

## SCIENTIFIC COMMUNITY

# Reducing achievement gaps in undergraduate general chemistry could lift underrepresented students into a “hyperpersistent zone”

R. B. Harris<sup>1\*</sup>, M.R. Mack<sup>2</sup>, J. Bryant<sup>2†</sup>, E. J. Theobald<sup>1</sup>, S. Freeman<sup>1‡</sup>

Students from underrepresented groups start college with the same level of interest in STEM majors as their peers, but leave STEM at higher rates. We tested the hypothesis that low grades in general chemistry contribute to this “weeding,” using records from 25,768 students. In the first course of a general chemistry series, grade gaps based on binary gender, race/ethnicity, socioeconomic status, and family education background ranged from 0.12 to 0.54 on a four-point scale. Gaps persisted when the analysis controlled for academic preparation, indicating that students from underrepresented groups underperformed relative to their capability. Underrepresented students were less likely than well-represented peers to persist in chemistry if they performed below a C–, but more likely to persist if they got a C or better. This “hyperpersistent zone” suggests that reducing achievement gaps could have a disproportionately large impact on efforts to achieve equity in STEM majors and professions.

## INTRODUCTION

Achievement gaps between well-represented and underrepresented students have been called “one of the most urgent and intractable problems in higher education” [(1), p. 99], and are increasingly recognized as an international issue (2). Grade gaps are particularly prominent in undergraduate science, technology, engineering, and mathematics (STEM) courses (3). In these disciplines, women and underrepresented minorities (URMs) actually underperform, on average, relative to well-represented peers with the same academic preparation (4, 5).

Achievement gaps are important because underrepresentation in STEM majors results from disproportionately high attrition—not from lack of interest. For example, the percentage of URM and non-URM students who enter U.S. colleges intending to complete a STEM major is about the same (6), but 6-year STEM-completion rates vary from 52% for Asian-Americans and 43% for Caucasians to 22% for African-Americans, 29% for Latinos/Latinas, and 25% for Native Americans (6). Although gendered bias in attrition is less severe, it exists: 52% of women and 48% of men who enter U.S. colleges intend to major in STEM fields, but 6-year completion rates for these STEM-interested students are 38% for women and 43% for men (6, 7).

Why do students drop out of STEM majors? For the overall student population, poor performance in first-year STEM courses is negatively correlated with persistence in STEM (8). But research has yet to link achievement gaps in a specific introductory course with the disproportionately high attrition from STEM majors observed for female, URM, and low-socioeconomic status (low-SES) students. Establishing an association is important for policy makers, because reducing the attrition of STEM-interested underrepresented students and increasing their recruitment into STEM majors could

help solve three major issues: (i) the need to supply qualified professionals to an increasingly STEM-dependent economy (9), (ii) maximizing the impact of diverse groups in solving particularly complex problems (10), and (iii) promoting economic mobility with the goal of reducing class distinctions (11).

In this study, we focus on achievement gaps in a particularly critical course: general chemistry. General chemistry is a year-long course sequence that most STEM-interested students begin in the first fall of their first year of college. It functions as a gateway or gatekeeper because it is required for many STEM majors, including virtually all of those offered in the life sciences and most in engineering, and has been shown to have an especially large impact on students who are interested in careers in medicine, dentistry, or pharmacy. For example, studies that followed cohorts of talented URM students who entered college on a premedical track found that for the individuals who abandoned that ambition, poor performance in general chemistry was the most important factor driving their decision (12).

The goal of this study was to test what we term the GenChem Hypothesis: the claim that achievement gaps in the first course of general chemistry have disproportionately large impacts on the attrition of underrepresented students from the STEM track. Although most of this work focused on general chemistry (GenChem), we included analyses of organic chemistry (OChem) as well because it represents a second year-long sequence required of students on the pre-health professional track.

We collected and analyzed data on final grades in GenChem and OChem courses offered at the University of Washington from 2001 to 2016. This institution is on the quarter system, so the complete, core chemistry series consists of GenChem 1, 2, and 3 and OChem 1, 2, and 3. In analyzing persistence in the general chemistry series, we excluded students in engineering programs that require only the initial course in the general chemistry series. The final dataset included 75,759 records from 25,768 unique students. We disaggregated these data to look at four possible types of achievement gaps: women and men, URMs and non-URMs, low-SES and higher-SES students, and first- and continuing-generation students. The data on binary gender, race/ethnicity, and first-generation status are self-reported and archived by the University of Washington upon

Copyright © 2020  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>Department of Biology, University of Washington, Box 355320, Seattle, WA 98195-5320, USA. <sup>2</sup>Department of Chemistry, University of Washington, Box 351700, Seattle, WA 98195-1700, USA.

\*Present address: Adaptive Biotechnologies, 1551 Eastlake Ave E, Seattle, WA 98102, USA.

†Present address: Department of Chemistry, University of Southern California, 3620 McClintock Avenue, Los Angeles, CA 90089-1062, USA.

‡Corresponding author. Email: srf991@uw.edu

admission. We defined individuals who self-identified as African-American, Latino/Latina, Native American, Native Hawaiian, or Pacific Islander in terms of either race or ethnicity as URM, and students who self-identified as Caucasian, Asian-American, or International as non-URM. We followed the literature in defining first-generation students as those who self-report that they do not have a parent who has completed a 4-year degree (8). Low-SES students were identified by admission to the university's Educational Opportunity Program (EOP), which serves economically and educationally disadvantaged individuals based on family income data and high school attended. Because our goal was to analyze broad patterns in STEM persistence for students from underrepresented groups, we treated the data on gender, URM status, SES status, and first-generation status as binary categories and did not explore issues related to intersectionality—for example, how the combination of gender and URM status affects persistence in the STEM-major track. During our study period, the average demographic profile of students in the initial general chemistry course was 52.3% female, 10.6% URM, 19.6% EOP, and 38.7% first-generation.

