What Moves the Needle on DFW Rates and Student Success in General Chemistry? A Quarter-Century Perspective

Vijay Vyas
Marquette University

Scott A. Reid
Marquette University, scott.reid@marquette.edu

Follow this and additional works at: https://epublications.marquette.edu/chem_fac

Part of the Chemistry Commons

Recommended Citation
What Moves the Needle on DFW Rates and Student Success in General Chemistry? A Quarter-Century Perspective

Vijay S. Vyas
Department of Chemistry, Marquette University, Milwaukee, WI

Scott A. Reid
Department of Chemistry, Marquette University, Milwaukee, WI

Abstract

Second-term general chemistry is a key gateway course at our institution for a variety of STEM (Science, Technology, Engineering, and Mathematics) majors. Herein, we examine data on DFW% (i.e., % of D and F
grades and withdrawals) in our general chemistry courses over roughly a quarter-century period, focusing on sections taught by the authors. The data are discussed in light of changes in curriculum, pedagogy, and assessments, which included introduction of recitation sections, classroom response systems (clickers), and online homework, restructuring of course content into a modular format, followed by a redesign of the course modules along an anchoring concepts scheme, introduction of the flipped classroom, and introduction of regular low-stakes quizzing. We find that the lowest DFW%, roughly 5%, occurs in classrooms combining an active learning pedagogy, core concepts curriculum, and low-stakes assessments with instructor intervention. Trends in the DFW populations are examined for White, non-Hispanic students, first-generation students, and by binary gender. We find that women are increasingly underrepresented in the DFW population, which may reflect the recent incorporation of lower stakes assessments. To the degree that our lowest DFW rates are consistent with that found by other researchers using different but related approaches, these may represent a baseline target for DFW rates in general chemistry II. Finally, while a trend toward lower DFW rates is found for both general chemistry I and II, the follow-on course, organic chemistry I, shows an inverse trend, which emphasizes the need for larger scale reform.

KEYWORDS:
Introductory Chemistry, General Chemistry, Student Success, Flipped Classroom, Low-Stakes Assessments, DFW

Introduction

Whether by design or not, general and organic chemistry courses serve as key gateways for a variety of STEM (Science, Technology, Engineering, and Mathematics) majors, as the literature shows that poor outcomes in these courses can lead students to drop their pursuit of STEM degrees. At our institution, general chemistry II and organic chemistry I are particularly important gateway courses for at-risk students. Avoiding a student deficit mindset to this problem requires us as educators to consider what we teach and how we teach it, as, despite calls for reform, the basic structure of general and organic chemistry has changed little over the past 30 years. In the 2016 report Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students’ Diverse Pathways, the authors emphasize both the importance of and challenges associated with reforming curriculum in gateway STEM courses, noting that “Adoption of reformed curriculum and reformed teaching practices remains difficult because of such barriers as little support from other faculty and the department, few incentives for improved teaching, inappropriate classroom infrastructure, limited awareness of research-based instructional practices, and lack of time.” This sobering analysis emphasizes the need for changes in both curricula and teaching practices and reform at levels beyond that of instructor and course. To do so requires moving away from a “student-deficit” mindset, where the problem is presumed to lie with inadequate student preparation. Supporting this view, a recent analysis of data from more than 100,000 students across six institutions showed that introductory courses can disproportionately push underrepresented minorities out of STEM majors, even when accounting for disparities in academic preparation.

Progress has been made in introductory chemistry, as highlighted in a recent symposium series, which provided examples of successful reforms that improved student success and lowered DFW% (i.e., the % of student outcomes corresponding to withdrawal or grades of D or F). Note that, while DFW% is not a strict measure of learning, it is an important metric used administratively across a variety of fields to examine student retention and success. Considering our institution, a recent four year analysis of outcomes in introductory biology and chemistry showed a one-year retention for all students of 90% in the university and 75% in a STEM major. For students receiving a DFW in general biology I, those numbers dropped to 65% and 46%, respectively. Considering students receiving a DFW in general chemistry I, the figures were 70% and 50%. Importantly, for students receiving a DFW in both courses, the one year retention rate was 51% in the university and 18% in a STEM major.
Over the past roughly 25 years, we have implemented a range of reforms in our general chemistry sequence, spanning the areas of curriculum, pedagogy, and assessments. These have included the addition of recitation sections, classroom response systems (clickers or web-based), and online homework, the restructuring of course content into a modular format followed by a redesign of the course modules along a core or anchoring concepts scheme, introduction of the flipped classroom, and a movement from reliance on high-stakes examinations to regular low-stakes quizzing. Here, we look at each of these through the lens of their impact on DFW%, in total and by demographic group. We focus on eight on-sequence sections of our second-term general chemistry course (CHEM 1002) taught by one of the authors (S.A.R.) over the period from Spring 1999 to Spring 2022, encompassing some 1500 students. We compare DFW% in sections taught by different instructors, and we frame the data by showing trends from the first-semester course. We also examine trends in DFW% in first-term organic chemistry.

