Chemistry Education Research and Practice

PAPER

View Article Online
View Journal | View Issue

Cite this: *Chem. Educ. Res. Pract.*, 2018, **19**, 783

Improving general chemistry performance through a growth mindset intervention: selective effects on underrepresented minorities

Angela Fink,[a] Michael J. Cahill,[a] Mark A. McDaniel,[ab] Arielle Hoffman[a] and Regina F. Frey *[ac]

Women and minorities remain underrepresented in chemistry bachelor's degree attainment in the United States, despite efforts to improve their early chemistry achievement through supplemental academic programs and active-learning approaches. We propose an additional strategy for addressing these disparities: course-based, social-psychological interventions. For example, growth-mindset interventions are designed to support students during challenging academic transitions by encouraging them to view intelligence as a flexible characteristic that can be developed through practice, rather than a fixed ability. Previous research has shown that such interventions can improve the overall performance and persistence of college students, particularly those who belong to underrepresented groups. We report a random-assignment classroom experiment, which implemented a chemistry-specific growth-mindset intervention among first-year college students enrolled in General Chemistry 1. Performance results revealed an achievement gap between underrepresented minority and white students in the control group, but no sex-based gap. Critically, after adjusting for variation in academic preparation, the mindset intervention eliminated this racial-achievement gap. Qualitative analysis of students' written reflections from the intervention shed light on their experiences of the mindset and control treatments, deepening our understanding of mindset effects. We integrate these results with the mindset and chemical education literatures and discuss the implications for educators seeking to support underrepresented students in their own classrooms.

Received 7th December 2017,
Accepted 13th April 2018

DOI: 10.1039/c7rp00244k

rsc.li/cerp

Introduction

In many science, technology, engineering, and mathematics (STEM) fields, women and certain ethnic and racial groups remain underrepresented in bachelor's degree attainment in the United States. For example, while women earned more than half of all bachelor's degrees in biological sciences in 2014, they received only 40% of degrees in physical sciences like chemistry (National Science Foundation, 2017). Such gender gaps are not unique to the U.S.; instead, they have been documented in 67 nations and regions, and they are counterintuitively larger in more gender-equal societies (Stoet and Geary, 2018). A similar pattern emerged among ethnic and racial groups with a history of underrepresentation in U.S. higher education, namely Blacks, Hispanics, and American Indians or Alaska Natives. Together, these underrepresented minority groups (URMs) made up 31% of the U.S. population in 2014; however, they earned only 21% of all STEM bachelor's degrees, and even fewer in chemistry (18%; National Science Foundation, 2017). In the current study, we present a novel approach for addressing such disparities within the field of chemistry. Specifically, we report a randomized classroom experiment that used a course-based, social-psychological intervention to help first-year college students respond effectively to the academic challenges they face in general chemistry.

Extensive chemical-education research has explored strategies for supporting students' early achievement during the transition to college. We adopt the same focus here, because early GPA is known to impact the retention of underrepresented groups in STEM (Chang *et al.*, 2014; Dika and D'Amico, 2016). In terms of previous research, numerous studies have evaluated diagnostic tools for identifying at-risk students in general chemistry (Wagner *et al.*, 2002; Lewis and Lewis, 2007; Mills *et al.*, 2009; Kennepohl *et al.*, 2010; Potgieter *et al.*, 2010; Shields *et al.*, 2012; Frey *et al.*, 2017), while others have assessed supplemental programs designed specifically to enhance at-risk students' academic skills and performance (Pienta, 2003; Bentley and Gellene, 2005; Botch *et al.*, 2007; Rath *et al.*, 2012; Shields *et al.*, 2012). Another body of work has explored the

[a] *Center for Integrative Research on Cognition, Learning, and Education (CIRCLE), Washington University in St. Louis, St. Louis, MO 63130, USA. E-mail: gfrey@wustl.edu; Tel: +1-314-935-7474*
[b] *Psychological and Brain Sciences, Washington University in St. Louis, USA*
[c] *Department of Chemistry, Washington University in St. Louis, USA*

impact of broader curricular changes, like the incorporation of more active and collaborative learning, on all general chemistry students (Báez-Galib *et al.*, 2005; Cooper *et al.*, 2008; Gafney and Varma-Nelson, 2008; MacArthur and Jones, 2008; Bunce *et al.*, 2010; Eichler and Peeples, 2016; Gregorius, 2017). It is now well-established that such techniques can generally improve students' STEM learning and performance outcomes (Prince, 2004; Freeman *et al.*, 2014). Finally, recent research has begun to look beyond students' cognitive and academic skills, examining how social-psychological factors like motivation and self-efficacy might influence their success in general chemistry (Ferrell and Barbera, 2015; Chan and Bauer, 2016; Ferrell *et al.*, 2016). The current study builds on that foundation, experimentally manipulating students' social-psychological processes to try and boost their general chemistry performance.