As a proxy for prior student performance and preparation, we gathered data on college entrance exams (SAT verbal and math scores, summed, or converted ACT scores) and high-school grade-point averages (GPAs). Last, we collected data on an array of instructor characteristics, including rank, binary gender, and student evaluation of teaching (SET) scores (table S1).

These data allowed us to address three broad questions.

1) How large are achievement gaps in chemistry? We initially calculated raw or “transcriptable” gaps throughout the general chemistry and organic chemistry series, as they reflect the grades that students experience and that are evaluated by employers and professional and graduate schools. We then modeled gaps that were controlled for indices of academic preparation and ability. If these modeled gaps are nonzero, then it suggests that the affected students are underperforming relative to peers with equivalent qualifications. Data on raw gaps document the student experience by quantifying differences in actual grades received; data on modeled gaps test the hypothesis that raw gaps in undergraduate STEM courses result from differences in preparation.

2) Do instructor characteristics predict grade gaps? In the literature, researchers have focused on instructor quality and instructor identity as potential predictors of student performance (13, 14). We tested the association between grade gaps and (i) instructor rank, receipt of a teaching award, and scores from SETs as traditional indices of quality and (ii) gender as an aspect of instructor identity.

3) Do gaps have consequences? Specifically, do the data support the GenChem Hypothesis' claim that the initial general chemistry course is a gatekeeper that “weeds out” diversity? To explore whether achievement gaps and attrition from the STEM track are linked, we analyzed (i) what percentage of the students who started GenChem 1 went on to each subsequent course in the general chemistry and organic chemistry series; (ii) failure rates for underrepresented versus well-represented students in the initial general chemistry course, measured as a D or F grade or withdrawal (DFW rate); and (iii) the probability that an underrepresented versus well-represented individual retook the initial course in the general chemistry series, dropped out of the chemistry sequence completely after the first course, or went on to the next course in the general chemistry sequence, both as a function of that individual's grade in the initial course and as a function of that individual's grade, SAT score, and high school GPA.

## RESULTS

### How large are gaps?

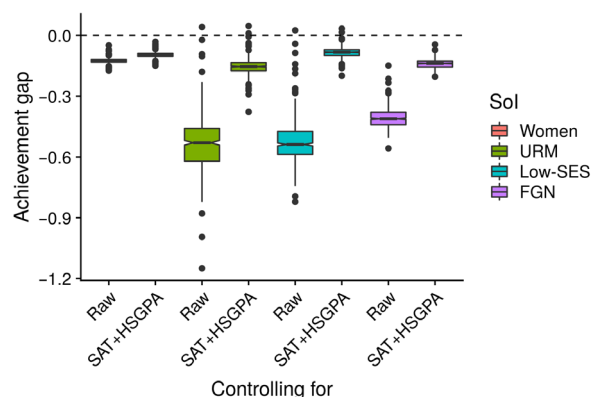
In terms of raw gaps, URMs received about 0.54 grade points less on average than non-URMs [95% confidence interval (CI),  $-0.81$  to  $-0.16$ ; Fig. 1]. In models that adjusted for variation in academic preparation by including SAT scores and high school GPA as predictors, the difference between URM and non-URM students narrowed to 0.16 grade points (95% CI,  $-0.28$  to  $-0.04$ ), a 70% gap reduction. The persistence of the modeled gap, however, indicates that URM students are underperforming in general chemistry relative to non-URMs matched in terms of academic preparation. Raw gaps for the other student subgroups range from 0.12 to 0.51 (Fig. 1 and table S2). In models that control for preparation, gaps persist for all four subgroups but are slightly smaller for female, low-SES, and first-generation students compared to URMs (Fig. 1 and table S2).

In organic chemistry, women experience much larger raw grade gaps than in the first two courses of general chemistry, indicating that women perform less well in organic chemistry, relative to men, than they do in general chemistry (fig. S1). We also documented consistently small differences between raw and modeled gaps for women in each of the six courses we analyzed (fig. S1), suggesting that overall academic preparation is less important in explaining how women perform in general and organic chemistry than it is in other student subgroups.

In contrast to the pattern for women in the five courses subsequent to general chemistry, URM, low-SES, and first-generation students continue to experience raw achievement gaps similar to those observed in GenChem 1. One notable exception to this pattern is a reduction in raw gaps for URM students in OChem 2 and especially OChem 3 (Fig. 1 and fig. S1). Although gaps that control for academic preparation are relatively consistent for first-generation students in the subsequent five courses, they are eliminated or even reversed for URM students in organic chemistry and for low-SES students in all five subsequent courses.

### Do instructor characteristics predict gaps?

Achievement gaps for all four subgroups of students did not vary as a function of instructor rank. In contrast, modeled achievement gaps for URM students were smaller if instructors got higher scores on



**Fig. 1. Achievement gaps by student subgroup in GenChem 1.** “Raw” indicates actual grades; “SAT+HSGPA” indicates estimated grades controlling for SAT scores and high school GPA. The black line at the center of each boxplot indicates the median, with the notch displaying the 95% CI around the median. “Sol” stands for “students of interest.”

SETs but larger if instructors were female or had received a teaching award. Teaching award winners also had larger achievement gaps for low-SES and first-generation students. We found no association between year and the size of grade gaps for any subpopulation (see the Supplementary Materials).

Although some of these predictors were associated with low  $P$  values, in all cases, the coefficients were small, meaning that the effect size was almost inconsequential (table S3). The size of the gaps reported in Fig. 1 were also remarkably stable over time, as well as among instructors (table S3). Thus, the instructor characteristics we analyzed had a negligible impact on the size of raw gaps and on the level of underperformance.