IRB Review
This study was submitted to the Institutional Review Board at our institution (IRB #4289) and was classified as not human subjects research and exempt from further review.

Methodology
Following IRB review, we obtained and examined deidentified student demographic and grade information for all on-sequence (i.e., fall general chemistry I, spring general chemistry II) sections of general chemistry I and II from the 1998–99 academic year to present, focusing in particular on general chemistry II sections taught by one of us (S.A.R.) over the period 1999–2022. The raw data set for general chemistry I included 16,890 records, of which 15,429 contained complete demographic data and were included in our analysis. For general chemistry II, excluding the Spring 2020 semester where a Pass/Fail grading option was offered, the raw data set contained 12,106 records, of which 11,379 contained complete data and were included in our analysis. We note that First-Generation status was available only in records from 2005 onward, and as we describe below, the data sets only encoded binary gender. All data was received as Excel data files and was analyzed using Excel and stored in password protected files.

Student Populations and Historical Trends
Setting the background for this study, we first consider historical trends in our on-sequence general chemistry I course over the past 25 years, illustrated in Figures S1–S7 (the S prefix for figures and tables denotes material in the Supporting Information (SI)). Figure S1 shows the historical trend in DFW% for our general chemistry I course, which has steadily decreased from an average of ∼17% to ∼12%. Here data both for individual sections (open circles) and aggregate data by semester (filled squares) are shown. Sections taught by the author (S.A.R.) are shown here in red, but these are not the focus of the present study. As shown in Figure S2, the White, non-Hispanic population has decreased from ∼85% to ∼65% over this period, and this population is increasingly underrepresented in the DFW%. Here, as in all related figures used in this article, we compare the difference in DFW% for a given demographic to the overall % of that demographic in the considered population. Thus, a negative value signifies underrepresentation of that demographic group in the DFW population, a positive value represents overrepresentation. Importantly, these differences will sum to zero within each considered category.

The largest growth in demographic group in our courses has been in the Hispanic/Latin-X population. As shown in Figure S3, in general chemistry I, this population has grown from ∼3% to roughly 20%, following a university initiative to achieve Hispanic-Serving Institution (HSI) status. At the same time, however, this demographic is increasingly overrepresented in the DFW population. Figure S4 shows data for the first-generation population,
which has held steady over time at ~25% of the population but is increasingly overrepresented in the DFW population.

Considering gender, we fully recognize that nonbinary gender should be considered; however, the data set available to us only included records of binary gender and so we followed that in our analysis. Figure S5 shows the trend for women in general chemistry I, a demographic that has remained stable over time, with a near zero DFW% differential. Finally, in Figure S6, we show historical data for the mean percentile of ACS exam scores in general chemistry I, which have remained stable.

Moving to the focus of this study, we consider the author’s sections of CHEM 1002, broken out in Table S1, which corresponded to a total of 1502 students over eight sections. Section sizes were rather uniform, varying between 117 and 205, with a mean of 167. Complete demographic information was available for 1428 of these students, and a summary of the demographic data for these sections is provided in Table S2. Considering academic measures, we find that mean PQPA (predicted first year QPA, an internal Marquette metric available from 2005 onward, which we have previously identified as a key diagnostic of student success\textsuperscript{15,16}) and mean ACT score are relatively constant. The percentage of first-generation students is also relatively constant, at around 20%, while the percentage of women varied between 47% and 69%. Over this time, the class race/ethnicity demographic has significantly shifted, as the percentage of White, non-Hispanic students in our classes decreased from ~85% to ~65%. Figure S7 shows historical data for the mean percentile of ACS exam scores in CHEM 1002, which have remained stable.