To our knowledge, this study is the first to implement a social-psychological intervention to facilitate students' transition into college-level chemistry and address the disparities that might emerge there (see Jordt *et al.*, 2017, for parallel work using a different type of social-psychological intervention in introductory biology). Such social-psychological interventions have proven successful at reducing educational achievement gaps (Yeager and Walton, 2011) and have been used to address societal issues and problematic behaviors in various domains (Walton, 2014). We selected this strategy because, separate from any potential issues of academic preparation, underrepresented groups can face unique psychological barriers to academic success (*e.g.*, Jury *et al.*, 2017; see the Literature review for discussion). We therefore argue that psychological interventions designed to mitigate those barriers provide a complementary approach to more traditional methods for supporting general chemistry students, especially those from underrepresented groups.

In particular, we designed a three-phase, "growth-mindset" intervention to manipulate students' implicit mindsets about intelligence, which have been shown to influence their academic goals, behaviors, and outcomes (Dweck and Leggett, 1988). We implemented the intervention as part of a random-assignment experiment among students enrolled in the first semester of a two-part General Chemistry course sequence. The students self-administered the growth-mindset treatment or a control intervention outside of class, as part of their online homework. This design minimized the administrative burden on instructors and the potential for instructor effects. It also increased our confidence that any observed effects were caused by our experimental manipulation, rather than differences among students' experiences in General Chemistry 1. In other words, our experimental design was intended to fit easily into the pre-existing course structure, while also providing a rigorous test of the mindset intervention.

Beyond its contribution to the chemical-education literature, this study expands the growth-mindset literature in several ways. As far as we know, ours is the first mindset intervention to target students' domain-specific views about intelligence in a particular subject, rather than their views about general intelligence (Aronson *et al.*, 2002; Good *et al.*, 2003; Blackwell *et al.*, 2007; Paunesku *et al.*, 2015; Yeager *et al.*, 2016a, 2016b). Following recent research, which shows the educational impact of students' domain-specific views about mathematics intelligence (Good *et al.*, 2012; Rattan *et al.*, 2012), we chose to focus students' attention specifically on their general chemistry abilities and performance. To achieve that focus, we synced the three phases of our intervention to the General Chemistry 1 exam schedule, asking students to consider the growth-mindset materials while reflecting on their study strategies for upcoming exams. We hope that this study will provide guidance to chemistry instructors and researchers interested in incorporating a mindset intervention into their own course structures.

Another contribution is our usage of a strong, positive control that might support students' achievement during the transition into college-level STEM courses like General Chemistry 1. Most mindset studies compare the intervention with an active control condition that is engaging but unlikely to influence students' course performance, like an antidrug treatment (Good *et al.*, 2003) or a lesson about brain anatomy and function (without mention of flexibility; Paunesku *et al.*, 2015; Yeager *et al.*, 2016b). Instead, we follow Blackwell *et al.* (2007) in using a control condition that includes information about study skills and time management, parallel to the mindset materials. This design provides a stringent test of our intervention, and it helps allay ethical concerns about providing only a subset of students with the treatment intervention.

Finally, we included a novel data type in our mindset investigation: students' written responses from the second and third phases of the intervention. Qualitative analysis of those responses provides a manipulation check, confirming whether students in the mindset and control conditions focused their attention on the unique themes of their assigned materials. Further exploration of the qualitative data may also strengthen our interpretation of the performance results, helping us understand how students' experiences of the mindset intervention impacted their approach to General Chemistry 1.