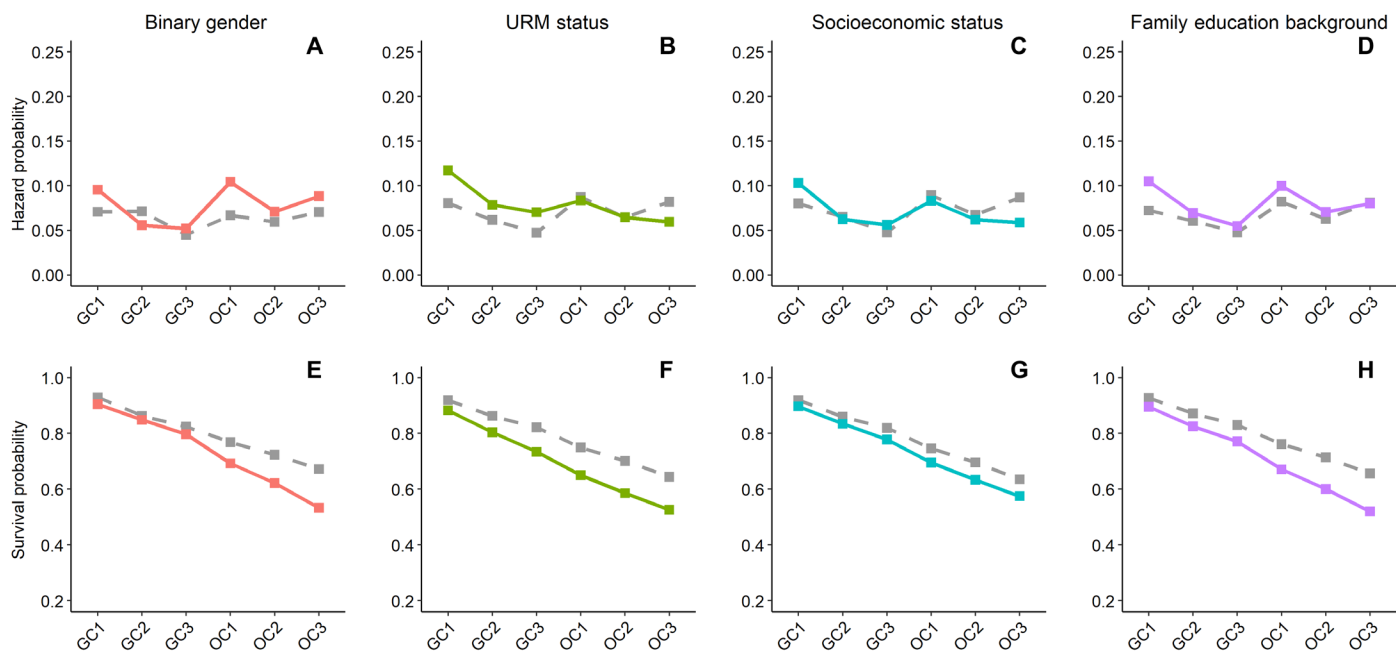
### Do gaps have consequences?

To evaluate the consequences of the observed achievement gaps, we began by examining attrition over the course of the general chemistry and organic chemistry series for each subgroup of students. Specifically, we used discrete-time logit-hazard models to estimate the risk of not advancing at each point in the curriculum and then to quantify the cumulative effects of attrition over time for each student subpopulation. For all four student subgroups, the hazard for not continuing was highest in the first general chemistry course, with female, low-SES, and first-generation students experiencing a second peak in risk in the first organic chemistry course (Fig. 2, fig. S2, and table S4). Although the general trend was for this risk of not advancing to decrease over time, in almost every course, the hazard was higher for underrepresented versus well-represented students, even controlling for indices of academic preparation. Attrition, measured as the survival probability over time, also differed sharply between all four categories of well-represented versus underrepresented

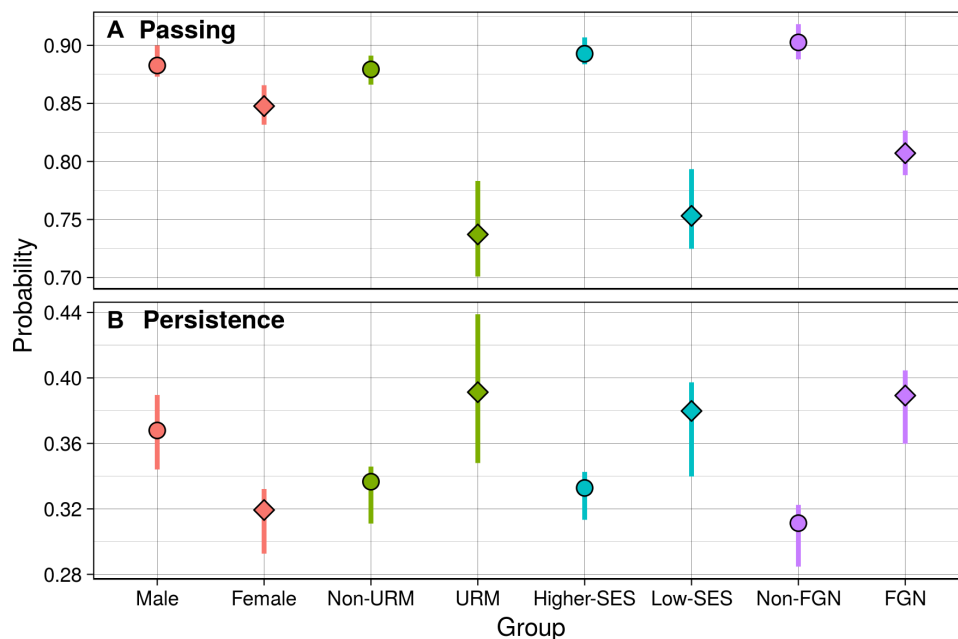
students and was most severe for URM, low-SES, and first-generation students (Fig. 2). These general patterns were even more pronounced when we ran the analysis without controlling for indices of academic preparation (fig. S2).

To test the GenChem Hypothesis more explicitly, we calculated (i) the probability of passing the initial general chemistry course, defined as completing and getting a grade of 1.7 or above, and (ii) the probability of a student continuing to the second general chemistry course if they had achieved a grade of 1.7 or above in the initial course and were eligible to progress in the series. We found that students from all four subgroups were more likely to fail than their well-represented counterparts and that women who passed the course were less likely to continue to the next course in the series than men (Fig. 3). These results suggest that grades in GenChem 1 make a major contribution to the attrition of underrepresented students in STEM majors.