Results and Discussion

Shown in Figure 1 are historical DFW% and W% for the author’s (S.A.R.) on-sequence (i.e., spring semester) sections of CHEM 1002. The figure includes a timeline representing the primary course modifications made over these years and when they were first introduced, the effects of which we will discuss in turn. These included, in historical order: (1) introduction of recitation sections, (2) introduction of classroom response systems (clickers), (3) introduction of online homework, (4) restructuring of course content into a modular format followed by a redesign of the course modules along a core or anchoring concepts scheme, (5) introduction of the flipped classroom, and (6) most recently, the introduction of regular low-stakes quizzing.

![Figure 1. DFW% (D’s, F’s, and Withdrawal) and W% for general chemistry II sections taught by one of the authors (S.A.R.). The sections average 166 students (see the SI for details), and vary between enrollments of 117 and 205. The timeline of notable modifications made to the course are indicated.](image-url)

1. Recitation Sections

Recitation sections have a long history in chemistry\textsuperscript{17} and other disciplines\textsuperscript{18−20}, going back more than 70 years.\textsuperscript{18} While they vary in format and implementation, such sections and related active learning approaches have...
positive demonstrated effects on student performance and success. At our institution, recitation sections were introduced into the CHEM 1002 course in the mid-1990s, just prior to our initial offering. These sections typically have student counts of around 30 and focus on small-group problem solving, led by a graduate or undergraduate teaching assistant. While attendance is voluntary, it typically is counted as a small part (i.e., 1%-2%) of the course grade. Over the years, the format of the recitation has moved from paper to clicker questions to web-based problem sets administered through the CHEM 101 platform.

Our experience is that recitation sections are well-regarded by students. Shown in Table S3 are the results of student evaluations of recitation sections in the author’s sections during the period that our Likert-scale based MOCES (Marquette Online Course Evaluation System) instrument has been in use. While TA specific questions, not shown, demonstrate some variance, questions concerning the sections themselves typically score in the “very good” range.

To examine the impact of these sections on DFW%, we must consider patterns of attendance. With the advent of clickers (below), we could more easily track student attendance, and our 2009 implementation was the first to incorporate clickers. Here, only lecture attendance was counted toward the grade, and the attendance rate of non-DFW students was 23% higher than for DFW students (83% vs 60%). Beginning in 2014, we began to count both lecture and recitation attendance as components of the grade (no more than 3% total). As for lecture, the recitation attendance rates of DFW students lagged significantly. For example, in our Spring 2015 lecture implementation, the recitation attendance rate of non-DFW students was 33% higher than for DFW students (86% vs 53%). In the most recent offering (Spring 2022), the difference was smaller, ~16% (85% vs 69%).

These findings point to an issue evidenced also in our study of the effects of a summer preparatory course, i.e., students who will likely benefit most from a given intervention are often less likely to participate. Thus, while we have found that counting attendance as a small component of the grade leads to higher attendance overall in both lecture and recitation, the attendance of DFW students still lags significantly behind.

2. Personal Response Systems (Clickers)

Across many fields, the use of personal or classroom response systems (“clickers” or CRS) has been successfully used by instructors to measure student learning, increase student engagement and develop problem solving skills. Studies generally show positive student response to clickers; however, the impact of clickers on student learning has shown mixed results. Consistent with these findings, a 2008 review of clicker use in the chemistry literature (56 publications) found generally positive student attitudes toward clickers, while measurable effects on student learning were found in some cases but not others. Importantly, studies reporting student learning gains included some form of collaboration.

