlin, & C. M. Nowik (Eds.), *Effective alternative assessment practices in higher education*. Research, Theory, and Practice Within Academic Affairs, Information Age Publishing.

LeHew, M. (2019). Specifications grading: Experimenting with a new form of student assessment. *Journal for Academic Excellence*, *6*, 12–15.

Leslie, P., & Lundblom, E. (2020). Specifications grading: What it is, and lessons learned. *Seminars in Speech and Language*, *41*, 298–309. [CrossRef] [PubMed]

Lewis, D. (2022). Impacts of standards-based grading on students' mindset and test anxiety. *JoSoTL*, *22*. [CrossRef]

Lillard, A. S., & Taggart, J. (2022). Reimagining assessment in a large lecture: An alternative approach inspired by Thomas Jefferson and Maria Montessori. *College Teaching*, *72*, 168–180. [CrossRef]

Lovell, M. D. (2018, June 24–27). *Board 44: Defining and assessing competencies in an undergraduate reinforced concrete design course*. 2018 ASEE Annual Conference & Exposition, ASEE Conferences, Salt Lake City, UT, USA.

Ludwigsen, D. (2017). Effective and efficient grading: Tips for assessing student learning. *Proceedings of Meetings on Acoustics*, *30*, 025001. [CrossRef]

Martin, K. M., Fernandez, T. M., & Mangum, R. (2021, July 26–29). *Work in progress: Examining engineering seniors students' perception of justice and fairness of grading practices*. 2021 ASEE Virtual Annual Conference Content Access, ASEE Conferences, Virtual.

Martin, L. J. (2019). Introducing components of specifications grading to a general chemistry I course. In *Enhancing Retention in Introductory Chemistry Courses: Teaching Practices and Assessments* (Vol. 1330, pp. 105–119). American Chemical Society.

Mattfeld, R. S. (2023). Improving student motivation through an alternative grading system. *Journal of Computing Sciences in Colleges*, *39*, 86–95.

McKnelly, K. J., Howitz, W. J., Thane, T. A., & Link, R. D. (2023). Specifications grading at scale: Improved letter grades and grading-related interactions in a course with over 1000 Students. *Journal of Chemical Education*, *100*, 3179–3193. [CrossRef]

McKnelly, K. J., Morris, M. A., & Mang, S. A. (2021). Redesigning a "writing for chemists" course using specifications grading. *Journal of Chemical Education*, *98*, 1201–1207. [CrossRef]

Mendez, J. (2018a, June 24–27). *Standards-based specifications grading in a hybrid course*. 2018 ASEE Annual Conference & Exposition, Salt Lake City, UT, USA.

Mendez, J. (2018b, June 24–27). *Standards-based specifications grading in thermodynamics*. 2018 ASEE Annual Conference & Exposition, Salt Lake City, UT, USA.

Mendez, J. (2019, November 9–15). *Development of a hybrid heat and mass transfer course*. ASME 2018 International Mechanical Engineering Congress and Exposition, Pittsburgh, PA, USA.

Mendez, J. (2023, April 1–2). *Specifications grading in undergraduate fluid mechanics*. 2023 ASEE IL-IN Section Conference, Edwardsville, IL, USA.

Mendez, J. (2024, April 13). *Work in progress: Specifications grading in mechanical engineering design courses*. ASEE IL-IN Section Conference, Chicago, IL, USA.

Mio, M. J. (2024). Alternative grading strategies in organic chemistry: A journey. *Frontiers in Education*, *9*, 1400058. [CrossRef]

Mirsky, G. M. (2018). Effectiveness of specifications grading in teaching technical writing to computer science students. *Journal of Computing Sciences in Colleges*, *34*, 104–110.

Mirth, J. A. (2017, June 25–28). *Applying lean thinking to the structure and delivery of a kinematics course*. 2017 ASEE Annual Conference & Exposition, Columbus, OH, USA.

Moster, C. A., & Zingales, S. K. (2024). Use of specifications-based grading in an online, asynchronous graduate organic chemistry course. *Frontiers in Education*, *9*, 1379216. [CrossRef]

Nilson, L. B., & Stanny, C. J. (2014). *Specifications grading: Restoring rigor, motivating students, and saving faculty time* (Reprint ed.). Stylus Publishing.

Noell, S. L., Rios Buza, M., Roth, E. B., Young, J. L., & Drummond, M. J. (2023). A bridge to specifications grading in second semester general chemistry. *Journal of Chemical Education*, *100*, 2159–2165. [CrossRef]

Pascal, J., Vogel, T. J., & Wagstrom, K. (2020, June 22–26). *Grading by competency and specifications: Giving better feedback and saving time*. 2020 ASEE Virtual Annual Conference Content Access, ASEE Conferences, Online.

Pope, L., Parker, H. B., & Ultsch, S. (2020). Assessment of specifications grading in an undergraduate dietetics course. *Journal of Nutrition Education and Behavior*, *52*, 439–446. [CrossRef] [PubMed]

Prasad, P. V. (2020). Using revision and specifications grading to develop students' mathematical habits of mind. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *30*, 908–925. [CrossRef]

Quintana, R., & Quintana, C. (2020). When classroom interactions have to go online: The move to specifications grading in a project-based design course. *Information and Learning Sciences*, *121*, 525–532. [CrossRef]

Reck, R. M. (2022, October 8–11). *Adopting alternative grading in an upper-level laboratory course in bioengineering*. 2022 IEEE Frontiers in Education Conference (FIE) (pp. 1–5), Uppsala, Sweden.

Ring, J. (2017). ConfChem conference on select 2016 BCCE presentations: Specifications grading in the flipped organic classroom. *Journal of Chemical Education*, *94*, 2005–2006. [CrossRef]

Rojas, C., & Quan, G. M. (2023, June 25–28). *Mastery grading in a software engineering course*. 2023 ASEE Annual Conference & Exposition, ASEE Conferences, Baltimore, MD, USA.

Rupakheti, C. R., Hays, M., Mohan, S., Chenoweth, S., & Stouder, A. (2018). On a pursuit for perfecting an undergraduate requirements engineering course. *Journal of Systems and Software*, *144*, 366–381. [CrossRef]

Saluga, S. J., Burns, A. M., Li, Y., Nguyen, M. M., & Edwards, K. D. (2023). A specifications-graded, spice-themed, general chemistry laboratory course using an argument-driven inquiry approach. *Journal of Chemical Education*, *100*, 3903–3915. [CrossRef]

Sanft, K. R., Drawert, B., & Whitley, A. (2021). Modified specifications grading in computer science: Preliminary assessment and experience across five undergraduate courses. *Journal of Computing Sciences in Colleges*, *36*, 34–46.

Santucci, A., & Golas, J. C. (2023). Grading for learning at the university of Rhode Island. *NEFDC Exchange*, *37*, 7–12.

Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE—Life Sciences Education*, *13*, 159–166. [CrossRef] [PubMed]

Seymour, E. (1997). *Talking about leaving: Why undergraduates leave the sciences* (1st ed.). Westview Press.

Shields, K., Denlinger, K., & Webb, M. (2019). Not missing the point (s): Applying specifications grading to credit-bearing information literacy classes. In M. N. Mallon, L. Hays, C. Bradley, R. Huisman, & J. Belanger (Eds.), *The grounded instruction librarian: Participating in the scholarship of teaching and learning*. Association of College and Research Libraries.

Smeding, A., Darnon, C., Souchal, C., Toczek-Capelle, M. -C., & Butera, F. (2013). Reducing the socio-economic status achievement gap at university by promoting mastery-oriented assessment. *PLoS ONE*, *8*, e71678. [CrossRef] [PubMed]

Sorensen-Unruh, C. (2024). The Ungrading learning theory we have is not the ungrading learning theory we need. *CBE Life Sciences Education*, *23*, es6. [CrossRef]

Streifer, A. C., & Palmer, M. S. (2021). Is specifications grading right for me?: A readiness assessment to help instructors decide. *College Teaching*, *71*, 244–251. [CrossRef]

Stutzman, R. Y., & Race, K. H. (2004). EMRF: Everyday rubric grading. *The Mathematics Teacher*, *97*, 34–39. [CrossRef]

Suresh, M. A. (2023, June 25–28). *Challenges and experiences in implementing a specifications grading system in an upper-division undergraduate computer networks course*. 2023 ASEE Annual Conference & Exposition, ASEE Conferences, Baltimore, MD, USA.

Talbert, R. (n.d.). *The EMRN rubric*. Available online: https://rtalbert.org/emrn/ (accessed on 25 October 2023).

Tamés, D. (2021, July 27–30). *Implementation and evaluation of specifications grading in a video production course* [Paper presentation]. 75th University Film and Video Association (UFVA) Conference: Empowering Spaces, Online.

Taylor, H. (1980). *Contract grading*. EPIC Clearinghouse on Tests, Measurement, and Evaluation.

Toledo, S., & Dubas, J. M. (2017). A Learner-centered grading method focused on reaching proficiency with course learning outcomes. *Journal of Chemical Education*, *94*, 1043–1050. [CrossRef]

Trachsler, T. A., Morris, E., & Mahoney, T. Q. (2023). Specifications grading in the sport management classroom: Breakdown of the System and Reflections Upon Implementation With Relation to Outcomes Assessment. *Sport Management Education Journal*, *17*, 121–128. [CrossRef]

Tsoi, M. Y., Anzovino, M. E., Lin Erickson, A. H., Forringer, E. R., Henary, E., Lively, A., Morton, M. S., Perell-Gerson, K., Perrine, S., Villanueva, O., Whitney, M., & Woodbridge, C. M. (2019). Variations in implementation of specifications grading in STEM courses. *Georgia Journal of Science*, *77*, 10.

Tuson, E., & Hickey, T. (2022). Mastery learning and specs grading in discrete math. In *Proceedings of the 27th ACM conference on innovation and technology in computer science education* (Vol. 1, pp. 19–25). Association for Computing Machinery.

Tuson, E., & Hickey, T. (2023, March 15–18). *Mastery learning with specs grading for programming courses*. 54th ACM technical symposium on computer science education V. 1 (pp. 1049–1054), Toronto, ON, Canada.

Villalobos, C., Garza, G., Setayesh, S., Fernandez, L., Balogh, A., Serbin, K. S., & Huber, T. J. (2024). Our never-ending pathway to innovate calculus 1: Course coordination and active learning to specifications grading and growth mindset. Problems, Resources, and Issues in Mathematics Undergraduate Studies, *Advance online publication*, 1–17. [CrossRef]

Vitale, S. E., & Concepción, D. W. (2021). Improving student learning with aspects of specifications grading. *Teaching Philosophy*, *44*, 29–57. [CrossRef]

Walden, P. R. (2022). Student motivation, anxiety and pass/fail grading: A SoTL project. *Teaching and Learning in Communication Sciences & Disorders*, *6*, 13. [CrossRef]

Wasniewski, E., Munro, T., & Tandon, T. (2021). Exploring Online Specifications Grading: An Undergraduate Course Case Study. In T. J. Bastiaens (Ed.), *Proceedings of the EdMedia + innovate learning* (pp. 179–184). Association for the Advancement of Computing in Education (AACE).

Williams, K. (2018). Specifications-based grading in an introduction to proofs course. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *28*, 128–142. [CrossRef]

Winstone, N., Bourne, J., Medland, E., Niculescu, I., & Rees, R. (2021). "Check the grade, log out": Students' engagement with feedback in learning management systems. *Assessment & Evaluation in Higher Education*, *46*, 631–643. [CrossRef]

Yang, M. A., & Korsnack, K. (2024). Pairing a bioinformatics-focused course-based undergraduate research experience with specifications grading in an introductory biology classroom. *Biology Methods and Protocols*, *9*, bpae013. [CrossRef] [PubMed]

Yik, B. J., Machost, H., Streifer, A. C., Palmer, M. S., Morkowchuk, L., & Stains, M. (2024). Students' perceptions of specifications grading: Development and evaluation of the perceptions of grading schemes (pgs) instrument. *Journal of Chemical Education*, *101*, 3723–3738. [CrossRef]

education sciences

MDPI

Review

Using Simulations and Screencasts in Online Preclass Activities to Support Student Building of Mental Models

Deborah G. Herrington [1,*,†] and Ryan D. Sweeder [2,†]

1 Department of Chemistry, Grand Valley State University, Allendale, MI 49401, USA
2 Lyman Briggs College, Michigan State University, East Lansing, MI 48825, USA; sweeder@msu.edu
* Correspondence: herringd@gvsu.edu
† These authors contributed equally to this work.

Abstract: As online learning and flipped classes become more important in chemistry instruction, the development of learning materials that can be used to support students' independent learning of conceptual chemistry content is critical. This paper summarizes the key findings from an eight-year investigation of effective practices for using simulations in preclass introductions to core chemistry concepts with a focus on supporting students' development of particulate-level models. Student learning gains for six core chemistry concepts were compared for students' independent use of a simulation using scaffolded instructions versus students' viewing a screencast of instructors modeling the use of the simulation to answer a series of questions. Though both approaches resulted in student learning gains and provided a solid foundation for subsequent instruction, the screencast approach provided additional benefits. These included avoiding potential simulation limitations and the ability to add instructional content to support student learning. Additionally, studying many iterations of assignments for several different topics yielded an assignment design framework that provides guidelines for instructors looking to create or use simulation-based preclass activities in the classroom to support student learning.