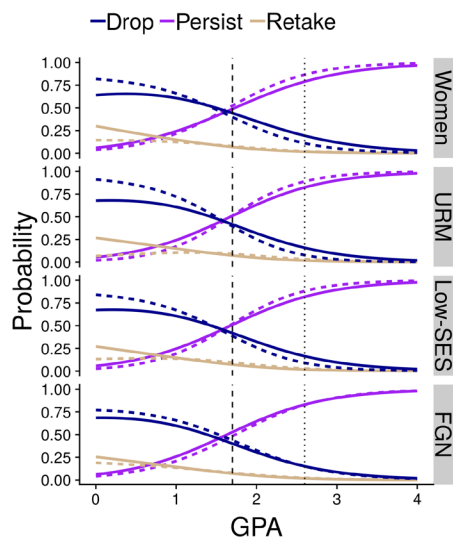
A more granular analysis indicates a much more nuanced pattern, however. We modeled the probability of three outcomes for female and male, URM and non-URM, low-SES and higher-SES, and first- and continuing-generation students as a function of their grade in GenChem 1. Did they retake the initial course? Did they drop out of chemistry and/or the university altogether—meaning that they never again appear in the dataset? Or did they persist—meaning that they took GenChem 2 at some point in the study period? In general, the data indicate the same pattern for all four subgroups of students (Fig. 4). If female, URM, low-SES, or first-generation students receive low grades in the initial course, then they are less likely than peers who received the same GenChem 1 grade and had the same indices of academic preparation to retake the initial course, more likely to drop out of the dataset, and less



**Fig. 2. Risk of not continuing in chemistry and attrition over time, by course and student subgroup, controlled for academic preparation.** (A to D) Proportion of students at the beginning of each general chemistry (GC) or organic chemistry (OC) course who did not advance to the subsequent course, controlled for indices of academic preparation. (E to H) Proportion of students who started general chemistry at the University of Washington and were retained at the end of each course, controlled for indices of academic preparation. In all of these graphs, underrepresented students (e.g., women) are represented by lines in color and well-represented students (e.g., men) by lines in grey.



**Fig. 3. Probability of passing and persisting in general chemistry as a function of student subgroup.** (A) Passing rate, calculated as  $1 - \text{DFW rate}$ , with DFW defined as either a grade  $< 1.7$  or withdrawal from GenChem 1 (at this institution, a 1.7 is considered a C-). (B) Persistence is defined as going on to GenChem 2 for students who took GenChem 1 and were eligible to go on in the series (received a grade of 1.7 or higher). In both graphs, vertical bars indicate SEs derived using bootstraps (number of replicates = 100; number of students per replicate = 5000). “FGN” indicates first-generation students.



**Fig. 4. Consequences of GenChem 1 grades, controlled for indices of academic preparation, for four student subgroups.** The vertical line at 1.7 indicates a threshold, below which a student is not allowed to move on to the next course in the series. The vertical line at 2.6 shows the “fixed” median grade used by the chemistry department to put scores from different sections on the same scale. A dashed line represents the underrepresented group identified on the right margin; a solid line represents the relevant comparison group (e.g., men in the top panel). “Drop” indicates students who took GenChem 1 but did not reappear in the dataset over the study period. “Retake” indicates students who took GenChem 1 again. “Persist” indicates students who took GenChem 2 during the study period. The estimates are from models that included SAT scores and high school GPA as covariates.

likely to persist. But if female, URM, or low-SES students pass the course with a grade of about 2.0 (C) or better, then they are less likely to drop out of the dataset than peers who received the same grade and more likely to take the second course. For each student subgroup, the sigmoidal relationship between grade and persistence has an inflection point just below the level of a C. The patterns for gaps that do not control for SAT and high school GPA are similar (fig. S3 and table S5).

## DISCUSSION

One of the most important results from our analysis is establishing a strong connection between grades in general chemistry and attrition from a course sequence required to continue in STEM. Students in all four underrepresented subpopulations have a higher probability of not progressing to the second course in the series than their well-represented peers. On the basis of the data in Figs. 3 and 4, this result appears to be due to students in all four subgroups having a disproportionately high probability of failing outright and/or entering the “drop” status after poor performance in the initial general chemistry course. Students who leave the introductory chemistry series are effectively prevented from pursuing a STEM major unless they complete the general chemistry series at a different institution. In addition, the hazard and survivorship data graphed in Fig. 2 indicate a disproportionately large impact of GenChem 1. Together, these observations support the GenChem Hypothesis: Poor performance in the initial general chemistry course is correlated with attrition of STEM-interested but underrepresented students.

A second and equally notable result is more optimistic. For women, URM, and low-SES students, the persistence data graphed in Fig. 4



show an important pattern: The responses of underrepresented and well-represented students switch at the sigmoidal curve's inflection point. Depending on the subgroup, this inflection point corresponds to a grade of 1.7 to 2.0 on a four-point decimal scale—with a 1.7 being the threshold for being allowed to go on in the series—or C− to C on the A to F scale. Thus, after poor performance—which we define as getting a grade of C− or worse—underrepresented students are more likely to leave the STEM-major track than well-represented peers who receive the same grade. But after adequate or good performance—which we define as getting the equivalent of a C or above—female, URM, and low-SES students are more likely to persist in the STEM-major track than their well-represented peers with the same grade. These results suggest that performing at the level of a C or above results in female, URM, and low-SES students who are “hyperpersistent” compared to their peers, whether or not they are matched for preparation and ability. This is important, because hyperpersistence is required for female, racial and ethnic minority, low-income, and first-generation individuals to reduce their underrepresentation in STEM majors and professions relative to well-represented and over-represented groups. “Hypersistence,” in contrast, causes underrepresentation.