In the chemistry community, surveys have revealed that the bulk of CRS adoptions occurred in large courses at the introductory level, and this has been true at our institution. In our general chemistry classes, clickers were first introduced in 2004, and several advantages were envisioned for both instructor and student. For the instructor, the CRS would provide real-time feedback on student understanding and the ability to review or revisit points of confusion. For the student, the CRS would provide real-time feedback on student understanding of lecture material, aiding in identifying points of confusion. Over the years, the format of response systems in our classes has moved from powerpoint-based clicker questions to web-based problem sets administered through the CHEM 101 platform. Considering the impact on DFW%, to benefit from this tool, students must attend and participate, and as discussed above, the classroom attendance for DFW students is significantly lower than their non-DFW counterparts. For example, in our most recent lecture implementation, the classroom attendance rate of non-DFW students was 84%, while that of the DFW students was 25%. Generally, we have found that counting lecture attendance for any part of the grade will push the overall attendance into the mid-80% range; however, this incentive apparently has less effect on DFW students.
3. Online Homework

In the early years, homework was not utilized formally in the course. Rather, students were assigned problems at the end of each chapter in the text and were told that they should complete these problems, at a minimum, to best prepare for exams. The instructor did not track student completion or performance. This system reflected the significant burden of manually graded paper homework sets and the lack of TA support for such grading.

With the advent of online homework systems, where assignments could be easily set up, assigned, and autograded, the role of homework in our general chemistry courses significantly expanded. Such systems were first introduced for purposes of formative assessment in 2009. Typically, homework sets were assigned weekly, with multiple attempts possible for each question, and these typically counted around 10% of the total course grade. Considering the impact on DFW%, in 2009, the average differential in homework score between non-DFW and DFW students was 39% (80% vs 41%). In subsequent offerings, this differential remained high, 32% (85% vs 53%) in 2015 and 56% (92% vs 36%) in 2022. Thus, the rate of homework completion for DFW students has significantly lagged that of their peers.

One measure of the effect of homework is the correlation with exam or quiz scores, as measured, e.g., with a Pearson correlation, which varies between values of −1 and +1.\(^52\) The correlation of homework with class assessments (exams, quizzes) in our classes has typically been modest, with Pearson correlations in the range 0.3–0.5, which are typical of those reported in the literature.\(^53\) However, this correlation, while not strong, is yet weaker for DFW students. For example, in 2009, the correlation of homework with hour exam scores was 0.41 for non-DFW students and 0.19 for DFW students—graphical representations of these correlations are shown in Figure S8. In 2015, the correlations were 0.38 and 0.09, respectively. As we typically give multiple attempts for the homework, reflecting its role as a formative assessment tool, this may reflect a “guess and check” approach, where students eventually obtain the correct answer without fully comprehending the problem. This issue is encapsulated in a recent request from one student in our classes: “I completed the homework by guessing, but I have no idea what is going on—can you help?”. This effect has been examined by Kortemeyer, who suggested that the number of attempts should be set around five, large enough to afford understanding and avoid cheating but small enough to discourage guessing.\(^54\)

There is ongoing debate in the literature regarding the effectiveness of online homework vs “paper and pencil” graded assignments;\(^55−58\) however, the positive impact of online homework on student performance (e.g., exam scores) has been shown in a number of studies.\(^53,59−61\) Interestingly, a controlled study of the effect of online homework on course grades in a precalculus class found that, while the exam scores in the treatment (homework) group were increased by roughly 15 points, the drop rate in the two sections (treatment and control) was similar. A recent report in general chemistry found that the majority of students receiving D and F grades completed less than 70% of the assigned homework assignments, while the majority of A students completed more than 90%.\(^62\) This differential is similar to that found here.

4. Course Restructuring

General chemistry courses often are criticized as being “miles wide and inches deep”.\(^63,64\) This is true of other STEM disciplines as well,\(^62,65−68\) and we have heard this criticism from our students, as illustrated in an evaluation comment from 2009: “This class is difficult and sometimes there really isn’t a flow of learning throughout the course, it’s a bunch of tiny pieces of information spliced together.” Chemistry educators have sought to address this through different avenues, including: (i) development of an “atoms-first” curriculum,\(^69,70\) (ii) variations in sequencing of general and organic chemistry,\(^71\) (iii) the “Chemistry Unbound” project,\(^72\) (iv) the CLUE (Chemistry, Life, Universe and Everything) and organic (OCLUE) projects,\(^73−76\) the Chemical Thinking curriculum,\(^77−81\) and (v) development of Anchoring Concepts Content Maps (ACCM) for general and other chemistry courses.\(^82−91\)
latter, a set of foundational or anchoring concepts, consistent across multiple courses, provides an overarching organizational structure, with differentiation occurring at the level of course detail. One advantage of the anchoring concepts framework, in our view, is that students are exposed to the same central concepts and questions across the discipline, which can efficiently link and build upon prior knowledge and reduce the fragmentation and compartmentalization of information.