In the remainder of the introduction, we review the STEM education and psychology research that informed our study. Drawing on this literature, we present four research questions, explaining our hypotheses and predictions about the impact of a mindset intervention on our student population. Next, we describe the General Chemistry 1 course context, our sample, and the design and implementation of our intervention in more detail. In the Results section, we address our four research questions in turn. We begin by examining students' performance data for effects of student sex,† race, and

† The terms "gender" and "sex" are utilized in this manuscript for distinct purposes. The term "gender," which refers to students' socially-constructed gender identity, is used throughout the literature review and in discussions linking our study to previous work. The term "sex," which refers to students' biological sex, is used when describing the results of the current study. We make this distinction because the social construct of gender and the existence of gender-based disparities are commonly discussed in the education and psychology literatures, and we sought to adopt the standard language of those fields. However, as reported in the Methods section, the university instrument used to gather demographic information about our sample asked for students' sex, and we wish to accurately represent those data. We therefore attempted to apply each term where appropriate and to highlight the distinction, in order to connect our work with the literature while also avoiding the implication that the terms are synonymous.

intervention condition (treatment *vs.* control). After establishing the quantitative results, we present the key observations from the qualitative analysis of students' intervention responses. In the general discussion, we discuss the implications of our findings for chemistry educators and researchers.

Literature review

Origins of underrepresentation

In order to counteract the international problem of disparities in STEM representation, we must strive to understand their origins. Importantly, current disparities cannot be attributed simply to differences in STEM interest. A recent report indicates that in 2009, underrepresented minority students in the U.S. matched their white and Asian peers in terms of STEM aspirations upon college entrance (Higher Education Research Institute, 2010). These data suggest that aspects of minority students' college experiences are driving them away from STEM fields like chemistry at higher rates than their peers. Research in psychology and education has explored a range of factors that might contribute to this pattern. For instance, it is well established that women and minority students in STEM may be subject to psychological stress in the form of ''stereotype threat,'' or a fear of confirming negative stereotypes about one's group, which can negatively impact performance (Steele, 1997). When negative stereotypes are salient, African American students demonstrate worse academic performance overall (*e.g.*, Steele and Aronson, 1995), and women struggle in specific domains where they are negatively stereotyped, like mathematics (*e.g.*, Spencer *et al.*, 1999).

Recent work suggests that the experience of stereotype threat during STEM classes and assessments reflects broader sociocultural conflicts between underrepresented groups and the scientific communities they seek to join. For example, as African Americans advance in STEM professions, they are more likely to report perceptions of racially-motivated microaggressions and feelings of poor alignment with the cultural norms of the scientific community (Brown *et al.*, 2016). Because of those experiences, an important factor in African Americans' scientific success is the development of effective strategies for dealing with racial bias. Managing such sociocultural conflicts may prove especially difficult for women of color (Carlone and Johnson, 2007) and for students who not only belong to underrepresented minority groups, but also come from a lower socioeconomic background (Jury *et al.*, 2017). For these individuals, multiple dimensions of their personal identities—their gender, ethnicity, race, and/or socioeconomic status—might be viewed as incompatible with a ''science identity,'' preventing them from receiving recognition and opportunities within the scientific community and potentially driving them away from their fields of interest (Carlone and Johnson, 2007).

Compounding these sociocultural pressures, the experience of academic adversity early in the transition to college may deter women and minorities from persisting in STEM. A recent study of STEM outcomes among first-generation college students, many of whom also belong to underrepresented ethnic and racial groups, showed that first-semester college GPA was the most robust predictor of STEM retention (Dika and D'Amico, 2016). A large-scale analysis of educational data from primarily minority-serving institutions provides further support for this hypothesis, demonstrating that academic preparation and STEM self-efficacy were key predictors of minority retention (Chang *et al.*, 2014). Students who received more rigorous pre-college training or felt more confident in their scientific skills were more likely to persevere in STEM fields, presumably because they encountered greater academic success than their less prepared counterparts. At the same time, the U.S. education system provides underrepresented minority and low-income students with disproportionately fewer college-preparation resources than their peers. For instance, a national study examining K-16 education policies and practices in six states across the country found substantial inequity in students' access to college counseling, advanced-placement course opportunities, and inter-institutional relationships with local colleges (Venezia and Kirst, 2005). Given such systemic inequality, students faced with identity-based threats may also contend with an exceptionally steep academic learning curve when they enter college, particularly in demanding introductory STEM courses.