The “hyperpersistence zone” documented here is consistent with previous studies indicating that URM students may be less grade-sensitive and more tenacious than non-URM students in general, and especially so if they are on the premedical track (5). Work on student affect suggests that this grittiness—defined as perseverance and passion in the pursuit of long-term goals—may spring from differences in motivation, with URM and first-generation students more likely to be driven by a commitment to help their families and broader communities than their well-represented peers (15, 16). The observation that URM and low-SES students are more likely to re-take the course after receiving poor or average grades is also consistent with the grittiness hypothesis.

Other results reported here reinforce broad patterns already established in the literature. The large raw gaps in general chemistry grades reported in Fig. 1 have been observed before for URM, low-SES, and female students (3–5), and the attrition results in Fig. 2 reflect data that have been reported in aggregate at the national level (6). The attrition results reported here, however, include a previously unobserved second “risk spike” during the initial organic chemistry course. The hazard posed at the start of organic chemistry may be particularly important for women, as they perform worse in organic chemistry relative to men than they do in general chemistry (see the Supplementary Materials).

Our data also confirm that women and URM students are underperforming in undergraduate STEM courses relative to their academic preparation and ability (4, 5). Here, we document the same trend for low-SES and first-generation students in general chemistry. These observations suggest that something about undergraduate STEM courses, beyond differences in preparation, is having a strong negative impact on underrepresented students.

What drives the underperformance observed in all four subgroups? We propose that sensitivity to evaluative situations plays a major role. For example, women tend to underperform relative to men on high-stakes STEM course exams, even though they outperform men on lower-stakes non-exam points (4). Women are also more grade sensitive in calculus and economics, meaning that they are less likely to take subsequent courses when achieving grades identical to those of male peers (17, 18). Low-SES students in STEM have higher “rejection sensitivity,” on average, than their higher-

SES peers (19). Among students from all demographic groups who intend to major in chemistry but later leave, researchers have documented higher performance-avoidance orientation and lower self-efficacy (20). Sensitivity to evaluative situations may be a shared characteristic of all four underrepresented subgroups analyzed here.

Why would underrepresented students be more sensitive to evaluative situations than their well-represented peers? The literature offers two hypotheses. The first focuses on self-efficacy—an individual's belief in their ability to succeed (21). Students with low self-efficacy do poorly in chemistry and other undergraduate STEM courses (17, 20, 22), and during a chemistry course, self-efficacy can show disproportionately large declines in URM students (23). The second hypothesis focuses on stereotype threat, which causes underperformance in evaluative situations due to the cognitive demands of coping with negative stereotypes about one's gender or race (24). In support of the second hypothesis, a values affirmation intervention that is designed to alleviate stereotype threat (25) has reduced achievement gaps for women in undergraduate physics (26) and for URM students in undergraduate biology (27).

In response to data like these, researchers are emphasizing the importance of designing general chemistry and other key courses for inclusion. This call focuses on the hypothesis that synergistic effects occur when an improved classroom culture is combined with elements of deliberate practice. More specifically, the hypothesis is that courses would better support underrepresented students if they encouraged belonging, science identity, and self-efficacy; emphasized active learning approaches that engage all students and increase exam scores and lower failure rates; and deemphasized inauthentic assessments such as high-stakes exams (2, 4, 28, 29, 30). Unfortunately, recent research across North American universities has documented that traditional lecturing still dominates in undergraduate STEM courses, with chemistry courses ranking as the most didactic of all the disciplines studied (31). The data on current practice are discouraging, given that calls for comprehensive reform of the general chemistry curriculum began over 65 years ago (32). A strictly didactic style dominated in the courses studied here, which may explain why we failed to observe either changes in achievement gaps over time or noteworthy effect sizes due to differences in instructor rank, gender, or teaching award status: These types of observable personal characteristics in instructors may be much less important than how faculty teach. The similarly small effect size observed based on SET scores is consistent with a recent meta-analysis showing that SETs have little or no correlation with measures of student learning (33).

Recent experiments suggest that intensive active learning and a more-inclusive classroom culture in general chemistry may lead to higher performance by low-SES and URM students (34, 35). However, the literature still lacks an example of a revised course design in general chemistry that results in reduced or no achievement gaps for female, URM, low-SES, and/or first-generation students—in terms of either raw gaps or gaps controlled for indices of academic preparation.

It is important to recognize that the results reported here are from a single institution. Before the GenChem Hypothesis and the existence of a hyperpersistence zone can be considered general features of STEM education, the relationship between general chemistry grades and retention needs to be evaluated in other programs—especially colleges and universities that differ in terms of the demographic makeup of their student populations. In addition, our data

reflect a period when all teaching in general chemistry and organic chemistry at the focal institution conformed to traditional lecturing. We do not yet know whether the performance sensitivity that drives the GenChem Hypothesis can be quantified at other institutions or whether the hyperpersistence zone documented here also occurs in course designs that include active learning. Last, retrospective studies like ours cannot test causality. A direct test of the GenChem Hypothesis might start with interventions that reduced achievement gaps, and test the prediction that subsequent attrition declined.

In terms of moving discipline-based educational research forward, this study supports calls for researchers to disaggregate data on student success and consider how specific subgroups of students are performing in terms of grades, affect, and attainment (36). These calls have two goals: making an “invisible” problem visible to instructors and administrators, and motivating changes in current practice with evidence (1, 12). For example, our data offer faculty, administrators, and policymakers an important prospect: Because of the “switched-sigmoidal” relationship in persistence documented in Fig. 4, small improvements in the course performance of underrepresented students could produce disproportionately large increases in persistence. Although reductions in achievement gaps could accelerate this effect, it would occur even if innovations in course design benefited all students equally in terms of increased grades. This impact would occur because more underrepresented students than well-represented students would rise above the inflection point into the hyperpersistent region of the curve. If so, then undergraduate STEM programs and the health professions may see increasing benefits from the resilience, cultural insights, linguistic fluency, and other assets that underrepresented students can bring (16).