Keywords: first-year undergraduate chemistry; online learning; simulations; screencasts

1. Introduction

The development and use of models in science and science instruction are critical in developing conceptual understanding and explanations of the natural world. Such models are particularly important for things that are not readily observable or may be too big, too small, too slow, or too fast to observe [1]. In chemistry, the development of scientifically accurate models for the motion and interactions of particles (atoms, molecules, ions, etc.) is critical to being able to predict and explain physical phenomena [2]. Given that it is not possible to directly observe atoms, ions, or molecules due to their small size, there has been increased development and use of conceptual simulations to support student development of particle-level mental models [3–5]. Conceptual simulations, unlike operational simulations that focus on teaching procedural skills, focus on the learning process, and are specifically designed to aid learners in constructing mental models [6]. In chemistry, conceptual simulations allow students to visualize the unobservable and explore how changing variables affect systems. The use of such simulations in science education has been shown to enhance science instruction and improve students' conceptual understanding and satisfaction [7,8]. When such simulations are used in classroom settings, instructors can provide students with guidance and direct their attention to the more salient aspects of the simulation. However, using simulations in class limits the amount of time students have to process information. As our students enter our classrooms with very different backgrounds, it is reasonable to expect that students may require different amounts of time interacting with a simulation to make sense of it and develop the desired mental models. The benefit of many of these online simulations is that students can access them outside of class where and when they want to and spend as much time as they need engaging with the material [9]. This is one rationale for "flipped" classrooms where students engage in some kind of activity on their own before class (readings, watching videos, etc.) and then spend class time working through problems [10].

The past decade has seen the increased use of the flipped classroom model for STEM instruction in response to the call for more active learning in the classroom [11]. This model frees up in-class time for instructors to support students' active engagement with tasks designed to reinforce or build upon the foundation of knowledge that students developed in the preclass activity [12]. The flipped classroom has been well studied with several meta-analyses supporting the conclusion that flipped instruction significantly improves student cognitive learning over traditional lecture methods across multiple subject domains [13–16]. Specifically in chemistry, most of the studies on flipped classrooms at the post-secondary level have focused on introductory and organic chemistry [12]. Further, a systematic review of several of these studies noted wide variations in flipped models and identified key features of flipped models resulting in significant student gains including the following: accountability for completing pre and in-class activities; employing responsive mini-lectures to address common student difficulties; and follow-up post-class practice [12]. Though the ability to increase the amount of active learning is often cited as the key reason for implementing flipped learning, a recent commentary highlights the importance of well-designed preclass activities for promoting conceptual learning [17].

1.1. Effective Preclass Activities

Though preclass activities can take many forms, including textbook reading, lecture videos are the most common forms of preclass activities used in chemistry [12,17]. Studies have shown that textbook readings or video lectures can promote learning gains [18,19], though one study that compared different formats for preclass activities found that video lectures had an advantage over textbook readings [19]. More importantly, studies have shown that students' active engagement in preclass activities is critical to student learning [12,13,16,17]. This is consistent with the ICAP Framework—Interactive, Constructive, Active, and Passive—which predicts that as students become more engaged with learning materials, their learning will increase [20].

Indeed, studies of flipped models that use video lectures with no methods for assuring students are actively engaged with the materials show small or no student gains [12,13,16]. The ability to embed required questions within a video can help to support student self-assessment and provide accountability for the completion of preclass activities, which may explain why students completing such activities were found to significantly outperform students completing textbook readings [21]. This idea is supported by a recent systematic review of flipped models that identified accountability for the completion of preclass activities as a feature of all flipped models that showed significant increases in course GPA [12]. The importance of the preclass activity is further established by a study comparing a traditional lecture class, with a class that incorporated collaborative in-class active learning activities, and a flipped model with collaborative in-class active learning activities. The results showed that after controlling for prior knowledge, the students in the flipped model significantly outperformed students in the other two classes and that there was no significant difference between the traditional class and the class that just incorporated active learning [18].

Another possible explanation for the positive effects of preclass activities is that they can reduce in-class cognitive load by introducing students to core concepts before discussing them in class. This was the basis of a study by Seery and Donnelly, where they looked at the impacts of preclass videos designed to introduce key concepts in forthcoming lectures that included a quiz to allow students to self-assess their understanding of the material [22]. They found that students with no prior chemistry knowledge improved to the point where there was no difference between their performance on the post-module test and final exam when compared to students who entered the class with prior knowledge. In designing such videos, it was also important for them to consider Meyer's theory of multimedia learning, which posits that students learn better when the material is explicitly chunked to help organization, only the content relevant to the learning objective is included, the material is presented both visually and verbally, and the means for student self-assessment are included [23].

1.2. Using Simulations in Preclass Activities

In chemistry, the use of conceptual simulations in preclass activities has value in helping students start to develop particle-level mental models for core concepts. This can allow students' active engagement with the material [24] and the opportunity for discovery learning [25–27]. One set of conceptual simulations that has been particularly well studied is the PhET simulations [3,27]. A recent review of 31 experimental or quasi-experimental studies noted that the use of PhET simulations was found to enhance student conceptual understanding of physics when included in inquiry-based activities, virtual labs, problem-based learning activities, and scaffolded learning activities [28]. Further, the use of PhET simulations in introductory concept development activities as part of a learning-cycle-based flipped model in a chemistry course resulted in significant learning gains from pre- to post-assessment and on the final exam for all concepts studied [8]. In this study, authors specifically chose to incorporate these concept development activities in class as opposed to using them as preclass activities as they noted that students typically only have a listening or note-taking role in online learning activities. Another study in physics incorporated PhET simulations into online preclass activities with the goal of enhancing reading assignments [29]. They found that though students found the preclass reading assignments with PhET simulation activities more enjoyable than the reading assignments alone, there was no significant difference in learning gains between the reading-only group and the group that had the reading assignment enhanced with the PhET simulation activity. The authors suggested that one reason for this may have been that did not provide enough scaffolding for students to productively explore the PhET simulation. This is supported by the fact that students reported spending on average about 10 min interacting with the PhET simulations and about 35 min reading.

The need for adequate scaffolding to effectively support student interactions with the simulations is an important consideration, especially as many of the chemistry simulations are quite complex. Though studies of PhET simulations conducted with students in individual interviews or small group settings found that engaged exploration with the simulation only occurred when students were provided with minimal or no guidance [25,30], a subsequent study specifically focused on student use of the PhET simulations individually outside of the classroom found that a higher level of scaffolding was required to meaningfully engage students with the simulations [31].

Even with appropriate scaffolding, especially if the simulations are quite complex, novice learners may not identify important features [14] or may misinterpret some features of the simulation [7,32,33]. Alternatively, a screencast, a screen capture video where an instructor leads students through a simulation, can address some of the disadvantages associated with students' independent simulation use. With a screencast, instructors can direct students' attention to key features, thus reducing their cognitive load, and can clarify or prevent misinterpretations of the simulation content or features [34]. However, as students are watching a video as opposed to directly interacting with the simulation, there is the potential for more passive engagement with the content when using screencasts [32].

The goal of the ChemSims Project [35] has been to develop structured support for students' use of simulations outside of the classroom to help them develop particle-level understanding of core chemistry concepts. To identify effective practices for the development of supporting materials, we compared students' independent, scaffolded use of simulations with student viewing of screencasts where an instructor demonstrated the use of the simulation to explore a concept. In both cases, students were expected to answer questions while completing the assignment to assess how well they were able to identify key elements of the simulation and apply these observations to the desired core chemistry concept. All activities were designed to be a preclass, initial introduction to a topic that could then provide a foundational experience upon which an instructor could build further understanding. Over the course of the project, materials to support student learning for six different foundational chemistry concepts were developed and studied [36–40]. This paper describes the key takeaways from the ChemSims project regarding the effective practices for using simulations and screencasts to support student learning outside of the classroom and how they can be used in the development of effective preclass activities. Such materials can be used to both support students' learning outside of the classroom and to support flipped classroom models that provide more in-class time for engaged student learning.

2. ChemSims Assignment and Study Design

In addition to important considerations such as ensuring active engagement of students and providing adequate scaffolding discussed above, the clarity and quality of instruction can certainly play a role in the efficacy of the screencast and simulations assignments. To ensure alignment of learning objectives, assessments, and instruction and provide a mechanism for evaluating areas of student confusion, the screencast and simulation assignments for each of six chemistry concepts were developed and tested using backward design [41] combined with an iterative revision approach (Figure 1), which has been described in more depth elsewhere [36]. Each activity was designed as an introduction to a topic with a specific focus of helping students develop particle-level mental models. Students were given a pretest aligned with the identified learning objectives to establish prior knowledge before engaging with either the simulation or screencast assignment; each assignment was designed to be equivalent in content and focus. Matched posttest questions were embedded in the assignment to allow for a pre–post comparison measure of student learning gains (Figure 2). Qualitative analysis of student written responses to pre and posttest questions as well as questions embedded within the assignment provided data regarding student challenges.

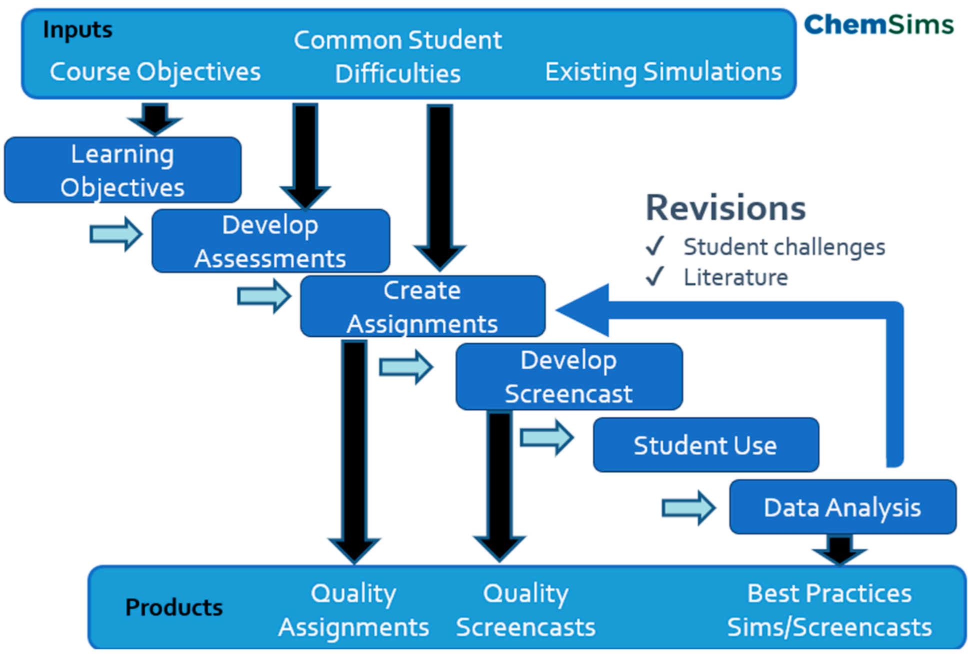


Figure 1. Scheme depicting the development process for each topic.

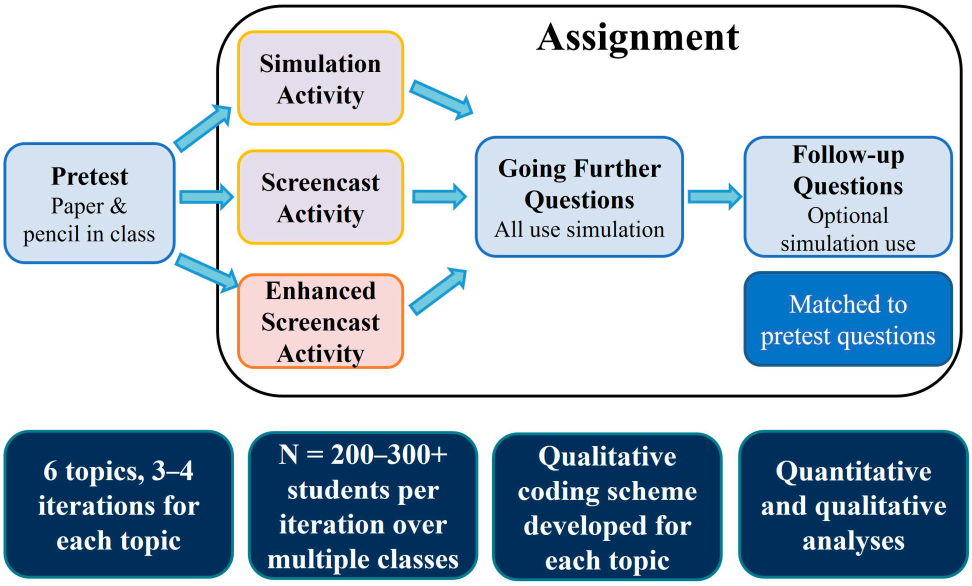


Figure 2. Classroom study design.

These data were used in the assignment revision process and as evidence for the efficacy of these assignments in supporting student development and use of accurate scientific particulate-level mental models. The revisions were a key aspect of the project and involved refining questions and instruction prompts to address the specific challenges or misinterpretations that the students were experiencing. For all the initial studies the screencasts were designed to parallel the scaffolded use of the simulations. However, for one particularly challenging topic, we incorporated additional instructional elements into the screencast assignment, allowing us to investigate the effects of an "Enhanced Screencast Activity". These enhancements are discussed in more detail in Section 3.1. Typically, three to four iterations of assignments for each topic were evaluated using mixed methods (Figure 2). Our study of these assignments was guided by two primary research questions:

1. How can we use simulations or screencasts to support students' conceptual understanding in chemistry outside of class?
2. What are the benefits and challenges of the guided interactive use of a simulation and the viewing of a screencast of the same topic?