Importantly, even among students who are well prepared, their subjective interpretations of academic adversity may contribute to underrepresentation. This argument gains support from the finding that women in graduate school are more likely than men to interpret high-effort expenditures as an indicator of their ill fit in STEM (Smith *et al.*, 2013). Such data suggest that negative stereotypes about their STEM abilities lead women to internalize academic struggles as evidence of their limitations, rather than surmountable challenges. In this way, sociocultural and academic factors can interact to undermine the performance and long-term persistence of underrepresented groups in STEM. Our study takes aim at the intersection of these forces, offering a relatively straightforward, low-cost option for helping underrepresented students respond more resiliently to academic adversity: course-based, social-psychological interventions.

Mindset interventions in education

The usage of social-psychological interventions to support students at risk of underachievement is not a new strategy (*e.g.*, Wilson and Linville, 1982, 1985), but it has gained traction in the past decade or so. Although such interventions vary greatly in methodology, they share several key characteristics (see Yeager and Walton, 2011, for a review). Effective interventions are persuasive but not pushy, encouraging students to reflect on ideas without explicitly instructing them to adopt them. They are sensitive to students' subjective experiences of their educational contexts, and they target recursive processes in those contexts, thereby effecting lasting changes in students' behaviors and perspectives. In other words, successful interventions are ''wise'' to the psychological mechanisms that underlie the societal issues or problematic behaviors they seek to alter, and they manipulate those mechanisms in a context-appropriate way (Walton, 2014).

Following these principles, we designed a growth-mindset intervention to manipulate students' perceptions of the academic setbacks they often face in introductory STEM courses (Aronson *et al.*, 2002; Good *et al.*, 2003; Blackwell *et al.*, 2007; Paunesku *et al.*, 2015; Yeager *et al.*, 2016a, 2016b). When we talk about students' mindsets, also called "implicit theories" or "lay theories," we refer to their beliefs about personal characteristics like intelligence or social competency. Growth-mindset interventions draw on a body of work arguing that students' mindsets about intelligence have wide-ranging cognitive, affective, and behavioral consequences (Dweck and Leggett, 1988; Dweck, 2006; Kennett and Keefer, 2006; Burnette *et al.*, 2013). Specifically, students who hold flexible views of intelligence, *i.e.*, growth mindsets, are likely to set goals that are focused on the learning and mastery of new skills. This is in contrast to students with fixed mindsets, who are likely to set performance goals; depending on the task at hand, they seek either to demonstrate their proficiency or to avoid demonstrating a lack of skill. As a result of their unique goals, students with growth *vs.* fixed mindsets approach academic challenges differently. For example, students with growth mindsets may embrace complex chemistry problems as learning opportunities, being willing or even excited to struggle with them and adaptively try out new strategies. On the other hand, students with fixed mindsets tend to experience performance anxiety and shy away from complex problems they have not mastered or do not recognize, a maladaptive response that hinders learning. Ultimately, the cognition-affect-behavior patterns that emerge from students' mindsets about intelligence lead to greater resiliency and higher achievement among those with growth mindsets (Yeager and Dweck, 2012).

Critically, an increasing number of studies have found that growth mindsets can be experimentally induced. Although the effects may vary across individuals, depending on their baseline skills and views of intelligence (Burns and Isbell, 2007), mindset interventions have been shown to improve academic performance and persistence, especially during educational transitions and among at-risk students. The majority of studies have demonstrated mindset effects during the transitions into junior high and high school (Good *et al.*, 2003; Blackwell *et al.*, 2007; Paunesku *et al.*, 2015; Yeager *et al.*, 2016a; but *cf.* Chao *et al.*, 2017, for more equivocal results), however studies have also found positive effects among students transitioning into college (Aronson *et al.*, 2002; Yeager *et al.*, 2016b). Our study will expand the empirical evidence for mindset effects at the collegiate level, extending the current literature into a new educational context (*i.e.*, chemical education) and utilizing a novel, domain-specific intervention.

Hypotheses and predictions

Before describing the methods, we review the research questions and hypotheses that guided our investigation.

Research question 1: After accounting for variation in academic preparation, does sex or race significantly influence students' first-semester performance in General Chemistry 1? We hypothesize that the diagnostic nature of typical course assessments will induce stereotype threat among females, who are negatively stereotyped in math-intensive fields like chemistry (Good *et al.*, 2008), as well as underrepresented minority groups like African Americans, who are negatively stereotyped in academic domains in general (Steele and Aronson, 1995). As a result, we expect the General Chemistry 1 achievement of control participants from those groups to lag behind their male and white counterparts, respectively.