Beginning in 2014, we sought to restructure our general chemistry courses around themes, which highlighted the key questions (“big ideas”) of the discipline. Initially, the CHEM 1002 course was reorganized into modules (Table S4)—these were not, however, significantly linked to larger themes. Subsequently, the modules were reorganized and tied to a set of anchoring concepts using the ACCM as a scaffold. This framework was developed in two successive summer offerings and refined in our general chemistry majors course sequence prior to being introduced to the larger classes. The current modular structure of our general chemistry courses is shown in Figure S9, with Table S5 providing a detailed set of learning objectives for each module. In this process, the laboratory curriculum was extensively revised to match the core learning outcomes.

At the beginning of each module, students are introduced to the core idea of that module, which is emphasized throughout the lecture and laboratory sessions. Although we have no direct evidence that these curricular changes have improved outcomes for DFW students, student response to these changes has been positive overall. For example, one student noted: “I also liked this module set up because the Equilibrium chapters, which is a majority of the course, was set up in a way that made logical sense. The organization of the course made it easier to follow the material, I was never confused about what was or was not on an exam.” Other comments followed along this line: “I enjoyed the organization of the topics. I think each topic helped build off the next, which aided in my comprehension of the material and CHEM 1002 as a whole.”

5. Flipped Classroom
Perhaps no change made to our CHEM 1002 course had as large of an impact on DFW% as the introduction of the flipped classroom. As with other modifications, the structure of the flipped class was developed in a small off-sequence (i.e., fall semester) CHEM 1002 class prior to implementation in the larger sections. As the results of these efforts have been published, we will only briefly recap them here. In our parallel controlled study, we compared students across sections using both a pretest and student percentile ranking in the preceding (general chemistry I) course. Our implementation of the flipped class incorporated short video lectures with paired homework sets and a weekly 75 min recitation section led by the instructor and a teaching assistant. The in-person meeting typically focused on problem solving but included occasional microlectures and demonstrations. Our study found a particular benefit for at-risk students, with a statistical difference in exam performance observed only for students in the bottom-third of the class, as identified by the pretest. This trend led to a significant decrease in DFW%. The correlation of homework completion with exam performance was stronger in the flipped section, particularly among the bottom demographic. It was suggested that this arose from the pairing of shorter homework assignments with each video lecture.

6. Low-Stakes Quizzing
The changes described thus far have been primarily pedagogical or curricular in nature. During the pandemic, we moved our focus to patterns of assessments. Historically, the bulk of assessments in our general chemistry courses were high-stakes exams (typically four in-semester exams worth 60% of the grade and a final worth 20%). Such an assessment pattern severely limits the instructor’s ability to identify and engage at-risk students early in the course. Moreover, there is abundant evidence that an assessment strategy based upon high-stakes examinations disproportionately impacts women. For example, in a study of a large ($N_{\text{students}} > 1000$) introductory biology course, Cotner and Ballen showed that women underperformed on high-stakes exams in comparison with men, even though such a discrepancy was not observed across other methods of
In classes where mixed assessment strategies were used and higher stakes assessments were minimized, this differential was significantly reduced.

Based upon these considerations, we decided to move from a pattern of higher stakes examinations to lower stakes and more frequent (typically weekly) quizzes, supplemented by a midterm and final exam. The quizzes and associated quiz library were developed over a summer session in 2020. Our initial findings demonstrated improved performance on similar questions in the quizzing environment, consistent with earlier results from other disciplines and with a theoretical framework centered around retrieval processes. In 2021–2022, we implemented the pattern of weekly quizzes (supplemented by a free-response midterm and final) in our on-sequence general chemistry courses. Together, the quizzes counted for 40% of the grade, with the free response midterm and final accounting for another 30%. The quizzes were administered via our course management system, using browser lockdown and webcam monitoring. The quizzes randomly drew from pools of questions, with typically 5–10 variants in each pool, and the order of the questions was also randomized. The majority of questions were multiple choice or multiselect.