3. Discussion

3.1. Key Takeaways

Based on the combined results of the detailed studies for each individual topic [36–40], we identified several key takeaways for the development of preclass activities that use simulations. They include the following:

- Students were able to learn content effectively from either direct simulation use or from engaging with screencasts paired with answering questions, and the activities raised the average understanding to a similar level regardless of prior knowledge;
- These preclass activities supported student development of particle-level models and provided a common experience that instructors could effectively build upon through classroom instruction;
- Screencasts provided several advantages over student-guided simulation use that included being able to avoid potential simulation limitations or seamlessly adding instructional content to support student learning (Figure 2: Enhanced Screencast);
- Assignment design is effective when following a pattern of orientation, exploration, and application of knowledge and is iteratively revised.

For all the topics we studied, statistical analysis showed significant pre–post gains for both simulation and screencast assignments and in almost all cases, there were no significant differences in the learning gains for students who used the simulation on their own with scaffolded instructions or viewed a screencast where an instructor manipulated the simulation and highlighted key features. In fact, we observed that the class average often rose to a similar level regardless of the starting prior knowledge level (as measured via the pretest question). Figure 3 illustrates this as we see greater variation in the average pretest scores on the (a) Equilibrium and (b) Kinetics assignments than on the posttest score. This suggests the value of using these activities as a way to help mediate differences in students' incoming background knowledge.

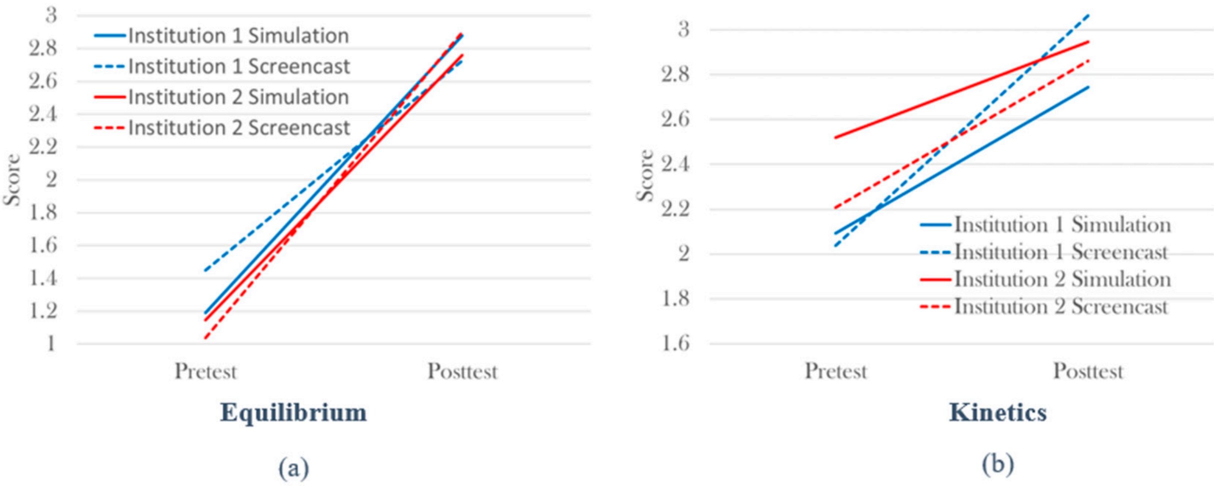
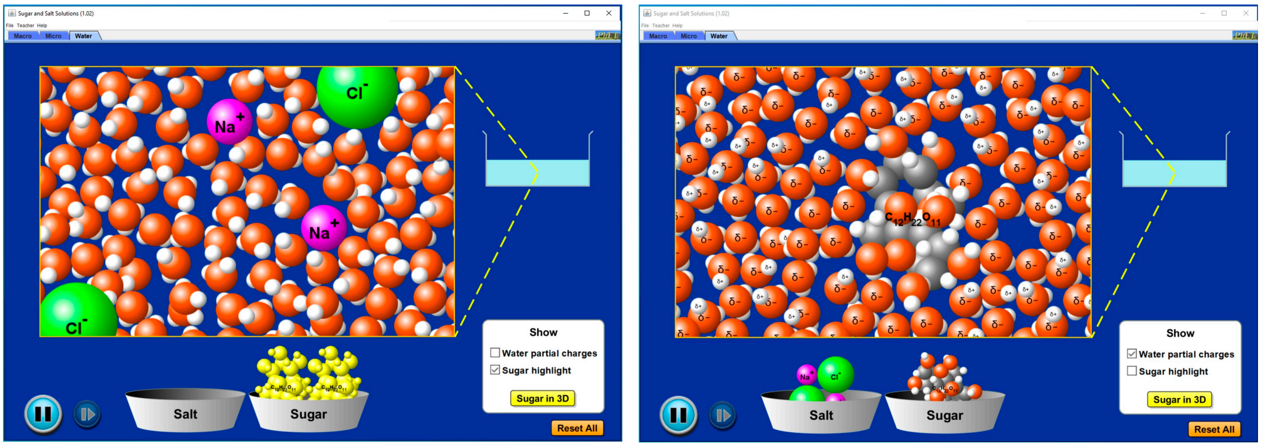


Figure 3. Pre–post class averages from two different institutions for: (**a**) Equilibrium; (**b**) Kinetics. Originally published by The Royal Society of Chemistry [40].

Both styles of preclass activities were effective at helping the students begin to build particle-level mental models. For example, in the Gas Laws pre-assessment, students were asked to use particle motion to explain why a helium-filled weather balloon gets larger as it ascends, but 42% of students provided a macroscopic-level explanation. When asked this same question after completing either the simulation or screencast activity 13.5% (simulation)—30% (screencast) of these students moved to a particle-level explanation [39].

Even though this work demonstrated that either student-guided use of simulations or screencasts can support students' conceptual learning and building of particle-level mental models, we also identified several advantages in using screencasts that might influence an instructor's choice. Screencasts can eliminate technology issues that students may encounter when trying to use simulations, for example, using simulations that run on Java, which no longer runs easily on many devices. Additionally, all models and simulations have limitations. Students may hit these limits and obtain "inaccurate" results when they are independently manipulating the variables in a simulation. However, as novice learners students may not recognize these limitations or the inaccuracy of the results leading to incorrect interpretations. For example, when using the PhET reaction and rates simulation [42], students were asked to heat up the system and observe what happened to the total energy. However, many students using the simulation stated that there was no change in the total energy for the system. This was possible because the students had "maxed out" the bar indicating total energy prior to heating the system. Thus, the simulation could not show a change in total energy [38]. This issue was not present for the screencast, as the instructor avoided such a potential source of confusion.

Further, for particularly complex simulations, screencasts allow instructors to better focus students' attention on key interactions. For example, Figure 4 shows two screens from the water tab of the PhET sugar and salt solutions simulation [43], which was used in our solubility activity. Though the simulation does a good job of illustrating how polar water molecules interact with ions (a) and with the polar sugar molecule (b), as these particles are moving around a lot and there are many of them on the screen, it is challenging for a novice learner to focus in on the specific interactions that best illustrate how water molecules orient themselves around ions and polar molecules. In the assignment, when asked about the interactions between water molecules and sucrose, most screencast students (85%) said that there were interactions between water and sucrose that were similar to those between two water molecules. However, most simulation students (88%) said that there were no interactions or indiscriminate interactions between water and sucrose. This suggests that on their own, even with substantial scaffolding that was revised three times, students were largely unable to discern the interactions between water and sucrose in this complex visual.



(a) (b)

Figure 4. Screenshot from the PhET Sugar and Salt Solutions simulation of water molecules interacting with: (**a**) ions; (**b**) a polar sugar molecule.

It is also generally easier to develop screencasts that highlight the important features, patterns, or interactions in a simulation than it is to provide written scaffolding that will get students to the same place. Though we were often able to find the right scaffolding eventually, it typically required several more iterations of the assignment revision cycle to get the scaffolding in the simulation assignments "right". This was especially true for complex simulations that have many different parts or variables to pay attention to, and in some cases, we were never able to obtain equal outcomes for simulation and screencast assignments [37,39]. Further, screencasts present material both visually and verbally. Since auditory and visual information is processed through different channels and simultaneously [44], this allows for dual coding, which has been shown to improve learning and retention of material [45,46].

Finally, screencasts allow instructors to supplement the simulations with additional content to further support student learning. For example, in a screencast, an instructor can provide side-by-side pictures of the simulation under separate conditions to better illustrate the effect of changing a variable. Alternatively, an instructor might provide additional images and verbal commentary. We found this to be important for particularly difficult or abstract concepts such as understanding how chemical potential energy is associated with bond breaking and bond forming [36]. Though the Atomic Interactions PhET simulation [47] used to illustrate this relationship is relatively simple with few moving parts, the concept is particularly challenging for students. The simulation focuses on the chemical potential energy changes associated with the formation and breaking of a single attractive force between two particles. Despite multiple iterations of both the simulation and screencast versions of this activity, it was only when we enhanced the screencast (Figure 2: Enhanced Screencast Activity) with additional images and supplemental instruction that explicitly connected the energy changes associated with the breaking and forming of individual bonds to the overall energy changes at the larger system level of a chemical reaction that we saw students better able to make connections between the simulation and the overall exo or endothermic nature of a chemical reaction [36].

3.2. Implications for Assignment Design

Based on these results from the ChemSims project and the use of such simulations and screencast assignments in our classrooms, we have identified three important guidelines in developing preclass activities that incorporate simulations:

1. The structure of the assignments should include (a) an orientation, (b) an opportunity for students to identify patterns and make connections, and (c) an opportunity for students to practice and assess their knowledge;
2. The activities should be viewed as the starting point of learning, which the instructor can build upon during in-class instruction and student work;
3. Employing an online format gives immediate access to student responses that allow the faculty to quickly identify challenges that students are experiencing.

Whether developing a simulation activity or screencast, one goal of the assignment is to help students understand key features of the simulation concerning what variables can be manipulated, how such variables are manipulated, the different types of visuals (e.g., graphs, vectors, particle motion, etc.) available, and how to interpret those visuals. This can be achieved by asking the students directed questions that require them to manipulate variables and make observations (simulation) or systematically demonstrating them (screencast) and asking students to answer related questions. PhET simulations are designed to support student construction of conceptual understanding through exploration with minimal guidance [30,48]. However, our experience was that when used outside the classroom (which lacks immediate instructor support), assignments missing scaffolded orientation would result in students missing critical features of the simulation. This was especially true for more complex simulations with several different variables and display options. After orientation to the simulation, it is important to provide scaffolding that will help students focus their attention on aspects of the simulation to help them identify key interactions or patterns. We found success in having students make a series of related or contrasting observations, sometimes having them summarize their results in a data table, and then asking them to develop conclusions based on the gathered information. For some concepts, it was then possible to have students investigate other relationships on their own. For example, in the gas laws activity, students explored how the pressure of a gas was affected by changes in the volume of the container or type of gas with more guided scaffolding or by watching the screencast and were then asked to determine relationships between other variables (e.g., temperature and pressure) on their own using the simulation. Finally, students' understanding of key concepts was assessed by asking them to explain the patterns or relationships they identified based on their observations from the simulation and to apply their newly acquired ideas to other relevant phenomena. This allows students to test their knowledge and self-assess their understanding and provides instructors with critical feedback about the level of student knowledge at the end of the activity.

Building scientifically accurate mental models for core science concepts is challenging and takes time and multiple exposures [49]. This is the basis for using these activities as the starting point for student learning. However, the key to building on these activities during in-class instruction is identifying challenges and gaps in student understanding from student responses (Guideline 2). This is made significantly easier when using a platform like Google Forms to collect student responses in an electronic format (Guideline 3). In our experience, quickly scanning through student responses to a particular question can indicate patterns. Example responses that highlight these patterns can be used to drive productive class discussion about the core concept. This can support students in refining and building on their initial ideas while simultaneously validating the time and effort that students put into completing the assignment. Further, it allows instructors an opportunity to normalize making mistakes and refining ideas as an important part of building science knowledge and the learning process in general. Identifying "common responses" given by students and discussing the strengths and weaknesses of each as a class provides important formative feedback to students in a non-threatening manner that does not require an instructor to give individual feedback to each student. We have found that this approach goes a long way toward helping develop a more learner-centered classroom environment and supporting student buy-in for employing preclass activities.

4. Conclusions

A 2012 Report from the National Academies of Sciences focused on the state of discipline-based education research stated that "In general, students have difficulty understanding phenomena and interactions that are not directly observable, including those that involve very large or very small spatial and temporal scales" [1]. Simulations can help students understand these phenomena and interactions and using simulations outside of the classroom can allow students with different incoming background knowledge to engage with the content for as long as they need. The ChemSims project allowed us to explore and evaluate different methods for using simulations to support student building of mental models and conceptual understanding of core chemistry concepts outside of the classroom. Through multiple iterations, we identified an effective activity design strategy of orientation, exploration, and application of knowledge. Findings from this project and previous studies indicate that self-exploration of simulations with appropriate scaffolding can be used to support student learning of core concepts. However, when using simulations outside of the classroom for preclass or homework activities, screencasts may provide several advantages for both students and the instructor. Importantly, screencasts can allow instructors to capitalize on the benefits of dual coding by providing simultaneous visual and auditory information [45] and provide supplemental instruction to extend what can be gleaned from the simulation alone. Further, in a screencast, the instructor can ensure that certain simulation conditions are examined, which may not happen during self-exploration if students are just trying to get through the activity.