Research question 2: Does a social-psychological intervention that manipulates students' mindsets about intelligence significantly improve any or all students' achievement in General Chemistry 1? Much of the work on growth-mindset interventions explicitly portrays them as a method for counteracting stereotype threat or others factors that disadvantage and prevent students from reaching their potential (Aronson *et al.*, 2002; Good *et al.*, 2003; Yeager *et al.*, 2016b). We therefore predict that our mindset intervention will selectively benefit females and underrepresented minority students, mitigating the effects of the additional psychological stressors they may face in General Chemistry 1, compared to their peers. In other words, we expect the students' intervention condition (mindset *vs.* control) to interact with their sex and race.

Research question 3: Does the mindset intervention offer a long-term benefit, such that mindset effects transfer to the second-semester course General Chemistry 2? At least two previous studies have documented long-term mindset effects, such that interventions administered prior to (Yeager *et al.*, 2016b) or at the middle of the academic year (Aronson *et al.*, 2002) positively impacted college students' end-of-year outcomes. This evidence suggests that any mindset effects in General Chemistry 1 might extend to performance in the second part of the course sequence, General Chemistry 2. If long-term benefits emerge, they would strengthen the argument that brief social-psychological interventions have the potential to improve students' transition into college-level chemistry, thereby increasing their chances of retention in the field.

Research question 4: What unique themes do mindset *vs.* control participants write about during the intervention, and what themes do they share in common? This question is more exploratory in nature, examining the contents of students' written responses during the intervention reflections. At a minimum, such qualitative analyses will provide a manipulation check, confirming whether our two conditions evoked different responses. In addition, they may clarify our interpretation of the performance data.

Methods

Course context

General Chemistry 1 is the first course in a two-part sequence, and it enrolls 700–800 students each fall. The course includes three 1 hour lectures per week, plus a mandatory weekly recitation, with the associated laboratory as a separate course. The lectures are divided into three sections but treated as a single unit. For instance, all sections utilize the same problems sets, quizzes, and exams; students from different sections are intermingled during recitations; and all sections are combined during the grading process. The graduate students who lead recitation sections are selected by the Department because of their experience and ability to teach group work. Prior to the course, they all participate in Department-led pedagogic training, which introduces best practices for developing mini-lectures, facilitating group discussions, and evaluating student work. During the course, the graduate assistants attend weekly meetings with their peers and the General Chemistry lecturers, where they receive detailed notes on teaching that week's recitation-problem set and discuss how to best facilitate that week's session.

Besides the required components of General Chemistry 1, several other supplemental learning opportunities are available to students. Instructor-led help sessions are offered daily, and students are invited to attend sessions with any of the different instructors. Students are encouraged to participate in department-sponsored Peer Led Team Learning (PLTL) groups that meet weekly to collaboratively solve practice problems (see Hockings *et al.*, 2008, for more detail). Finally, a Transition Program is available to students who score in the lowest 25% on a composite measure of academic preparation. This measure incorporates students' performance on the Chemistry Department's Online Diagnostic (OD) exam, which assesses their incoming chemistry content knowledge; their scores on the math portion of the ACT, which is a standardized college admissions test in the United States (ACT, 2018); and their scores on STEM-related advanced-placement (AP) exams, which evaluate students' field-specific academic skills after year-long, college-level courses that are taken during high school (The College Board, 2018a; see Shields *et al.*, 2012, for discussion of the OD exam and Transition Program). The Transition Program includes extended-length recitations, mandatory PLTL, and participation in smaller peer-mentored study groups.

Student sample

Study participants were first-year students enrolled in General Chemistry 1 during fall semester of 2015 or 2016. Information about participants' sex, race, and academic preparation were obtained from the Office of the University Registrar after they provided informed consent. Recruitment of participants necessarily spanned two years in order to obtain a large enough sample of underrepresented minority students, who comprised approximately 15% of the university's student body at the time of this study. Only first-year students were included in the sample, because the intervention and control materials were specifically intended to improve students' transitions into college.