To initially examine the effect of the quizzing pattern on DFW rates in the large lecture class, we conducted the classes in a lecture format, where lecture material was interspersed with sets of problems pushed out to students via the CHEM 1001 platform. Each class was attended by a graduate TA, and the instructor and TA walked around the room during problem solving sessions to assist students. In this larger implementation, quizzes were open during a window from Friday evening to Sunday evening, and students had one attempt to take the quiz. To protect the integrity of the quiz and quiz library, no feedback was initially provided upon quiz submission; however, later in the week, an alternate view opened that showed students the questions that they had missed. On Monday of each week, any student failing the quiz (50% or below) was invited to meet with the instructor and review the quiz in person. Following this meeting, the student was allowed to retake the quiz.

As example data, in our Fall 2021 general chemistry I course, a total of 54 quizzes were retaken, with a 40% retake rate for students initially failing the quiz. Students retaking the quiz improved their score on average by 32%. In our Spring 2022 general chemistry II course, a total of 58 quizzes were retaken, again with a 40% response rate for students initially failing the quiz. Here, students retaking the quiz improved their score on average by 30%. In both courses, the average improvement for students retaking the quiz was usually sufficient to move them into the range of a passing score.

Viewing the impact of this strategy of low stakes assessments with intervention, the DFW% in the Spring 2022 course was similar to that observed in the flipped courses taught in 2015 and 2016. Interestingly, the W% dropped to nearly zero, yet the overall DFW% remained similar. Thus, regular low stakes quizzing may aid in lowering DFW rates, even in a large lecture class. Supporting this are data from the Fall 2021 CHEM 1001 sections (Figure S1) where the DFW% in the author’s class was 7.7%, while the other three sections had DFW% of 20.6%, 14.7%, and 15.8%, respectively.

To dig deeper into the data, we ask whether the trend shown in Figure 1 is consistent across all demographic groups. Given the smaller DFW% in the more recent sections, and the lower diversity in the earlier sections, we decided to aggregate the data into two groups: 1999–2010 and 2011–present. A comparison of demographic data for students in these two groups in the author’s sections shown in Table S6 in the SI, while Table S7 compares the data across all sections for comparison. Consistent with the picture presented above, the mean academic measures (ACT score and PQPA) of these groups are similar.

We compared data across three demographic categories: (1) race/ethnicity, (2) binary gender, and (3) first-generation status. In each of the grouped data sets, we compare the percentage of each group in the DFW category to the percentage of that same group in the overall population. This is expressed as a difference, as
described above. It is important to emphasize again that, within each category, the sum of differentials will add to zero, which indeed we used as a check in our data analysis. Thus, if one subgroup within a demographic category is underrepresented in the DFW population, other group(s) within that category must be overrepresented.

As it is often difficult to generalize findings such as these across different instructors and institutions, we included in the comparison data from all other sections of the course during this period, grouped in the same manner. We note that not all reforms shown in Figure 1 have yet been adopted by all faculty, yet a number are now consistent across all sections of the course. These include recitation sections, classroom response systems, and online homework. A comparison of the DFW rates across all general chemistry II sections is shown in Figure S10. As found in the general chemistry I course, the DFW rate has somewhat steadily decreased over time, reaching a recent mean near 10%.

The demographic comparison among sections is shown in Figure 2. As our earlier classes primarily consisted of White, non-Hispanic students, when considering the race/ethnicity category, we restricted our analysis to this largest demographic subgroup. This population is historically underrepresented in the DFW population across all sections, a trend that has increased in the more recent data. However, in the author’s sections, an inverse trend is observed, suggesting that other demographics groups are now less overrepresented in the DFW population. Statistics for the author’s sections are too small to make definitive claims about which groups have benefited.

![Figure 2. Over- or underrepresentation of different groups in the DFW population. Here, data is aggregated over the years shown, and the x-axis reflects the difference in DFW% in comparison with the overall population of that group, as described in the text. Within each group, i.e., race/demographic, binary gender, and first generation, the differences sum to 0. Thus, e.g., considering binary gender, if women are underrepresented by 10% in the DFW population, men are overrepresented by 10%.