5. Future Work

Much research on the flipped learning model has so far focused on the benefits of the increased level of in-class work and collaboration. Yet, recent studies and reviews suggest that the preclass instruction is a critical element of effective flipped models [12,17] and that online options can have some benefits over more traditional reading assignments [19]. One area of future work is to examine what benefits may exist for incorporating these types of particle-level simulations into preclass activities for flipped classrooms. Additionally, though we did not focus on laboratory instruction in this paper, laboratory instruction has been frequently critiqued with respect to its support of student learning [50–52]. One reason for this may be the broad set goals for laboratory that range from learning skills to learning concepts. Though conceptual simulations such as the ones discussed in this paper have been used for laboratory investigations, most laboratory simulations are operational, focusing on laboratory procedures. A possible area for future research is looking at the use of conceptual simulations in pre-lab activities to provide a conceptual background. Does having a background understanding of the chemistry concept being studied in a laboratory activity help students get more out of the laboratory activity?

Other future research should focus on using best practices in multimedia learning for the development of preclass activities. As a result of COVID-19, many people gained significant technical savvy in creating online materials, particularly screencasts, so making screencast assignments may not seem like a new concept. However, in most cases, this was done quickly out of immediate necessity and thus, these materials were often not created with best practices in supporting independent student-engaged learning or the principles for effective multimedia learning in mind. Thus, in addition to the considerations we identify above, if planning to develop screencast preclass activities, we strongly recommend exploring Mayer's 12 Multimedia Principles for Learning [23]. Mayer's definition of multimedia learning can also be viewed as dual-code or dual-channel learning grounded in dual-coding theory, which posits that the mind processes verbal and visual information through separate channels [45]. Further, he takes the perspective of multimedia learning as knowledge construction. Thus, the goal of multimedia presentations is not just to present information, but also to support the processing of the information by cueing what to pay attention to and how to organize the material, and how to relate the material to prior knowledge. Within this context, several principles stand out as critical in the production of video materials to support student conceptual learning or core chemistry concepts. First, the multimedia principle suggests that people learn best from a combination of words (verbal) and pictures (visual). This means that it is very important to include relevant visuals to support student learning. Second, the coherence principle indicates that learning is more effective when unnecessary information is excluded. This suggests limiting text on the screen to key terms and eliminating any extraneous images or animations that are not core to the concept being discussed, flashy transitions, story-based sidebars, or images of the narrator on the screen during the learning process. These things, as well as background sound or music, which should also be avoided, can distract a learner's attention away from the key content. A third principle is signaling, which suggests that learning is enhanced when cues such as highlighting, arrows, or circling are used to draw attention to important information. The last highly relevant principle to keep in mind when developing videos is the segmenting principle. Mayer found that learning is more effective when content is broken down into smaller well-articulated units. Features that support this are clear introductions that indicate what will be covered in the video or summaries that summarize what was covered, and title slides or headers that match the wording used in the introduction or summarizing organizers.

Another important consideration in developing screencasts is accessibility. In particular, closed captioning should be accurate, especially for technical words, and contain proper punctuation. Closed captioning is something that many students who do not have auditory challenges use to support their viewing of videos, especially if they use words students are not familiar with. Additionally, if analogies are used in supplementing the content in a simulation to help explain a topic, it is important to ensure all learners can connect to the analogy. One way to achieve this is to provide a visual alongside a verbal explanation that explicitly links the analogy to the science concept.

Author Contributions: D.G.H. and R.D.S. were equally involved in the writing, reviewing, and editing of this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The ChemSims project was funded by the National Science Foundation, grant numbers 1705365 and 1702592.

Institutional Review Board Statement: The ChemSims project was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Michigan State University (x15-799e, 12 August 2015) and Grand Valley State University (16-012-H, 17 July 2015).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the ChemSims Project.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors would like to thank the contributions of our collaborating researchers including Jessica VandenPlas, Stella Archiyan, Marissa Biesbrock, Shanna Hilborn, Brianna Martinez, Lauren Miling, Alec Shrode, and Elizabeth Sielaff.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. National Research Council. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*; Singer, S.R., Nielson, N.R., Schweingruber, H.A., Eds.; National Academies Press: Washington, DC, USA, 2012.
2. Johnstone, A.H. Macro and Microchemistry. *Sch. Sci. Rev.* **1982**, *64*, 377–379.
3. PhET Website. Available online: http://phet.colorado.edu/ (accessed on 1 September 2013).
4. Molecular Workbench. Available online: http://mw.concord.org/modeler/index.html (accessed on 19 December 2023).
5. ChemCollective. Available online: https://chemcollective.org/ (accessed on 19 December 2023).
6. Figueiredo, M.; Rafael, C.; Neves, J.; Vicente, H. Assessing the Impact of Computer Simulations on Physics and Chemistry Learning. In *Methodologies and Intelligent Systems for Technology Enhanced Learning, 13th International Conference*; Milrad, M., Otero, N., Sánchez-Gómez, M.C., Mena, J.J., Durães, D., Sciarrone, F., Alvarez-Gómez, C., Rodrigues, M., Vittorini, P., Gennari, R., et al., Eds.; Springer Nature: Cham, Switzerland, 2023; pp. 34–44.
7. Rutten, N.; van Joolingen, W.R.; van der Veen, J.T. The Learning Effects of Computer Simulations in Science Education. *Comput. Educ.* **2012**, *58*, 136–153. [CrossRef]
8. Wu, H.-T.; Mortezaei, K.; Alvelais, T.; Henbest, G.; Murphy, C.J.; Yezierski, E.F.; Eichler, J. Incorporating Concept Development Activities into a Flipped Classroom Structure: Using PhET Simulations to Put a Twist on the Flip. *Chem. Educ. Res. Pract.* **2021**, *22*, 842–854. [CrossRef]
9. Keengwe, J.; Kidd, T.T. Towards Best Practices in Online Learning and Teaching in Higher Education. *MERLOT J. Online Learn. Teach.* **2010**, *6*, 533–541.
10. Abeysekera, L.; Dawson, P. Motivation and Cognitive Load in the Flipped Classroom: Definition, Rationale and a Call for Research. *High. Educ. Res. Dev.* **2015**, *34*, 1–14. [CrossRef]
11. Turan, Z. Evaluating Whether Flipped Classrooms Improve Student Learning in Science Education: A Systematic Review and Meta-Analysis. *Scand. J. Educ. Res.* **2023**, *67*, 1–19. [CrossRef]
12. Bancroft, S.F.; Jalaeian, M.; John, S.R. Systematic Review of Flipped Instruction in Undergraduate Chemistry Lectures (2007–2019): Facilitation, Independent Practice, Accountability, and Measure Type Matter. *J. Chem. Educ.* **2021**, *98*, 2143–2155. [CrossRef]
13. Gong, J.; Cai, S.; Cheng, M. Exploring the Effectiveness of Flipped Classroom on STEM Student Achievement: A Meta-Analysis. *Technol. Know Learn* **2023**, 1–22. [CrossRef]
14. Hew, K.F.; Bai, S.; Huang, W.; Du, J.; Huang, G.; Jia, C.; Khongjan, T. Does Flipped Classroom Improve Student Cognitive and Behavioral Outcomes in STEM Subjects? Evidence from a Second-Order Meta-Analysis and Validation Study. In Proceedings of the Blended Learning. Education in a Smart Learning Environment, Bangkok, Thailand, 24–27 August 2020; Cheung, S.K.S., Li, R., Phusavat, K., Paoprasert, N., Kwok, L., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 264–275.
15. Cheng, L.; Ritzhaupt, A.D.; Antonenko, P. Effects of the Flipped Classroom Instructional Strategy on Students' Learning Outcomes: A Meta-Analysis. *Educ. Technol. Res. Dev.* **2019**, *67*, 793–824. [CrossRef]
16. Ezeh, C.; Adesope, O.; Kehinde, O.; Jaiyeola, E. Effects of Flipped Instruction on College Students' Learning in STEM Subject Domains: A Meta-Analysis. *J. STEM Educ.* **2022**, *23*, 33–48. [CrossRef]
17. Eichler, J.F. Future of the Flipped Classroom in Chemistry Education: Recognizing the Value of Independent Preclass Learning and Promoting Deeper Understanding of Chemical Ways of Thinking During In-Person Instruction. *J. Chem. Educ.* **2022**, *99*, 1503–1508. [CrossRef]
18. Rau, M.A.; Kennedy, K.; Oxtoby, L.; Bollom, M.; Moore, J.W. Unpacking "Active Learning": A Combination of Flipped Classroom and Collaboration Support Is More Effective but Collaboration Support Alone Is Not. *J. Chem. Educ.* **2017**, *94*, 1406–1414. [CrossRef]
19. Jensen, J.L.; Holt, E.A.; Sowards, J.B.; Heath Ogden, T.; West, R.E. Investigating Strategies for Pre-Class Content Learning in a Flipped Classroom. *J. Sci. Educ. Technol.* **2018**, *27*, 523–535. [CrossRef]
20. Chi, M.T.H.; Wylie, R. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educ. Psychol.* **2014**, *49*, 219–243. [CrossRef]
21. Pulukuri, S.; Abrams, B. Improving Learning Outcomes and Metacognitive Monitoring: Replacing Traditional Textbook Readings with Question-Embedded Videos. *J. Chem. Educ.* **2021**, *98*, 2156–2166. [CrossRef]
22. Seery, M.K.; Donnelly, R. The Implementation of Pre-Lecture Resources to Reduce in-Class Cognitive Load: A Case Study for Higher Education Chemistry. *Br. J. Educ. Technol.* **2012**, *43*, 667–677. [CrossRef]
23. Mayer, R.E. *Multimedia Learning*, 2nd ed.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2009; ISBN 978-0-521-73535-3.
24. Hegarty, M. Dynamic Visualizations and Learning: Getting to the Difficult Questions. *Learn. Instr.* **2004**, *14*, 343–351. [CrossRef]
25. Chamberlain, J.M.; Lancaster, K.; Parson, R.; Perkins, K.K. How Guidance Affects Student Engagement with an Interactive Simulation. *Chem. Educ. Res. Pract.* **2014**, *15*, 628–638. [CrossRef]
26. Akaygun, S.; Jones, L.L. How Does Level of Guidance Affect Understanding When Students Use a Dynamic Simulation of Liquid–Vapor Equilibrium? In *Learning with Understanding in the Chemistry Classroom*; Devetak, I., Glažar, S.A., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 243–263, ISBN 978-94-007-4366-3.
27. Wieman, C.E.; Adams, W.K.; Perkins, K.K. PhET: Simulations That Enhance Learning. *Science* **2008**, *322*, 682–683. [CrossRef]
28. Banda, H.J.; Nzabahimana, J. Effect of Integrating Physics Education Technology Simulations on Students' Conceptual Understanding in Physics: A Review of Literature. *Phys. Rev. Phys. Educ. Res.* **2021**, *17*, 023108. [CrossRef]