A total of 565 first-year students belonged to our two target groups, underrepresented minority and white students. Asian first-year students were excluded from the analysis, because they do not fit clearly into our critical comparison groups. They are slightly overrepresented in STEM degree attainment in the U.S. (National Science Foundation, 2017), and historical data at our institution suggest that they tend to outperform white students in this course. A total of 17 first-year students were also excluded because they did not report racial or ethnic information. The underrepresented minority group (n = 136, 24% of sample) included students who self-identified as Black or African American, Hispanic or Latinx, American Indian or Alaska native, or Native Hawaiian or Pacific Islander. In terms of sex, students could identify as female or male; an intersex option was not available on the university's demographic survey, and gender identity was not requested by the university. The sample was 57% female (n = 324), and the distribution of females was similar across racial groups (62% of URMs, 56% of whites; see Table 1 for more detail). No students were excluded due to unreported sex data.

The majority of participants also enrolled in General Chemistry 2 during the spring semester immediately following their General Chemistry 1 course (n = 506, 90% of sample). As in the complete sample, this subset comprised 24% underrepresented minority students and 57% females. Students who did not continue into the second part of the course sequence predominantly belonged to engineering majors that only require General Chemistry 1.

Procedure

The intervention was implemented as a course requirement, so all students enrolled in General Chemistry 1 were pseudo-randomly

Table 1 Characteristics of the first-year sample enrolled in General Chemistry 1 in Fall 2015 or Fall 2016

	Complete sample		URMs		Whites	
	Mindset	Control	Mindset	Control	Mindset	Control
N	275	290	65	71	210	219
% female	56.4	57.6	55.4	66.2	56.7	54.8
% URM	23.6	24.5	—	—	—	—
Mean ACT math[a] (SE)	32.8 (0.13)	32.8 (0.15)	31.9 (0.31)	31.2 (0.34)	33.0 (0.14)	33.3 (0.14)
Mean AP proportion[b] (SE)	0.37 (0.02)	0.37 (0.02)	0.22 (0.03)	0.25 (0.03)	0.42 (0.02)	0.41 (0.02)
% in PLTL	66.1	66.9	67.7	67.6	65.6	66.7
% in Transition Program[c]	9.1	7.6	12.3	15.5	8.0	5.0

Note: SE indicates standard error. t-Tests and chi-squared tests were used to compare the characteristics of mindset and control participants within each group (complete sample, URMs, and whites). No significant differences were found ($ps > 0.05$). [a] For students who reported SAT math scores, concordance tables (Dorans, 1999) were used to convert those data into ACT math scores. [b] AP proportion reflects students' performance on the advanced placement exams for 4 STEM subjects: biology, calculus, chemistry, and physics. For each exam where a student earned a score of 4 or 5 (out of 5), their AP proportion score increased by 0.25. Thus, this proportion is an indicator of how many of the four AP STEM exams a student excelled at (scores of 4 or 5). [c] These data underestimate Transition Program participation, because Fall 2016 program participants were excluded from our sample after receiving an additional social-psychological intervention from their recitation instructor.

assigned to the growth mindset and control conditions when they logged into the online homework. Specifically, each student's university ID number was divided by 4, and the remainder was used to sort them into a condition. ID numbers with a remainder of 0 or 1 received the mindset intervention, while those with a remainder of 2 or 3 received the control (*e.g.*, 100 would have a mindset assignment because 100/4 has a remainder of 0; 102 would have a control assignment because 102/4 has a remainder of 2). This automated process grouped students into the same condition for each phase of the intervention.

We then analyzed data only from participants who provided informed consent and met the requirements outlined above (*i.e.*, first-year, URM or white, female or male). Consent forms were administered during the separate General Chemistry laboratory, where students are divided into twenty-person sections, so that a representative of the research team was better able to address individual students' questions. The consent form described the research study as a broad investigation of the teaching and learning techniques being utilized in lower- and upper-level STEM courses. Given this general framing, we have no reason to expect that the consent process influenced students' responses during the intervention.

Thus, our experimental sample represents only a subset of all randomly-assigned students, excluding those who (a) did not provide consent for us to access their data (15% of all students enrolled in the course), (b) consented but failed to meet our criteria (37% of students, mostly Asians and upper-year students), and (c) consented and met our criteria, but did not perform any experimental tasks (5% of students). Because these exclusion processes occurred after the randomization procedure, they may have introduced bias into our sample (Murnane and Willett, 2010). Specifically, such exclusions may have affected the comparability of the two groups, creating an imbalance in the attributes of the treatment and control participants. They may also have compromised the generalizability of our sample, differentiating our sample from the general population of all students enrolled in General Chemistry 1. To address these concerns, we conducted two sets of comparisons.