Considering binary gender, we find that the population of women is underrepresented in the DFW population overall, but more so in the author’s sections. This trend has increased in the more recent data, which may reflect the recent implementation of a mixed assessment strategy that emphasize low-stakes assessments. Finally, across all sections, first-generation students are overrepresented in the DFW population, and this difference has increased in recent years, as shown in Figure 3, which is a worrisome trend.
Given the importance of organic chemistry, it is also of interest to compare DFW rates in general chemistry II with those in the follow-up course, organic chemistry I, which is also a key gateway course at our institution. Despite calls for reform, organic faculty at our institution have resisted adopting similar changes in pedagogy, curriculum, or assessments. As pointed out by Cooper and co-workers, this is not uncommon, as approaches to the teaching of organic have changed little in recent decades. In Figure S11, we compare historical DFW rates from the organic chemistry I course with those of general chemistry II. Before 2014, the organic chemistry I DFW rates were uniformly lower than those observed in general chemistry II; however, in recent years, the organic chemistry I DFW rates have trended upward and now exceed those of general chemistry II.

This trend may be taken by the skeptical reader to suggest that improving student outcomes in general chemistry II has led students to failure in organic chemistry I. However, this is countered by the historical trends in DFW% in general chemistry I vs II—compare Figure S1 with Figure S10. As the DFW% in general chemistry I has decreased over time, a mirrored decrease in DFW% has occurred in general chemistry II, which we consider a more challenging course. Thus, improving student success in general chemistry I has not led to failure in the follow-on course. Rather, the trend shown in Figure S11 may highlight that, to be successful, reform must extend beyond the introductory course sequence. We are now focusing our attention on the organic curriculum and have developed a similar curriculum centered around core or anchoring concepts, which like the OCLUE approach of Cooper and co-workers emphasizes mechanistic reasoning with an early treatment of spectroscopy to develop connections between structure, stability, and reactivity. Our approach also incorporates changes in pedagogy and assessments, using many of the successful strategies identified here.

Finally, we recognize that instructors may be concerned regarding the effect of the various interventions described here on student evaluations of their courses. Thus, we examined trends in student evaluations of the courses. Shown in Table S8 in the SI are the means of questions drawn from our standard student evaluation instrument, which uses the indicated Likert scales. As can been seen, from 2009, when this evaluation instrument was first introduced, to 2022 there has been little variation in the student evaluations of these courses.

**Limitations**

There are two significant limitations to this work. First, historical trends limited the amount of data for certain demographic subgroups in the author’s sections. This is because, in the early years, student demographics were highly skewed toward White, non-Hispanic students, while in later years the lower DFW% was a limiting factor. Thus, conclusions that can be drawn regarding trends for certain subpopulations are limited. A second primary limitation is the length of time over which the study was conducted. The author’s courses were not static, but
neither was the student population nor the college experience, which has certainly changed over time. Where data is available, we find that academic measures such as ACT score, predicted first year QPA, and ACS exam scores are similar across this time period; however, we have applied no measures in the affective domain. Linking to student success, a variety of studies have demonstrated the influence of effective domain characteristics such as motivation, interest, self-concept, self-efficacy, and attitude. In particular, a strong link has been demonstrated between self-efficacy and academic success. In developing his Attitudes toward the Subject of Chemistry Inventory, Bauer noted that this dimension inherently embodies a range of factors such as attitude, self-concept, self-efficacy, beliefs and interests, and values, none of which we examined here. Very recently, we have incorporated affective domain assessments in our diagnostic exams; however, there is no historical data available.

Positionality Statement
To place this study in context, we briefly describe our own positionality. Reid is a White US-born scholar, a first-generation college student, and a late-career academic who served as the primary instructor for the courses described here. Vyas is a first-generation immigrant scholar who identifies as Asian and a middle-career academic who has served as general chemistry instructor and director of the laboratory program, with major contributions to curriculum, pedagogy, and assessment redesign.

Conclusions and Future Prospects
In this article, we have examined data on DFW rates in second-term general chemistry courses taught by one of the authors (S.A.R.) over a roughly quarter-century period. The data are discussed within a timeline of changes made to the course in terms of curriculum, pedagogy, and assessments, which included introduction of recitation sections, introduction of classroom response systems (clickers), introduction of online homework, restructuring of course content into a modular format followed by a redesign of the course modules along a core or anchoring concepts scheme, introduction of the flipped classroom, and introduction of regular low-stakes quizzing. We find that the lowest DFW rates, roughly 5%, are found in flipped classroom sections or in a large lecture section that incorporated low-stakes assessment with instructor intervention for at-risk students. Our findings suggest that a combination of active learning pedagogy, core concepts curricula, and incorporation of low-stakes assessments is a strategy capable of moving the needle to improve DFW rates in second-term general chemistry.