29. Stang, J.B.; Barker, M.; Perez, S.; Ives, J.; Roll, I. Active Learning in Pre-Class Assignments: Exploring the Use of Interactive Simulations to Enhance Reading Assignments. *arXiv* **2016**, arXiv:1607.04588.
30. Adams, W.K.; Paulson, A.; Wieman, C.E.; Henderson, C.; Sabella, M.; Hsu, L. What Levels of Guidance Promote Engaged Exploration with Interactive Simulations? In Proceedings of the AIP Conference Proceedings, Edmonton, AB, Canada, 23–24 July 2008; pp. 59–62.
31. Adams, W.K.; Armstrong, Z.; Galovich, C. Can Students Learn from PhET Sims at Home, Alone? In Proceedings of the International Conference on Physics Education, College Park, MD, USA, 18 December 2015; pp. 23–26.
32. Keehner, M.; Hegarty, M.; Cohen, C.; Khooshabeh, P.; Montello, D.R. Spatial Reasoning With External Visualizations: What Matters Is What You See, Not Whether You Interact. *Cogn. Sci.* **2008**, *32*, 1099–1132. [CrossRef] [PubMed]
33. Rieber, L.P. Effects of Visual Grouping Strategies of Computer-Animated Presentations on Selective Attention in Science. *Educ. Technol. Res. Dev.* **1991**, *39*, 5–15. [CrossRef]
34. Mayer, R.E.; Moreno, R. Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educ. Psychol.* **2003**, *38*, 43–52. [CrossRef]
35. ChemSims. Available online: https://chemsims.com/ (accessed on 16 January 2024).
36. VandenPlas, J.R.; Herrington, D.G.; Shrode, A.D.; Sweeder, R.D. Use of Simulations and Screencasts to Increase Student Understanding of Energy Concepts in Bonding. *J. Chem. Educ.* **2021**, *98*, 730–744. [CrossRef]
37. Herrington, D.G.; Sweeder, R.D.; VandenPlas, J.R. Students' Independent Use of Screencasts and Simulations to Construct Understanding of Solubility Concepts. *J. Sci. Educ. Technol.* **2017**, *26*, 359–371. [CrossRef]
38. Sweeder, R.; Herrington, D.; VandenPlas, J.R. Supporting Students' Conceptual Understanding of Kinetics Using Screencasts and Simulations Outside of the Classroom. *Chem. Educ. Res. Pract.* **2019**, *20*, 685–698. [CrossRef]
39. Martinez, B.L.; Sweeder, R.D.; VandenPlas, J.R.; Herrington, D.G. Improving Conceptual Understanding of Gas Behavior through the Use of Screencasts and Simulations. *Int. J. STEM Educ.* **2021**, *8*, 5. [CrossRef]
40. Herrington, D.G.; Hilborn, S.M.; Sielaff, E.N.; Sweeder, R.D. ChemSims: Using Simulations and Screencasts to Help Students Develop Particle-Level Understanding of Equilibrium in an Online Environment before and during COVID. *Chem. Educ. Res. Pract.* **2022**, *23*, 644–661. [CrossRef]
41. Wiggins, G.P.; McTighe, J. *Understanding by Design*, 2nd ed.; Association for Supervision and Curriculum Development: Alexandria, VA, USA, 2005.
42. Reactions & Rates. Available online: https://phet.colorado.edu/en/simulations/reactions-and-rates (accessed on 19 December 2023).
43. Sugar and Salt Solutions. Available online: https://phet.colorado.edu/en/simulations/sugar-and-salt-solutions (accessed on 19 December 2023).
44. Paivio, A. *Imagery and Verbal Processes*; Holt, Rinehart, And Winston: New York, NY, USA, 1971.
45. Clark, J.M.; Paivio, A. Dual Coding Theory and Education. *Educ. Psychol. Rev.* **1991**, *3*, 149–210. [CrossRef]
46. Mayer, R.E.; Anderson, R.B. The Instructive Animation: Helping Students Build Connections between Words and Pictures in Multimedia Learning. *J. Educ. Psychol.* **1992**, *84*, 444–452. [CrossRef]
47. Atomic Interactions. Available online: https://phet.colorado.edu/en/simulations/atomic-interactions (accessed on 19 December 2023).
48. Adams, W.K.; Reid, S.; LeMaster, R.; McKagan, S.B.; Perkins, K.K.; Dubson, M.; Wieman, C.E. A Study of Educational Simulations Part 1—Engagement and Learning. *J. Interact. Learn. Res.* **2008**, *19*, 397–419.
49. Brown, P.C.; Roediger, H.L.; McDaniel, M.A. *Make It Stick: The Science of Successful Learning*; Harvard University Press: Cambridge, MA, USA, 2014; ISBN 978-0-674-41937-7.
50. Bretz, S.L. Evidence for the Importance of Laboratory Courses. *J. Chem. Educ.* **2019**, *96*, 193–195. [CrossRef]
51. Hofstein, A. The Laboratory in Chemistry Education: Thirty Years of Experience with Developments, Implementation, and Research. *Chem. Educ. Res. Pract.* **2004**, *5*, 247–264. [CrossRef]
52. Seery, M.K. Establishing the Laboratory as the Place to Learn How to Do Chemistry. *J. Chem. Educ.* **2020**, *97*, 1511–1514. [CrossRef]

education sciences

MDPI

Essay

Re-Envisioning Classroom Culture in an Introductory General Chemistry Course: Description of a Course Redesign Project

Song Wang [1] and Thomas J. Bussey [2,*]

[1] Department of Chemistry and Biochemistry, California State University-Dominguez Hills, Carson, CA 90747, USA; sowang@csudh.edu
[2] Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, USA
* Correspondence: tbussey@ucsd.edu

Abstract: In the U.S., the retention of students in STEM degree pathways has been an issue that many higher education institutions have and continue to face. Many of us in the chemistry education community have been reflecting on our own roles and responsibilities to create a more inclusive learning environment for all students in chemistry. Culturally relevant pedagogy (CRP) and culturally responsive teaching (CRT) are two influential frameworks that informed efforts in promoting inclusivity in chemistry classrooms. However, the current literature focuses primarily on isolated interventions, highlighting a need for theoretical development that articulates the synergy between the two frameworks and synthesizes them in the context of chemistry education. In this essay, we present a framework for re-envisioning chemistry classroom culture consisting of four tenets: culturally relevant chemistry knowledge, cultural validation, collectivist cultural orientations, and humanized chemistry learning environments. We identified five course redesign entry points: amplifying student voice, emphasizing group work, contextualizing content knowledge, scaffolding technical language, and revising assessment structures. We hope to present both a framework and a set of course redesign entry points for chemistry educators interested in re-envisioning their classroom culture. We will also discuss the evaluation plan of this project and future work to sustain student cultural assets in chemistry classrooms.

Keywords: chemistry education; culturally relevant pedagogy; culturally responsive teaching; introductory chemistry; course redesign; student engagement

1. Introduction

In a world that is increasingly shaped by science and technology, an equitable and high-quality science education is paramount to the sustainment and advancement of countries and their citizens. For decades, the United States (U.S.) has recognized, although not necessarily prioritized, this goal in the context of innovation capital and economic prosperity (National Academy of Sciences et al., 2010; National Academies of Sciences, Engineering, and Medicine, 2021). Not only is scientific literacy crucial for fostering the U.S. gross domestic product, but many reform efforts have also focused on meeting the demands for producing a diverse science, technology, engineering, and mathematics (STEM) workforce that is equipped to push scientific fields to pursue important and relevant issues for all communities. In recent years, there has been widespread interest in increasing the number of students completing bachelor's degrees in STEM disciplines, with the President's Council of Advisors in Science and Technology (PCAST) calling for one million additional STEM-educated university graduates (Olson & Riordan, 2012). Particularly, PCAST has stressed the importance of expanding the participation of individuals from diverse populations in STEM fields in order to reach the nation's full innovation potential (President's Council of Advisors on Science and Technology, 2021).

In the U.S., retention of students in STEM degree pathways has been a perennial issue that many higher education institutions continue to face. Particularly, the retention of racially, culturally, and linguistically minoritized STEM students in higher education has compounded the existing educational debt, i.e., the systemic inequities and injustices that persist within educational systems, that has continued to accumulate (Ladson-Billings, 2006). There are documented equity gaps for minoritized students in undergraduate chemistry courses (Estrada et al., 2016). These equity gaps for minoritized individuals extend into the U.S. workforce, with only 6.2% of chemists and material scientists, chemical engineers, and chemical technicians identified as Black or African American, and only 7.0% identified as Hispanx or Latinx, as compared to their U.S. population size of 13.7% and 19.5%, respectively (Vargas et al., 2023). Research has reported that although racially minoritized students choose STEM majors at rates similar to those of their White and Asian peers, their rate of degree completion was lower, highlighting an equity gap in higher education STEM pathways (Estrada et al., 2016).

Scholars researching how minoritized students experience STEM learning environments have uncovered several factors that contributed to their experience of exclusion. Black and Latinx undergraduate students experienced exclusion from STEM learning environments in four main ways: (1) faculty perpetuating exclusionary classroom culture, (2) study group peers perpetuating stereotype threat, (3) nuances in undergraduate student experiences with their cultural peers, and (4) unaddressed discrimination (Flores et al., 2024). Narratives from Black and Latinx STEM students have highlighted that the exclusionary classroom culture created by faculty trickles down to students and negatively impacts their efforts to develop productive relationships with both the faculty and their peers.

In recent years, following the murder of George Floyd and the rise of the Black Lives Matter movement, many of us in the chemistry education community began asking questions about our own roles and responsibilities to create a more diverse and equitable learning environment for all students in chemistry. Over the last several years, we have been pondering this question: "What does an anti-racist and decolonialized chemistry curriculum look like?" We think it is important to note that this is a question that many have been asking long before the broader community (ourselves included) began to engage in this work. In a September 2021 editorial, Greta Glugoski-Sharp, 2021–2022 President for the American Association of Chemistry Teaches, noted that "[c]ulturally responsive teaching is asking chemistry [educators] to "re-envision" their classrooms—not by throwing out lessons or activities that we know work, but rather by building on successful practices already in place". Similarly, discussions about anti-racist pedagogy have included conversations about culturally relevant pedagogy centering academic success, cultural competence, and critical consciousness (Gay, 2008; Ladson-Billings, 1995, 2006).

In this essay, we describe an integrated course redesign framework for an introductory general chemistry course that resulted from our cycle of research, reflection, discussion, implementation, and evaluation. Our goal was to bring together the two complementary frameworks, namely culturally relevant pedagogy (CRP) and culturally responsive teaching (CRT), to re-envision college general chemistry courses. In synthesizing the two frameworks, we propose an approach for re-envisioning chemistry classroom culture using two tenets of culturally relevant pedagogy (culturally relevant chemistry knowledge and cultural validation) and two tenets of culturally responsive teaching (collectivist cultural orientations and humanized chemistry learning environments). We then applied this approach in the context of introductory general chemistry and have identified five entry points for course redesign. The implementation and evaluation of the course redesign is still ongoing, so the current essay focuses on a reflective discussion of our redesign approach with examples from the resulting activities of course redesign.

1.1. Culturally Relevant Pedagogy and Culturally Responsive Teaching

One important motivation for us to engage with culturally relevant pedagogy and culturally responsive teaching is our desire to think about teaching and learning beyond the boundaries of course content material. We wanted to expand our perspective on what it means to practice inclusive teaching. Ladson-Billings ignited the movement of culturally relevant pedagogy (CRP), where academic success, cultural competence, and critical consciousness to understand and critique social issues were weaved together to support student learning (Ladson-Billings, 1995). Gay proposed culturally responsive teaching (CRT) as using the cultural knowledge, lived experiences, frames of reference, and styles of performance of racially, ethnically, and linguistically diverse students to tailor learning interactions to suit their learning needs and make learning environments more inclusive and effective for them (Gay, 2008). The goal for science education through the adoption of culturally responsive teaching is to foster in students a new perspective on science and its impact on society.

Both culturally relevant pedagogy and culturally responsive teaching were proposed to improve educational experiences for students from diverse backgrounds, particularly students who are historically marginalized in education. As shown in Figure 1 above, CRP and CRT have a complementary relationship. The central consideration of both CRP and CRT is teaching from an asset-based perspective that includes students' experiences, interests, and backgrounds. CRP focuses on incorporating cultural references in the course content to empower students intellectually, emotionally, socially, and politically (Ladson-Billings, 1995). CRT focuses on creating inclusive learning environments and engaging students through relationship building (Gay, 2008). CRP's emphasis on empowering students is complemented well by CRT's focus on adapting instructional methods to create more inclusive learning environments. The two approaches share a commitment to affirming students' identities and making learning experiences inclusive and meaningful for students.

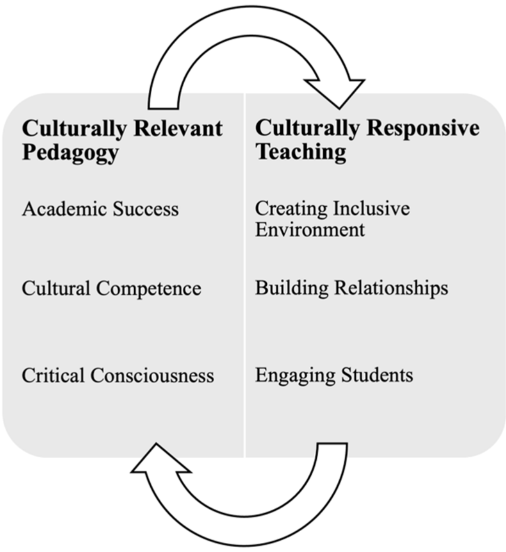


Figure 1. Principles of Culturally Relevant Pedagogy and Culturally Responsive Teaching. The principles of the two frameworks are mutually reinforcing.

1.2. CRP and CRT in Chemistry Courses

Chemistry educators and education researchers have been advancing equitable and inclusive chemistry teaching to reduce equity gaps in undergraduate chemistry courses (Aoki et al., 2022; Scanlon et al., 2018; Stoddard, 2022; White et al., 2021). The role of culture in chemistry has been discussed by scholars with a focus on its marginalizing effect (Jumarito & Nabua, 2021; Oladejo et al., 2022; Rahmawati et al., 2017). Western culture tends to dominate the traditional chemistry curriculum and subsequent student learning, which may lead students from other cultural backgrounds to become disengaged in chemistry (Oladejo et al., 2022; Spencer et al., 2022). Implementing CRP and CRT can help to show the relevance of chemistry to students' lived experiences and demonstrate how developing their expertise in chemistry can help them tackle problems of relevance to their communities (Roque-Peña, 2024; Winstead et al., 2022). The central goal of implementing culturally relevant pedagogy and culturally responsive teaching is to create learning environments that help students recognize, honor, and develop their own cultural beliefs and identity.

CRP and CRT have been applied in course redesign projects around the globe and in various educational settings. Internationally, there has been great interest in the Global South to connect chemistry content knowledge with local cultural practices to incorporate culturally relevant chemistry knowledge and validate student culture backgrounds. Chemistry education researchers and practitioners from many countries have produced valuable insights into ways in which chemistry learning environments can be adapted to different local cultural contexts.

Bridging learning in the classroom and at home illustrates the tenets of cultural validation and humanizing the chemistry learning environment. For example, Rodenbough and Manyilizu developed chemistry lesson plans that incorporated readings about *Ocimum kilimandscharicum*, a plant found primarily in Kenya, in lesson plans focused on using virtual laboratory tools to teach topics related to chemical bonding and molecular geometry (Rodenbough & Manyilizu, 2019). In Nigeria, Oladejo et al. (2022) reported a culture-techno-contextual approach for incorporating an assignment on indigenous knowledge and cultural practices in a lesson on radiation. Students were tasked with asking their elders about cultural practices and knowledge about radiation and communicating with their elders what they have learned in the classroom (Oladejo et al., 2022).