## MATERIALS AND METHODS

### Experimental design

The goals of this study were to explore patterns in achievement gaps that affect female, URM, low-SES, and first-generation students and to create a baseline for a planned series of experiments at the University of Washington, focused on evaluating the impact of evidence-based course designs in general chemistry and organic chemistry. The specific objectives of the work were to (i) evaluate the impact of instructor characteristics, including rank and gender, on achievement gaps; (ii) test the hypothesis that achievement gaps in college chemistry are a continuation of gaps in academic preparation, as indexed by SAT score and high school GPA; and (iii) document the impact of achievement gaps on retention of underrepresented students in STEM majors that require completing the general chemistry series. The pre-specified outcome variables were final course grade and DFW rate—each associated with student demographic data—and retention in general chemistry, indexed as the probability of continuing to the next course in the introductory series. The study was conducted under application 00001169 to the University of Washington Institutional Review Board, which approved the work and waived informed consent.

### Data collection

We obtained data from the University of Washington Registrar’s office for all students enrolled in general chemistry and organic chemistry during the 2000–2016 academic years. The information retrieved included chemistry course enrollment and final grade, overall college GPA, matriculation year, graduating major(s), high

school GPA, college entrance exam scores, binary gender, race and ethnicity, family income, parental education attainment, and EOP status (table S1). We obtained data on instructor rank, binary gender, SETs, and teaching award status from the University of Washington Chemistry Department records.

### Data coding and filtering

We coded a student as follows: (i) URM if citizenship was in the United States and either ethnicity or race was indicated as black, Latino/Latina, Native American, or Native Hawaiian/Pacific Islander; (ii) low-SES if enrolled in the University of Washington EOP; and (iii) first-generation (FGN) if neither parent had completed a 4-year degree. We used high school GPA and college entrance exam scores as indicators of academic preparation and achievement. Using College Board concordance tables, we converted ACT scores to corresponding SAT scores. To correct for any year-over-year changes in entrance exam scores, we centered both high school GPA and SAT by matriculation year.

Course grades at the University of Washington are recorded as W for withdrawal and 0.0 for failing, and then in 0.1-point intervals between 0.7 and 4.0. Students who complete the course but receive a grade less than 1.7 are prohibited from continuing to the next general chemistry course in the series. We also created a categorical variable that characterized a student’s trajectory following enrollment in the initial general chemistry course as (i) persisted in the chemistry series, meaning that the student received a grade in the initial course and subsequently enrolled in the next course; (ii) retook the initial course; or (iii) left the University of Washington chemistry curriculum altogether. In analyzing other aspects of course performance, we considered only a student’s first attempt at the course in question.

Before analysis, we removed any student with a major that only required the initial course in the general chemistry series, such as all engineering majors except bioengineering and chemical engineering. This step allowed us to infer that all students in the initial general chemistry course intended to complete a STEM major that required the full general chemistry sequence. We also removed students with credit/no credit grades instead of a numerical grade and students with hardship withdrawals, which result from extraordinary circumstances such as family or medical emergencies.

### Statistical analysis

We developed predictive models of course grade, risk and survivorship, failure rates, and next-step decisions as a function of student demographic variables and academic preparation. We began by calculating the raw, uncorrected relationship between each student subgroup and final numeric grade, over 70% of which is determined by performance on high-stakes exams. The Department of Chemistry has a policy of curving final grades to a common, mean GPA of  $2.6 \pm 0.2$  to adjust for variation across instructors and academic quarters. To control for additional variation across sections, we standardized final grades by section using *z*-scores.

We developed a linear mixed-effect model of final grades to account for the variation that can be attributed to students versus those that are a product of variation through time. Mixed-effect models partition analyses according to fixed and random effects: Fixed effects refer to variables that are explicitly tested (e.g., URM status), while random effects refer to variables that are not necessarily of interest (e.g., section, term, and year effects) (37).

We conducted model selection in four phases to identify which fixed and random effects best explained the variation in final grades (38). First, we designated full models using all variables of potential interest. To choose the appropriate random-effects structure, we calculated the restricted maximum likelihood for all possible combinations of random effects while holding fixed effects constant, and ranked those models according to the adjusted Akaike Informative Criterion (AICc). We considered the best-fit random-effect structure to be that which contained the smallest number of parameters while being within two points of the lowest AICc score. In the third step, we used maximum likelihood estimation to compare the best-fit random-effects model to the same model with all random effects removed, and chose the model with the lowest AICc score. In the final step, we took the model selected in the third step and conducted backwards elimination, using a restricted maximum likelihood criterion, to determine which fixed effects should be included in the final model. We removed predictors from the model until all individual predictors reached significance at the 0.05 threshold for type I error. Model selection results for each outcome are provided in tables S3 and S6.

Grade gaps for each model were calculated at the section level by summing model coefficients and then subtracting the predicted course grade for each reference group from the grade for each subgroup. For each model, this resulted in a distribution of 125 gaps corresponding to the 125 sections of GenChem 1 sampled between 2001 and 2016. To obtain units of actual grade points, we rescaled the grade difference using the section-specific SD in final grade

$$\text{gap} = (\text{grade}[\text{Group} = 1] - \text{grade}[\text{Group} = 0]) * \sigma_{\text{section}}$$

When plotting these achievement gaps, we accounted for weighting. Model estimates may be biased by sample size because student composition varied across sections, quarters, and years. For example, in our dataset, the proportion of URM students in each section ranged from 0 to 25%. A section without URM students would deceptively appear to have no achievement gap. To correct for this, we weighted all models by the inverse of the SE in the proportion of URM students per section.