Recently, Hayes and Randall reported on a successful lowering of DFW rates in general chemistry using a combination of factors which included: (a) a math placement exam, (b) a class meeting pattern of 5 days per week, which included both “lecture” and “recitation” components, and (c) frequent homework assignments. They find DFW rates that average around 5% in second-term general chemistry and suggest that this may represent a “natural” limit that reflects a baseline level of student attrition. While this baseline might be expected to vary significantly across institutions, it is noteworthy that our “baseline” value is similar to theirs.

Considering future prospects, it will be important to continue to refine our assessments, using, the 3D-LAP protocol developed by Cooper. In our most recent implementation, a significant fraction of questions on the midterm and final exams addressed all three dimensions of this protocol, and we are working to revise our selected response question library. To better understand the role of the affective domain in student success, we also are implementing student attitude and self-concept measures using the ASCIv2 instrument and the CSCI (Chemistry Self-concept Inventory). Overall, there remains much work to be done in improving student outcomes for all students in general chemistry I/II, which are key gateway STEM courses.
Supporting Information
The Supporting Information is available at https://pubs.acs.org/doi/10.1021/acs.jchemed.2c01121.

- Figures of DFW% across all sections of the on-semester CHEM 1001 and 1002 courses, population and DFW% for White, non-Hispanic students, Hispanic and Latin-X students, first-generation students, and women in the on-semester CHEM 1001 course from 1998 to 2021, ACS exam scores, graphical representations of the correlation between homework score and hour exam average for the Spring 2009 course, current course modules for CHEM 1001/1002, and comparison of DFW rates in organic chemistry 1 and second-term general chemistry by semester and tables of enrollments in on-semester CHEM 1002 courses, demographic data for students in the authors sections, results of recitation section evaluations from the authors sections, initial course modules from Fall 2014, module learning outcomes for most recent CHEM 1002 implementation, demographic data for students in the grouped sections of the author, demographic data for students in all grouped sections, and summary of student evaluations (PDF, DOCX)

- ed2c01121_si_001.pdf (1.35 MB)
- ed2c01121_si_002.docx (5.53 MB)

Terms & Conditions
Most electronic Supporting Information files are available without a subscription to ACS Web Editions. Such files may be downloaded by article for research use (if there is a public use license linked to the relevant article, that license may permit other uses). Permission may be obtained from ACS for other uses through requests via the RightsLink permission system: http://pubs.acs.org/page/copyright/permissions.html.

Acknowledgments
The authors gratefully acknowledge the assistance of Laura MacBride (Office of Institutional Research) and Seth Zlotocha (Office of the Registrar) in obtaining the data used in this study. Helpful conversations with Barbara Reisner and Adam Fiedler are gratefully acknowledged.

References
5. Hatfield, N.; Brown, N.; Topaz, C. M. Do introductory courses disproportionately drive minoritized students out of STEM pathways?. PNAS Nexus 2022, 1 (4), pgac167, DOI: 10.1093/pnasnexus/pgac167


42. Nielsen, K. L.; Hansen, G.; Stav, J. B. Teaching with student response systems (SRS): teacher-centric aspects that can negatively affect students’ experience of using SRS. *Res. Learn Technol.* 2013, 21 (1), 18989, DOI: 10.3402/rlt.v21i0.18989


53. Arasasingham, R. D.; Martorell, I.; McIntire, T. M. Online Homework and Student Achievement in a Large Enrollment Introductory Science Course. *Journal of College Science Teaching* 2011, 40 (5), 70–79


78. Talanquer, V. DBER and STEM Education Reform: Are We Up to the Challenge?. *J. Res. Sci. Teach* **2014**, *51* (6), 809–819, DOI: 10.1002/tea.21162


120. Zusho, A.; Pintrich, P. R.; Coppola, B. Skill and will: the role of motivation and cognition in the learning of college chemistry. *Int. J. Sci. Educ.* 2003, 25 (9), 1081–1094, DOI: 10.1080/0950069032000052207