In addition to the African context, chemistry educators in Southeast Asian and South American countries also reported projects that incorporated culturally relevant chemistry knowledge (Rahmawati et al., 2017; Jumarito & Nabua, 2021). In the Philippines, Jumarito and Nabua (2021) integrated indigenous practices such as Subanen practices of sourcing water with bamboo water drains into chemistry instructional materials to incorporate culturally relevant chemistry knowledge in the curriculum. High school chemistry teachers in Indonesia implemented culturally responsive teaching using ethnochemistry texts that highlighted the relationship between cultural practices such as the tradition of burning frankincense and related chemistry concepts (Rahmawati et al., 2017). Lastly, in Brazil, chemistry educators incorporated culturally relevant chemistry knowledge in their curriculum by integrating the traditional use of cactus pear by rural communities in the Brazilian northeast to create a lesson on the saponification reaction (Pereira Gomes et al., 2024).

In the U.S., projects that utilized CRP and CRT had been primarily conducted to engage culturally minoritized students in chemistry to advance diversity, equity, and inclusion in chemistry education (Pickering et al., 2023). For example, chemistry education researchers in Michigan collaborated with Iḷisaġvik College in Alaska to develop a unit on snow chemistry that incorporated the traditional knowledge of the Iñupiaq community into the chemistry curriculum (Spencer et al., 2022). CReST, a new high school curriculum that paired chemistry and world history, incorporated the tenet of cultural validation by including a case study about the fresco (a form of mural art) lifecycle in the curriculum to relate chemistry content knowledge to students' cultural heritage to increase student interest and performance in chemistry (Ferri & White, 2024).

Storytelling was also a popular format for integrating CRP and CRT in chemistry classrooms. Sanders Johnson reported an organic chemistry course redesign project at Spelman College that incorporated the story of Uncle Nearest Premium Whiskey (Sanders Johnson, 2022). The story centered around Nearest Green's great-great-granddaughter who serves as the Master Blender. Presenting a narrative that centers Black women provides culturally relevant learning material that connects to a student demographic that is often excluded from the traditional chemistry curriculum. Similarly, chemistry instructors at a Historically Black College and University (HBCU) in Baltimore explored ways to humanize the chemistry learning environment in remote general chemistry laboratory courses by using storytelling (Winstead et al., 2022). Through the creation of "The Mystery of Mr. Johnson" series, the redesigned laboratory curriculum leveraged storytelling to illustrate the role chemistry can serve in advancing equity in local community. These examples highlight the broad interest from chemistry educators in the potential for transforming how chemistry is taught both in K-12 and college setting through re-envisioning chemistry classroom culture and practices.

1.3. Framework for Re-Envision Chemistry Classroom Culture

The applications of CRP and CRT discussed above had a common theme of re-envisioning chemistry classroom culture to address the sense of disconnection that makes students, particularly those from minoritized cultures, become disengaged in chemistry (Oladejo et al., 2022; Spencer et al., 2022). Shown in Figure 2, we synthesized the two frameworks and their applications in different contexts into a single framework for re-envisioning chemistry classroom culture that consists of four tenets. The four tenets are culturally relevant chemistry knowledge, cultural validation, humanized chemistry learning environment, and collectivist cultural orientation.

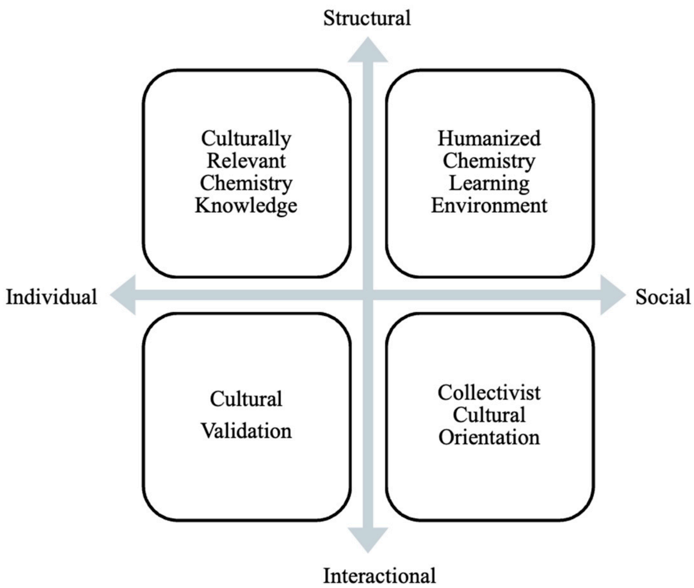


Figure 2. Four tenets of re-envisioning chemistry classroom culture. These tenets include culturally relevant chemistry knowledge, humanized chemistry learning environment, cultural validation, and collectivist cultural orientation. The four tenets are mapped onto two axes: the individual–social axis and the structural-interactional axis.

The four tenets can be thought of as four quadrants organized along two axes: the horizontal individual–social axis and the vertical structural-interactional axis. The individual–social axis describes whether a tenet focuses on individual students or the social environment, respectively. The structural–interactional axis describes whether a tenet focuses on curricular and pedagogical structures of a course or instructor-student and student-student interactions during class sessions, respectively. Taken together, the axes describe various aspects of chemistry classroom culture that are the focus of each tenet. Cultural validation focuses on interactions that acknowledge and honor individual students' cultural backgrounds. Culturally relevant chemistry knowledge focuses on creating course structures that incorporate individual cultural knowledge. Humanized chemistry learning environments focus on developing course structures that create an inclusive educational setting. And lastly, collective cultural orientation focuses on enacting social discourses that orient student interactions towards community building. In our course redesign project, the application of this framework in the context of introductory general chemistry course produced five entry points for enacting course redesign. We will discuss the entry points and provide an example for each in the next section.

2. Approach to Course Redesign

Applying our framework, in Fall 2024, we began a course redesign project in an introductory chemistry course designed to support students as they build chemical intuition and quantitative reasoning about observable, natural phenomena. This course serves to prepare students to move into and through the general chemistry sequence. General chemistry courses are foundational gateway courses for many majors and careers. However, equity gaps in these courses are well documented in the literature, and they often present barriers for minoritized students that prevent them from continuing their academic paths, and in some cases, lead to their departure from STEM fields or higher education altogether (Estrada et al., 2016; Goethe & Colina, 2018). With a substantial increase in students placing into preparatory college mathematics courses, enrollments in preparatory general chemistry courses have grown as well. This provides a unique opportunity to re-envision chemistry classroom culture for a student population potentially vulnerable to dropping out of STEM pathways early on in their academic journey.

2.1. Entry Points for Course Redesign

The goal of our course redesign project is to design and implement instructional practices to validate student's identity as someone who is developing their expertise in chemistry to address important social justice issues facing their community and create opportunities for students to discuss chemistry concepts in meaningful ways that connect concepts to relevant and authentic situations in their lives. In the previous section, we presented a review and synthesis of existing literature on CRP and CRT in chemistry course redesign projects. We then proposed a framework for re-envisioning chemistry classroom culture. In this section, we present a set of five entry points for course redesign that emerged by applying the framework in an introductory college chemistry course. Figure 3 below shows an overview of the five entry points for the course redesign. For each entry point, we will describe the implementation of the course redesign that can be produced, provide an example from our own course redesign project, and discuss how the redesign connects to our framework.

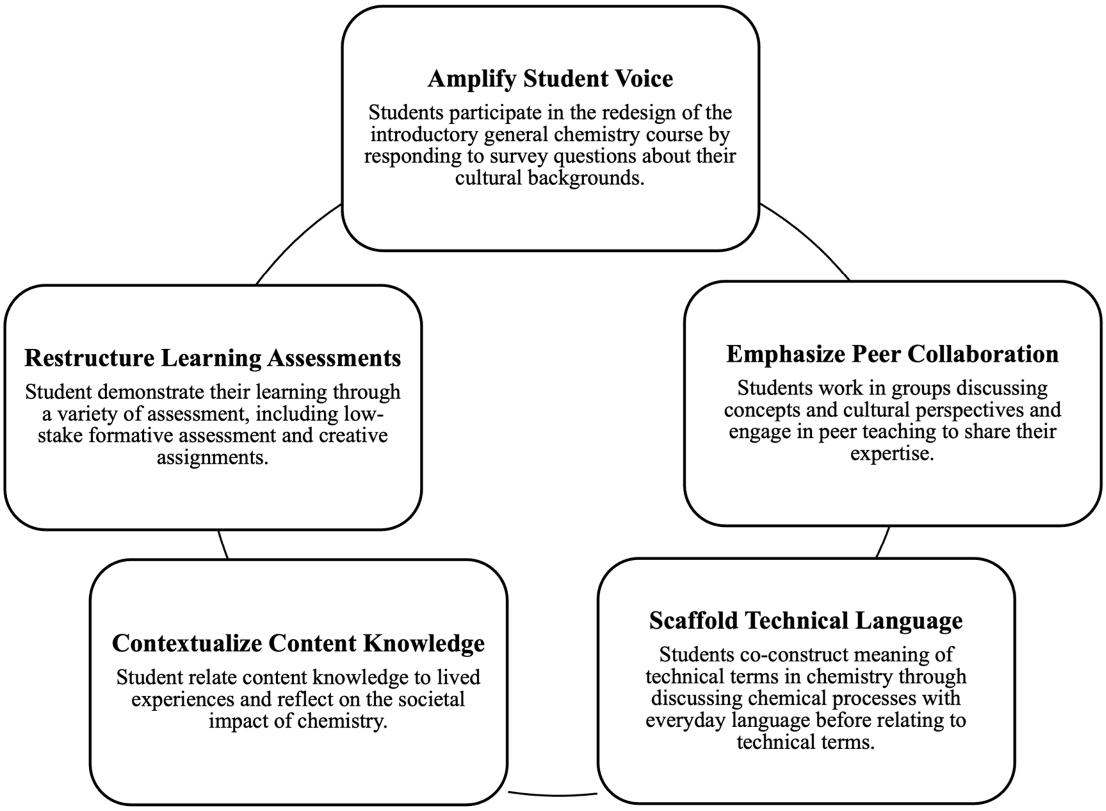


Figure 3. Five entry points for re-envisioning classroom culture. These entry points include: (1) amplify student voice, where students participate in course redesign by sharing their cultural backgrounds; (2) emphasize peer collaboration, which fosters group discussion and peer teaching; (3) scaffold technical language, helping students connect everyday language with scientific terminology; (4) contextualize content knowledge, linking chemistry concepts to students' lived experiences and societal impacts; and (5) restructure learning assessment, providing students with various formats to demonstrate their learning. These five entry points are threaded together by a focus on leveraging students' cultural knowledge as an asset for learning chemistry.

2.1.1. Amplify Student Voice

As the first entry point, amplifying student voice in the redesign process is central to realizing the four tenets we discussed above. Incorporating culturally relevant knowledge, practicing cultural validation, fostering collective cultural orientation, and creating a humanized learning environment are all predicated on centering students' lived experiences. Re-envisioning chemistry classroom culture must, therefore, start with getting to know the students and the cultural knowledge they bring into the classroom. When students feel comfortable and encouraged to share their lived experiences with the instructor and their peers, the chemistry classroom becomes a space that provides students with opportunities to learn and exchange knowledge about their cultural backgrounds. This, in turn, creates a learning environment that values and validates the cultural backgrounds and identities of students. Moreover, minoritized students often experience a lack of agency in science learning where they need to fit into a learning environment and persist against that environment (Flores et al., 2024). It is imperative to address and attenuate this unbalanced power dynamic in the classroom when changing classroom culture.

From this entry point, we created channels for students to voice their needs and concerns. We used first day of class surveys to get to know our student beyond the information provided by the college registrar (see Supplemental Material). Before the first day of class, we administered a survey to ask students about the communities to which they felt a sense of belonging, the identities that were salient to them, the cultural practices that they found intriguing, and the issues that were affecting their communities. Student responses to these questions provided the foundation for us to implement culturally relevant pedagogy. As the course went into full swing, our effort to include student voice in the class redesign continued. After each exam, a short survey, similar to an exam wrapper, was sent out to promote metacognition and seek student feedback on course material and pedagogy (Hodges et al., 2020; Gezer-Templeton et al., 2017). We asked students about the instructional practices that they find most beneficial to their learning, as well as suggestions for changes to how the course is being facilitated (see Supplemental Material).

Instructors can create channels to amplify student voice such as a first day of class survey (Schmitt et al., 2013) or exam wrappers (Hodges et al., 2020). The goal of implementing a student survey at the beginning was to invite students to contribute to the course redesign proactively, rather than being positioned as recipients of our educational intervention. These surveys can also help instructors develop course material that include culturally relevant chemistry knowledge and the relationship between chemistry and cultural practices. The first day of class survey asked students to reflect on chemistry related cultural practices that they find intriguing and chemistry-related issues that their communities face. These reflections can serve as a starting point for meaningful conversations and discussions about chemistry during the course.