To study the risk of not continuing after each course and to document attrition over time, we analyzed students who began the series in the initial general chemistry course and excluded students who transferred into a later course or entered with advanced placement credit. We also excluded students who took the initial general chemistry course in 2016 in analyses that involved subsequent events, to avoid classifying them as a nonpersister if, in fact, they continued in general chemistry after 2016. We simplified the analysis by examining students' first attempt at each course in the series. In this way, all students in the sample shared the same history, e.g., all students advancing to the third course in general chemistry shared the experience of passing the first two courses on their first attempt. We then constructed a multilevel, discrete-time logit-hazard model with group membership as a predictor at the student level and clustered by student to account for repeated measures due to each student enrolling in multiple courses throughout the series. Details on model selection and the final model for the risk and survival analysis are given in the Supplementary Materials.

We developed logistic mixed-effect models of failure (DFW) rates to account for the variation in this outcome that can be attributed to section, quarter, and year effects. We followed the same model se-

lection procedure described above. We calculated bootstrapped CI empirically by randomly sampling 5000 individuals from the full dataset 100 times and calculating the mean. Relative risks were estimated by the odds ratios, which we obtained from the logistic regression coefficients.

To estimate the impact of grade disparities in GenChem 1 on enrollment in GenChem 2, we quantified the probability of a student (i) enrolling in GenChem 2, (ii) retaking GenChem 1, or (iii) leaving the chemistry series and then used multinomial regression to analyze variation in each outcome. As described above, we conducted backwards elimination using maximum likelihood estimation and removed predictors from the model until all individual predictors reached significance at the 0.05 threshold for type I error (table S6C). However, this model did not include random effects due to low inter-class correlations.

We conducted all statistical analyses in R (39). We used the packages lme4, nnet, and muMIn for multilevel linear models, multinomial logistic regressions, and model selection, respectively.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/24/eaaz5687/DC1>

## REFERENCES AND NOTES

1. E. M. Bensimon, Closing the achievement gap in higher education: An organizational learning perspective. *New Dir. High. Educ.* **2005**, 99–111 (2005).
2. B. Spitzer, J. Aronson, Minding and mending the gap: Social psychological interventions to reduce educational disparities. *Brit. J. Educ. Psychol.* **85**, 1–18 (2015).
3. D. C. Haak, J. HilleRisLambers, E. Pitre, S. Freeman, Increased structure and active learning reduce the achievement gap in introductory biology. *Science* **332**, 1213–1216 (2011).
4. R. L. Matz, B. P. Koester, S. Fiorini, G. Grom, L. Shepard, C. G. Stangor, B. Weiner, T. A. McKay, Patterns of gendered performance differences in large introductory courses at five research universities. *AERA Open* **3**, 1–12 (2017).
5. C. Alexander, E. Chen, K. Grumbach, How leaky is the health career pipeline? Minority student achievement in college gateway courses. *Acad. Med.* **84**, 797–802 (2009).
6. National Academies of Sciences, Engineering, and Medicine, *Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Students' Diverse Pathways* (The National Academies Press, 2016).
7. National Center for Educational Statistics, *Status and Trends in the Education of Racial and Ethnic Groups, Indicator 26: STEM Degrees* (U.S. Department of Education, 2019).
8. S. L. Dika, M. M. D'Amico, Early experiences and integration in the persistence of first-generation college students in STEM and non-STEM majors. *J. Res. Sci. Teach.* **53**, 368–383 (2016).
9. President's Council of Advisors on Science and Technology, *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics* (Executive Office of the President, 2012).
10. S. S. Levine, E. P. Apfelbaum, M. Bernard, V. L. Bartelt, E. J. Zajac, D. Stark, Ethnic diversity deflates price bubbles. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18524–18529 (2014).
11. P. M. DiBartolo, L. Gregg-Jolly, D. Gross, C. A. Manduca, E. Iverson, D. B. Cooke III, G. K. Davis, C. Davidson, P. E. Hertz, L. Hubbard, S. K. Ireland, C. Mader, A. Pai, S. Raps, K. Siwicki, J. E. Swartz, Principles and practices fostering inclusive excellence: Lessons from the Howard Hughes Medical Institute's capstone institutions. *CBE Life Sci. Educ.* **15**, ar44 (2016).
12. D. A. Barr, J. Matsui, S. F. Wanat, M. E. Gonzalez, Chemistry courses as the turning point for premedical students. *Adv. Health Sci. Educ.* **15**, 45–54 (2010).
13. R. W. Fairlie, F. Hoffmann, P. Oreopoulos, A community college instructor like me: Race and ethnicity interactions in the classroom. *Am. Econ. Rev.* **104**, 2567–2591 (2014).
14. M. M. Chingos, Instructional quality and student learning in higher education: Evidence from developmental algebra courses. *J. Higher Educ.* **87**, 84–114 (2016).
15. M. C. Jackson, G. Galvez, I. Landa, P. Buonara, D. B. Thoman, Science that matters: The importance of a cultural connection in underrepresented students' science pursuit. *CBE Life Sci. Educ.* **15**, ar42 (2016).
16. L. I. Rendón, A. Nora, R. Bledsoe, V. Kanagala, *Científicos Latinxs: The Untold Story of Underserved Student Success in STEM Fields of Study* (Center for Research and Policy in Education, University of San Antonio, 2019).