2.1.2. Emphasize Peer Collaboration

One entry point for fostering a collective cultural orientation and practicing cultural validation in introductory chemistry courses is increasing the amount of class time for in-class group work (Luzyanin, 2024; White et al., 2021). Fostering a collectivist cultural orientation requires that students explore the nature of science as a social, tentative, creative activity. Group work can also create opportunities for students to share their culture with their peers and feel a sense of validation and belonging (Rendon, 1994). In our project, the format of class sessions was structured to incorporate an increase in group collaborations using a jigsaw-like activity structure. Lecture notes were provided to students before the start of each class session and served as a guide for group work. At the start of the group activity, each student picked one of the topics covered in the module. Based on student interests, the instructor made minor reshufflings to ensure that each topic was covered by a similar number of students. The group activities consisted of two stages. In the first stage, students develop expertise on one of the content topics by engaging with scaffolded instructional materials where a set of guiding questions provide structure when students read text on the content topics. In the second stage, students form new groups consisting of at least one person who had developed expertise on each topic. These new groups then allowed students to have agency and be positioned as authorities to teach each other about the material they learned in the first stage.

As an example, we implemented a group activity on chemical reactions that started with students being assigned with one of the three topics: describing chemical changes with chemical equations, interpreting and balancing chemical equations, and performing quantitative analysis with chemical equations. In the first stage of the group activity, students formed groups based on their interests. Guiding questions for each topic encouraged students to relate chemistry content knowledge to cultural practices that were familiar to them. During this process, they prepared notes to share with students engaging with other topics. After students have completed their notes to answer all the guiding questions, they formed new groups consisting of at least one student who engaged with each topic. The second stage of group activity focused on peer teaching. Students took turns sharing their notes with their peers, communicating their chemistry and cultural expertise both orally and visually. After the second stage of the group activity, each student developed notes on all topics covered in a class module.

Emphasizing peer collaboration is an entry point related to the tenets of collectivist cultural orientation and cultural validation. The two-stage group activities were intended to facilitate interactions among students that lead to meaningful discussions about the chemistry content material (Karacop, 2017; Nolan et al., 2018; Tarhan et al., 2013). Students were encouraged to share the ways in which they can relate course content material and their lived experiences. The format of peer teaching was intended to encourage students to share their newly developed expertise with their peers. These social interactions could orient classroom culture towards collaboration and community building. Moreover, students could feel a sense of validation from sharing their cultural knowledge and building a learning community with their peers.

2.1.3. Scaffold Technical Language

From the tenets of culturally relevant chemistry knowledge and humanized learning environment, we identified scaffolding technical language as an entry point for course redesign. Chemistry content knowledge is often conveyed through highly technical language and symbolic representations (Tang, 2019). These technical terms and representations often seem far removed from daily life and give the impression that chemistry is a realm separated from the human experience. The language of chemistry presents additional barriers for students to develop and communicate their content knowledge expertise (Tang, 2019). To support conceptual understanding and add a human aspect to chemistry concepts, rather than introducing the definition of technical vocabulary upfront, the redesigned course material used chemical phenomenon and practices as a starting point to introduce the context in which the meaning of technical vocabulary emerges. After introducing the macroscopic chemical phenomenon, diagrams and simulations were provided to connect macroscopic phenomenon to microscopic representations. Technical nomenclature was developed once experiential knowledge of a phenomenon was established. Scaffolding the introduction of technical vocabulary can facilitate deeper understanding that goes beyond rote memorization of definitions and support emerging bilingual students (Afitska, 2016; Jung, 2019; Symons, 2021).

As an example from this entry point, we revised how technical terms such as enthalpy, calorimetry, exothermic, endothermic, were introduced during the class module on thermochemistry. The introduction of these technical terms began by introducing fundamental concepts that were already familiar to the students, such as energy, heat, and temperature. Animations and diagrams of heat transfer were introduced to illustrate the idea of a system, its surroundings, and the flow of heat energy during endothermic and exothermic processes (see Supplemental Material). Simpler language such as “heat content” were also used to introduce the concept of enthalpy. Additionally, food labels were presented to provide a familiar daily experience that led to the introduction of calorimetry experiments.

Similarly to technical language, symbolic representations in chemistry were introduced and explicitly related to the corresponding macroscopic chemical phenomenon and the submicroscopic processes. For example, students were already describing the changes they saw when they dropped a bath bomb in water or as they interpreted a baking recipe and balancing the ingredients. Symbolic representations were then introduced and used to represent these descriptions of phenomena. Scaffolding the learning of technical language and symbolic representations in chemistry is intended to relate technical terms to these experiences and encourage students to form personally meaningful connections between technical terms and their lived experiences (Jung, 2019; Tang, 2019).

2.1.4. Contextualize Content Knowledge

Course content in introductory chemistry classes often focuses on describing submicroscopic entities that, along with technical, symbolic language, can give students the impression of chemistry being too abstract and removed from meaningful, personal experiences (Sjöström & Talanquer, 2014). From the tenets of culturally relevant chemistry knowledge and cultural validation, we identified contextualizing content knowledge as an entry point for course redesign. Providing contextualized examples allows students to relate chemistry concepts and practices to their lived experiences (Sjöström & Talanquer, 2014; Tshojay & Giri, 2021; Urban et al., 2017). At the beginning of each class module, we introduced a scenario or practice that would lead to discussions about the topics in the class module. Student responses from the first day of class survey guided the preparation of these short lectures that provided context to the chemistry topics. In addition, we included social contexts and issues such as healthcare, pollution, climate change, etc., into the lectures before students form groups to situate the chemistry content and provide students with an opportunity to think critically about the applications of this content knowledge.

As an example from this entry point, we revised our module on classification of matter by contextualizing it in the important role pure substances play in our lives. Using the production of oxygen for treating COVID-19, we introduced the concept of a pure substance and the practice of isolating substances. The sociopolitical issue of oxygen shortages due to the recent pandemic provided context for students to not only think about the role of chemical practices in our society but to also critically reflect on the political process of how resources created through these practices were being allocated. Similarly, our redesigned module on chemical bonds started with an introduction of the importance of understanding chemical bonding in drug discovery. The context of drug discovery also prompted discussions about healthcare disparities that communities have experienced, particularly how less affluent and minoritized communities have been historically burdened with the risk of drug discovery but could not access the benefits of newly developed treatments.

Providing context for chemistry content knowledge before the introduction of the concepts can meaningfully situate the chemistry concepts within culture and society, emphasizing the role of chemistry in cultural practices (Sjöström & Talanquer, 2014). The content material covered in a general chemistry course often consists of largely abstract concepts relating to submicroscopic entities and processes, so contextualizing the content knowledge is an entry point for course redesign to change how culture is included in the chemistry classroom. Contextualizing chemistry concepts and practices in introductory general chemistry course can also serve to facilitate discussion about the societal impacts of chemical practices (Broman & Parchmann, 2014; Urban et al., 2017). These discussions can validate students' cultural experiences and practices in the discipline of chemistry. When students make connections between their cultural backgrounds and chemistry content knowledge, their cultural knowledge can become further validated as not only personal lived experiences but also scientific expertise.

2.1.5. Restructure Learning Assessment

From the tenets of humanizing the chemistry learning environment and collective cultural orientation, we identified the structure of learning assessments as an entry point. In our project, the structure of assessments was revised to encourage collaboration and peer support. High-stakes, individual exams often induce a great amount of stress for students, and their learning often becomes oriented towards performing on high-stake assessments instead of finding personally meaningful connections to course content material (Bardi et al., 2011; Willson-Conrad & Kowalske, 2018). In addition, revising the structure of assessment can create opportunities for students to demonstrate their learning in ways that are meaningful to them.

Our restructured exams adopted a two-stage format. During the first part of the exam, students spent time solving close-ended questions individually. The close-ended questions include explaining the submicroscopic process of macroscopic chemical phenomenon (e.g., explaining the phenomenon of an iron bike frame rusting after being left in the rain) and performing quantitative analysis of chemical processes (e.g., calculating the amount of propane required to produce enough heat to raise the temperature of 1 kg of water from 25 degrees Celsius to 75 degrees Celsius). During the second part of the exam, students engaged in open-ended activities in groups such as creating concepts maps for the course modules that were covered by the exam. Students were encouraged to work together and utilize resources available to them to demonstrate their learning in creative ways.

Revising the structure of the assessments was intended to expand the ways in which students can demonstrate their learning. Students could demonstrate their competency not only through solving exam problems but also creatively and collectively through making connections between chemistry concepts, cultural practices, and social issues. The revised assessment structure can humanize the chemistry learning environment by reducing student anxiety and valuing personal meaning-making in addition to problem solving, critical thinking, and scientific communication skills (Rempel et al., 2021; Sjöström & Talanquer, 2014). In addition to developing a scientifically accurate understanding of chemistry concepts, students can spend more time thinking about how chemistry is impacting their lives and how they can leverage what they learned to contribute to their communities (Broman & Parchmann, 2014). Revising the structure of learning assessments can also serve to reinforce the other redesigned class activities to create a learning environment that encourages students to lean on their collectivist cultural backgrounds as resources for learning.

3. Discussion

In this essay, we presented a theoretical synthesis of two frameworks, CRP and CRT, in the context of chemistry education to propose a framework for re-envisioning chemistry classroom culture. We then described five entry points for course redesign as we applied our framework to guide our work to transform an introductory general chemistry course. In Figure 4 below, we illustrate the relationship between the five entry points to course redesign and the four tenets of our framework. With amplifying student voice as the central focus, we identified each entry points by intersecting two tenets in our framework. We also want to recognize that every chemistry classroom is a phenomenon produced by its own unique intersections of social, cultural, historical, and political relationships, so re-envisioning chemistry classroom culture requires attending to the diversity of students in particular classrooms instead of taking a one-size-fits-all approach. The entry points described here are not meant to be an exhaustive list of all possible actions to re-envision classroom culture but a set of actions that we adopted in our educational and institutional context. Recognizing that education cannot be considered in a vacuum, we hope to present a theoretical framework and a set of entry points for redesign approaches that chemistry educators can tailor to their specific context. We also described here some examples of course redesign that we implemented from the entry points discussed above.

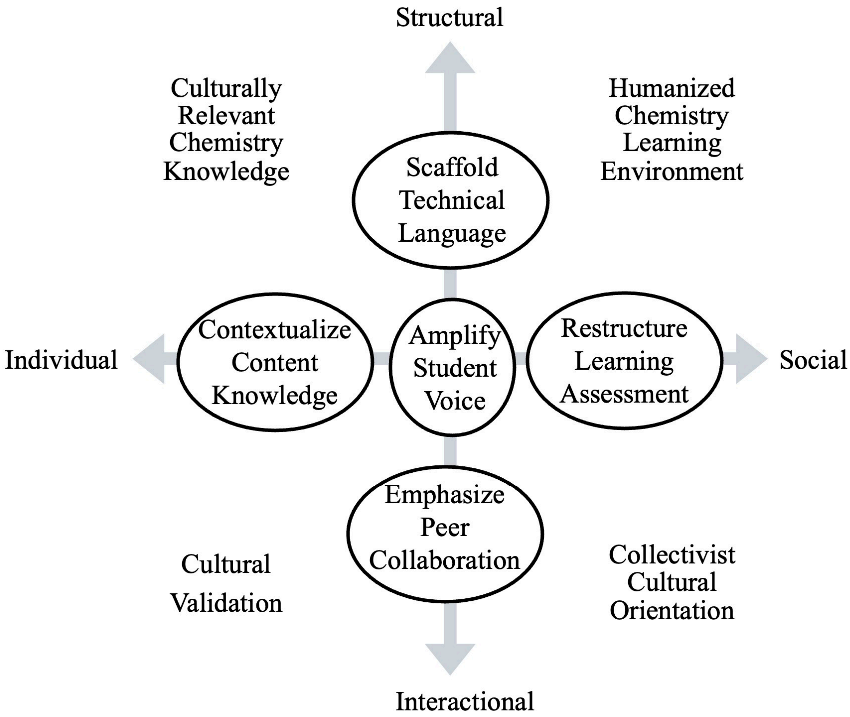


Figure 4. Connections between the five course redesign entry points and the four tenets of re-envisioning chemistry classroom culture. Central to our approach is the entry point of amplifying student voice. Each of the remaining four entry points are related to two tenets of re-envisioning classroom culture.

Johnson and Elliott highlighted that the transformation of STEM departments to become more inclusive entails teaching in ways that may be very different from the ways we, as faculty, originally learned our discipline (Johnson & Elliott, 2020). It requires educators to reflect on problems within the culture of chemistry in which we learned to succeed. Self-reflection on teaching practices has also been shown to be an important strategy for adopting teaching practices that are more culturally responsive (Civitillo et al., 2019). For chemistry educators interested in reflecting on their own teaching, it is productive to make culturally responsive chemistry teaching explicit by reflecting on the intersection between the culture of chemistry, the culture of higher education, and the cultures of minoritized communities (Aguirre & Del Rosario Zavala, 2013; Xie & Ferguson, 2024). For example, instructors can engage with reflections questions such as: how does my lesson make student's chemical thinking visible? How does my lesson foster and sustain student's participation in chemistry practices? How does my lesson make the hidden curriculum of higher education visible in critical ways? And how does my teaching validate student identity as someone who is developing their chemical thinking to critique and change important equity or social justice issues facing their community? Engaging with these reflection questions not only creates entry points for course redesign but also fosters a reflexive perspective towards the transformation of learning environments to be more inclusive. The goal is to continuously build our capacity to respond to the changes in cultural, social, and historical contexts where we encounter our students.