17. J. Ellis, B. K. Fosdick, C. Rasmussen, Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLoS ONE* **11**, e0157447 (2016).
18. K. Rask, J. Tiefenthaler, The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Econ. Educ. Rev.* **27**, 676–687 (2008).
19. C. S. Rozek, G. Ramirez, R. D. Fine, S. L. Beilock, Reducing socioeconomic disparities in the STEM pipeline through student emotion regulation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1553–1558 (2019).
20. R. Shedlosky-Shoemaker, J. M. Fautsch, Who leaves, who stays? Psychological predictors of undergraduate chemistry students' persistence. *J. Chem. Educ.* **92**, 408–414 (2015).
21. A. Bandura, The explanatory and predictive scope of self-efficacy theory. *J. Soc. Clin. Psychol.* **4**, 359–373 (1986).
22. J. Y. K. Chan, C. G. Bauer, Identifying at-risk students in general chemistry via cluster analysis of affective characteristics. *J. Chem. Educ.* **91**, 1417–1425 (2014).
23. S. M. Villafaña, C. A. Garcia, J. E. Lewis, Exploring diverse students' trends in chemistry self-efficacy throughout a semester of college-level preparatory chemistry. *Chem. Educ. Res. Pract.* **15**, 114–127 (2014).
24. C. M. Steele, J. Aronson, Stereotype threat and the intellectual test performance of African-Americans. *J. Pers. Soc. Psych.* **69**, 797–811 (1995).
25. G. L. Cohen, J. Garcia, N. Apfel, A. Master, Reducing the racial achievement gap: A social-psychological intervention. *Science* **313**, 1307–1310 (2006).
26. A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, T. A. Ito, Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science* **330**, 1234–1237 (2010).
27. H. Jordt, S. L. Eddy, R. Brazil, I. Lau, C. Mann, S. E. Brownell, K. King, S. Freeman, Values affirmation intervention reduces achievement gap between underrepresented minority and white students in introductory biology classes. *CBE Life Sci. Educ.* **16**, ar41 (2017).
28. E. J. Theobald, M. J. Hill, E. Tran, S. Agrawal, E. N. Arroyo, S. Behling, N. Chambwe, D. Laboy Cintrón, J. D. Cooper, G. Dunster, J. A. Grummer, K. Hennessey, J. Hsiao, N. Iranon, L. Jones II, H. Jordt, M. Keller, M. Lacey, C. E. Littlefield, A. Lowe, S. Newman, V. Okolo, S. Olroyd, B. R. Peacock, S. B. Pickett, D. L. Slager, I. W. Caviedes-Solis, K. E. Stanchak, V. Sundaravandan, C. Valdebenito, C. R. Williams, K. Zinsli, S. Freeman, Active learning narrows achievement gaps for underrepresented students in science, technology, engineering, and math. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 6476–6483 (2020).
29. M. Estrada, A. Eroy-Reveles, J. Matsui, The influence of affirming kindness and community on broadening participation in STEM career pathways. *Soc. Iss. Pol. Rev.* **12**, 258–297 (2018).
30. K. Rainey, M. Dancy, R. Mickelson, E. Stearns, S. Moller, A descriptive study of race and gender differences in how instructional style and perceived professor care influence decisions to major in STEM. *Int. J. STEM Educ.* **6**, 6 (2019).
31. M. Stains, J. Harshman, M. K. Barker, S. V. Chasteen, R. Cole, S. E. DeChenne-Peters, M. K. Eagan Jr., J. M. Esson, J. K. Knight, F. A. Laski, M. Levis-Fitzgerald, C. J. Lee, S. M. Lo, L. M. McDonnell, T. A. McKay, N. Michelotti, A. Musgrove, M. S. Palmer, K. M. Plank, T. M. Rodela, E. R. Sanders, N. G. Schimpf, P. M. Schulte, M. K. Smith, M. Stetzer, B. Van Valkenburgh, E. Vinson, L. K. Weir, P. J. Wendel, L. B. Wheeler, A. M. Young, Anatomy of STEM teaching in North American universities. *Science* **359**, 1468–1470 (2018).
32. M. Cooper, M. Klymkowsky, Chemistry, life, the universe, and everything: A new approach to general chemistry and a model for curricular reform. *J. Chem. Educ.* **90**, 1116–1122 (2013).
33. B. Uttl, C. A. White, D. W. Gonzalez, Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Stud. Educ. Eval.* **54**, 22–42 (2017).
34. M. A. Deri, P. Mills, D. McGregor, Structure and evaluation of a flipped general chemistry course as a model for small and large gateway science courses at an urban public institution. *J. Coll. Sci. Teach.* **47**, 68–77 (2018).
35. C. A. Stanich, M. A. Pelch, E. J. Theobald, S. Freeman, A new approach to supplementary instruction narrows achievement and affect gaps for underrepresented minorities, first-generation students, and women. *Chem. Educ. Res. Pract.* **19**, 846–866 (2018).
36. K. D. Gibbs Jr., P. Marsteller, Broadening participation in the life sciences: Current landscape and future directions. *CBE Life Sci. Educ.* **15**, ed1 (2016).
37. E. Theobald, Students are rarely independent: When, why, and how to use random effects in discipline-based education research. *CBE Life Sci. Educ.* **17**, rm2 (2018).
38. A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, G. M. Smith, *Mixed Effects Models and Extensions in Ecology with R* (Springer, 2009).
39. R Core Team, *A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
40. D. F. Halpern, C. P. Benbow, D. C. Geary, R. C. Gur, J. S. Hyde, M. A. Gernsbacher, The science of sex differences in science and mathematics. *Psychol. Sci. Public Interest* **8**, 1–51 (2007).

**Acknowledgments:** We thank K. Quigley and D. M. Heinekey for assistance with obtaining records on teaching assignments and instructor characteristics. **Funding:** This work was supported by grants from the Howard Hughes Medical Institute (52008126) and from the University of Washington Office of the Dean of Arts and Sciences. **Author contributions:** All authors designed the research. M.R.M. obtained the data from the University of Washington Registrar's Office, and J.B. obtained the data from the University of Washington Chemistry Department. R.B.H. and M.R.M. performed statistical analyses with input from E.J.T. S.F. wrote the manuscript with input from R.B.H., M.R.M., and E.J.T. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The raw, de-identified data used in the study are available at [https://github.com/rebzy/UW\\_chem\\_gap](https://github.com/rebzy/UW_chem_gap). Additional data related to this paper may be requested from the authors.

Submitted 19 September 2019

Accepted 13 April 2020

Published 10 June 2020

10.1126/sciadv.aaz5687

**Citation:** R. B. Harris, M. Mack, J. Bryant, E. J. Theobald, S. Freeman, Reducing achievement gaps in undergraduate general chemistry could lift underrepresented students into a "hyperpersistent zone". *Sci. Adv.* **6**, eaaz5687 (2020).