4. Future Work

As our course redesign project is still ongoing at the time of preparation of this essay, we are still in the process of collecting data that evaluates the impact of our interventions. We plan to evaluate student experiences of these interventions through student interviews and evaluation of student learning artifacts. Future plans for this instructional improvement project also aim to engage with research literature on culturally sustaining teaching in STEM disciplines to not only change classroom instructional practices to leverage students' cultural capital for learning chemistry but to also create chemistry learning activities that further extend students' cultural heritage. For example, place-based pedagogies such as virtual field trips may help humanize the educational environment by fostering deeper connections between classroom learning and local culture. These connections humanize the formal educational environment and leverage students' lived experiences outside classroom as resources for learning.

Since its conception, pedagogy that considered not only the content material but also students' cultural backgrounds have been framed with four main terms: culturally sensitive pedagogy, culturally relevant pedagogy, culturally responsive pedagogy, and culturally sustaining pedagogy. These terms are closely related but also have their own distinct ways of incorporating students' cultural backgrounds into pedagogy. Culturally sensitive pedagogy is primarily concerned with the awareness of cultural differences. Culturally relevant pedagogy is primarily concerned with linking students' cultural heritage and community cultural practices with the learning that takes place in the classroom. Culturally responsive pedagogy is primarily concerned with adapting the classroom learning environment to students' cultural capital and community cultural wealth. Culturally sustaining pedagogy is primarily concerned with ways to honor, explore, and extend students' cultural heritage (Ladson-Billings, 2021). As we continue our work to re-envision chemistry classroom culture, we aim to develop a pedagogy of response-ability, i.e., a pedagogy that focuses on our capacity to respond to students' diverse lived experiences, and we hope to galvanize chemistry educators to transform the culture of chemistry learning and teaching into one that celebrates the wealth of cultural knowledge brought by our students (Bozalek et al., 2018).

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci15030307/s1, First Day of Class Survey; Post-Exam Reflection; Example Group Activity.

Author Contributions: Conceptualization, S.W. and T.J.B.; course redesign, S.W. and T.J.B.; writing—original draft preparation, S.W. and T.J.B.; writing—review and editing, S.W. and T.J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Afitska, O. (2016). Scaffolding learning: Developing materials to support the learning of science and language by non-native English-speaking students. *Innovations in Language Learning and Teaching*, *10*(2), 75–89. [CrossRef]

Aguirre, J. M., & Del Rosario Zavala, M. (2013). Making culturally responsive mathematics teaching explicit: A lesson analysis tool. *Pedagogy: International Journal*, *8*(2), 163–190. [CrossRef]

Aoki, E., Rastede, E., & Gupta, A. (2022). Teaching sustainability and environmental justice in undergraduate chemistry courses. *Journal of Chemical Education*, *99*(1), 283–290. [CrossRef]

Bardi, M., Koone, T., Mewaldt, S., & O'Connor, K. (2011). Behavioral and physiological correlates of stress related to examination performance in college chemistry students. *Stress*, *14*(5), 557–566. [CrossRef]

Bozalek, V., Bayat, A., Gachago, D., Motala, S., & Mitchell, V. (2018). A pedagogy of response-ability. In *Socially just pedagogies in higher education: Critical posthumanist and new feminist materialist perspectives* (pp. 97–112). Bloomsbury Publishing.

Broman, K., & Parchmann, I. (2014). Students' application of chemical concepts when solving chemistry problems in different contexts. *Chemistry Education Research and Practice*, *15*(4), 516–529. [CrossRef]

Civitillo, S., Juang, L. P., Badra, M., & Schachner, M. K. (2019). The interplay between culturally responsive teaching, cultural diversity beliefs, and self-reflection: A multiple case study. *Teaching and Teacher Education*, *77*, 341–351. [CrossRef]

Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., Hurtado, S., John, G. H., Matsui, J., McGee, R., Okpodu, C. M., Robinson, T. J., Summers, M. F., Werner-Washburne, M., & Zavala, M. (2016). Improving underrepresented minority student persistence in STEM. *CBE—Life Sciences Education*, *15*(3), es5. [CrossRef] [PubMed]

Ferri, J. K., & White, R. S. (2024). Culturally relevant STEM (CReST): An integrated support curriculum for high school chemistry and world history. *Education Sciences*, *14*(2), 182. [CrossRef]

Flores, G. M., Bañuelos, M., & Harris, P. R. (2024). "What are you doing here?": Examining minoritized undergraduate student experiences in STEM at a minority-serving institution. *Journal of STEM Education Research*, *7*(2), 181–204. [CrossRef]

Gay, G. (2008). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.

Gezer-Templeton, P. G., Mayhew, E. J., Korte, D. S., & Schmidt, S. J. (2017). Use of exam wrappers to enhance students' metacognitive skills in a large introductory food science and human nutrition course. *Journal of Food Science Education*, *16*(1), 28–36. [CrossRef]

Goethe, E. V., & Colina, C. M. (2018). Taking advantage of diversity within the classroom. *Journal of Chemical Education*, *95*(2), 189–192. [CrossRef]

Hodges, L. C., Beall, L. C., Anderson, E. C., Carpenter, T. S., Cui, L., Feeser, E., Gierasch, T., Nanes, K. M., Perks, H. M., & Wagner, C. (2020). Effect of exam wrappers on student achievement in multiple, large STEM courses. *Journal of College Science Teaching*, *50*(1), 69–79. [CrossRef]

Johnson, A., & Elliott, S. (2020). Culturally relevant pedagogy: A model to guide cultural transformation in STEM departments. *Journal of Microbiology and Biology Education*, *21*(1), 5. [CrossRef]

Jumarito, E., & Nabua, E. (2021). Integrating indigenous practices in chemistry instructional material for a culturally responsive STEM curriculum. *Asia Research Network Journal of Education*, *4*, 1–15.

Jung, K. G. (2019). Learning to scaffold science academic language: Lessons from an instructional coaching partnership. *Research in Science Education*, *49*(4), 1013–1024. [CrossRef]

Karacop, A. (2017). The effects of using the jigsaw method based on the cooperative learning model in undergraduate science laboratory practices. *Universal Journal of Educational Research*, *5*(3), 420–434. [CrossRef]

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, *32*(3), 465–491. [CrossRef]

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, *35*(7), 3–12. [CrossRef]

Ladson-Billings, G. (2021). Three decades of culturally relevant, responsive, and sustaining pedagogy: What lies ahead? *The Educational Forum*, *85*(4), 351–354. [CrossRef]

Luzyanin, K. V. (2024). Workshops redeveloped: Group work to nurture a more inclusive learning environment. *Development in Academic Practice (Special)*, *2024*, 9–17. [CrossRef]

National Academies of Sciences, Engineering, and Medicine. (2021). *Call to action for science education: Building opportunity for the future*. The National Academies Press.

National Academy of Sciences, National Academy of Engineering & Institute of Medicine. (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5*. The National Academies Press.

Nolan, J. M., Hanley, B. G., DiVietri, T. P., & Harvey, N. A. (2018). She who teaches learns: Performance benefits of a jigsaw activity in a college classroom. *School Teaching and Learning Psychology*, *4*(2), 93. [CrossRef]

Oladejo, A. I., Okebukola, P. A., Olateju, T. T., Akinola, V. O., Ebisin, A., & Dansu, T. V. (2022). In search of culturally responsive tools for meaningful learning of chemistry in Africa: We stumbled on the culturo-techno-contextual approach. *Journal of Chemical Education*, *99*(8), 2919–2931. [CrossRef]

Olson, S., & Riordan, D. G. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Report to the President. Executive Office of the President.

Pereira Gomes, J., Silva de Lima, M., & Dantas Filho, F. F. (2024). Integrating local knowledge on cactus pear from Brazilian northeastern communities for culturally responsive chemistry teaching. *Journal of Chemical Education*, *101*(8), 3578–3583. [CrossRef]

Pickering, C., Lopez, M., Craft, E., Belknap, S., VanIngen-Dunn, C., Miller McNeill, L., & Rodriguez, J. (2023, June 25–28). *Theory to practice: Faculty professional development to integrate culturally responsive pedagogy and practices in STEM education to improve success of underserved students in STEM*. ASEE Annual Conference & Exposition Proceedings, Baltimore, MD, USA. [CrossRef]

President's Council of Advisors on Science and Technology. (2021). *Industries of the future institutes: A new model for American science and technology leadership*. Executive Office of the U.S. President.

Rahmawati, Y., Ridwan, A., & Nurbaity, N. (2017). Should we learn culture in the chemistry classroom? Integration of ethnochemistry in culturally responsive teaching. *AIP Conference Proceedings*, *1868*(1), 030009. [CrossRef]

Rempel, B. P., Dirks, M. B., & McGinitie, E. G. (2021). Two-stage testing reduces student-perceived exam anxiety in introductory chemistry. *Journal of Chemical Education*, *98*(8), 2527–2535. [CrossRef]

Rendon, L. I. (1994). Validating culturally diverse students: Toward a new model of learning and student development. *Innovative Higher Education*, *19*, 33–51. [CrossRef]

Rodenbough, P. P., & Manyilizu, M. C. (2019). Developing and piloting culturally relevant chemistry pedagogy: Computer-based VSEPR and unit cell lesson plans from collaborative exchange in East Africa. *Journal of Chemical Education*, *96*(6), 1273–1277. [CrossRef]

Roque-Peña, J. (2024). Student engagement in an introductory chemistry course at a Hispanic-serving institution: An exploratory case study. *Journal of Chemical Education*, *101*(8), 2966–2975. [CrossRef]

Sanders Johnson, S. (2022). Embracing culturally relevant pedagogy to engage students in chemistry: Celebrating Black women in the whiskey and spirits industry. *Journal of Chemical Education*, *99*(1), 428–434. [CrossRef]

Scanlon, E., Legron-Rodriguez, T., Schreffler, J., Ibadlit, E., Vasquez, E., & Chini, J. J. (2018). Postsecondary chemistry curricula and universal design for learning: Planning for variations in learners' abilities, needs, and interests. *Chemistry Education Research and Practice*, *19*(4), 1216–1239. [CrossRef]

Schmitt, K. R., Larsen, E. A., Miller, M., Badawy, A. H. A., Dougherty, M., Sharma, A. T., Hrapczynski, K., Andrew, A., Robertson, B., Williams, A., Kramer, S., & Benson, S. (2013). A survey tool for assessing student expectations early in a semester. *Journal of Microbiology & Biology Education*, *14*(2), 255–257.

Sjöström, J., & Talanquer, V. (2014). Humanizing chemistry education: From simple contextualization to multifaceted problematization. *Journal of Chemical Education*, *91*(8), 1125–1131. [CrossRef]

Spencer, J. L., Maxwell, D. N., Erickson, K. R. S., Wall, D., Nicholas-Figueroa, L., Pratt, K. A., & Shultz, G. V. (2022). Cultural relevance in chemistry education: Snow chemistry and the Iñupiaq community. *Journal of Chemical Education*, *99*(1), 363–372. [CrossRef]

Stoddard, S. V. (2022). The benefits of enlightenment: A strategic pedagogy for strengthening sense of belonging in chemistry classrooms. *Education Sciences*, *12*(7), 498. [CrossRef]

Symons, C. (2021). Instructional practices for scaffolding emergent bilinguals' comprehension of informational science texts. *Pedagogy: International Journal*, *16*(1), 62–80. [CrossRef]

Tang, K. S. (2019). The role of language in scaffolding content and language integration in CLIL science classrooms. *Journal of Immersion and Content-Based Language Education*, *7*(2), 315–328. [CrossRef]

Tarhan, L., Ayyıldız, Y., Ogunc, A., & Sesen, B. A. (2013). A jigsaw cooperative learning application in elementary science and technology lessons: Physical and chemical changes. *Research in Science & Technological Education*, *31*(2), 184–203.

Tshojay, P., & Giri, N. (2021). Contextualized approach in teaching chemistry: The perspectives and practices of teachers and students. *International Journal of Applied Chemistry and Biological Sciences*, *2*(4), 78–91.

Urban, S., Brkljaca, R., Cockman, R., & Rook, T. (2017). Contextualizing learning chemistry in first-year undergraduate programs: Engaging industry-based videos with real-time quizzing. *Journal of Chemical Education*, *94*(7), 873–878. [CrossRef]

Vargas, E. A., Scherer, L. A., Fiske, S. T., Barabino, G. A., & National Academies of Sciences, Engineering, and Medicine. (2023). Population data and demographics in the United States. In *Advancing antiracism, diversity, equity, and inclusion in STEMM organizations: Beyond broadening participation*. National Academies Press.

White, K. N., Vincent-Layton, K., & Villarreal, B. (2021). Equitable and inclusive practices designed to reduce equity gaps in undergraduate chemistry courses. *Journal of Chemical Education*, *98*(2), 330–339. [CrossRef]

Willson-Conrad, A., & Kowalske, M. G. (2018). Using self-efficacy beliefs to understand how students in a general chemistry course approach the exam process. *Chemistry Education Research and Practice*, *19*(1), 265–275. [CrossRef]

Winstead, A. J., McCarthy, P. C., Rice, D. S., & Nyambura, G. W. (2022). Linking chemistry to community: Integration of culturally responsive teaching into general chemistry I laboratory in a remote setting. *Journal of Chemical Education*, *99*(1), 402–408. [CrossRef] [PubMed]

Xie, J., & Ferguson, Y. (2024). STEM faculty's perspectives on adopting culturally responsive pedagogy. *Teaching in Higher Education*, *29*(5), 1215–1233. [CrossRef]

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Education Sciences Editorial Office
E-mail: education@mdpi.com
www.mdpi.com/journal/education

MDPI

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-4418-0