First, we compared the characteristics of sample participants in the mindset and control conditions (see Table 1 for descriptive statistics). Collapsing across race and within each racial group, *t*-tests confirmed that the mindset and control participants possessed comparable academic preparation, and chi-squared tests confirmed their equal participation in PLTL and the Transition Program‡ ($ps > 0.05$). These data indicate that our random assignment remained internally valid, despite students' subsequent self-selection into the study and exclusion of non-target groups by the research team. Next, we compared the experimental sample with first-year students from the target groups who consented but did not participate (*i.e.*, group c above). Table 2 indicates some significant differences between these two groups: non-participants were more likely to be male and less likely to participate in PLTL than participants in our sample. These findings suggest that our sample selection processes weakened the external validity of our study: our sample is not representative of all first-year students enrolled in General Chemistry I at our institution, which limits the generalizability of our results. We return to this issue in the Discussion section (see Limitations).

Table 2 Statistical comparisons assessing generalizability of the experimental sample

	Sample	Non-participants[b]	Statistic[a]
N	565	78	—
% female	57.0	39.7	7.55*
% URM	24.1	33.3	2.65
Mean ACT math (SE)	32.8 (0.10)	32.2 (0.31)	1.85
Mean AP proportion[c] (SE)	0.37 (0.01)	0.31 (0.03)	1.68
% in PLTL	66.4	44.9	12.80*
% in Transition Program[d]	8.3	9.0	<0.001

Note: SE indicates standard error. * indicates $p < 0.05$. [a] Pearson's chi-squared test (df = 1) for the categorical variables PLTL and Transition Program participation, independent t-tests (df = $n - 2$) for the continuous variables ACT math and AP proportion. [b] Non-participants include first-year students from our target groups (*i.e.*, URM or white, female or male) who consented to participate in the study but completed no experimental tasks. [c] AP proportion reflects students' performance on the advanced placement exams for 4 STEM subjects: biology, calculus, chemistry, and physics. For each exam where a student earned a score of 4 or 5 (out of 5), their AP proportion score increased by 0.25. [d] These data underestimate Transition Program participation, because Fall 2016 program participants were excluded from our sample after receiving an additional social-psychological intervention from their recitation instructor.

The experimental procedure was the same for both intervention conditions, requiring students to complete three online homework problems distributed over the course of the semester. The problems were framed as short reading and writing tasks that were not related directly to chemistry, but that covered information the instructors believed would help students navigate the course and their college careers more generally. This common framing downplayed the existence of two separate conditions, although students may have become aware of the manipulation through conversation with their peers.

Phase 1 of the intervention was embedded in the students' second online homework assignment (of 13) for the semester, which fell approximately 2 weeks before the first unit exam. During phase 1, students read a short article and completed a comprehension quiz. Mindset participants received an article entitled ''You Can Grow Your Brain'' (Yeager *et al.*, 2016b), which describes the brain as malleable, such that new connections can be grown and strengthened through effortful practice. It argues that even if someone has always struggled in a given domain, they can improve their abilities by taking on challenges and developing new learning strategies with the help of others. In contrast, control participants received ''Transition Tips'' for how to succeed in college, which did not reference the brain or mindsets about intelligence (see Appendix 1). This document

‡ Our descriptive statistics underestimate the rate of Transition Program participation, because Fall 2016 participants in the Program were not included in our sample. Their extended recitation instructor administered another social-psychological intervention (a self-affirmation task), which we expected to interfere with our experimental manipulation. While the exclusion of 2016 Transition Program participants impacted our sample size, there is no evidence that it influenced the success of our random assignment.

encouraged students to (i) get organized and manage their time, (ii) maintain their health and work-life balance, (iii) be an active course participant, and (iv) use available resources (*e.g.*, instructor-led help sessions).

Phases 2 and 3 of the intervention were included in the students' eighth and thirteenth online homework assignments, which fell 1 week before the second unit exam and the cumulative final exam, respectively. During these phases, students were reminded of the key points of their assigned article and prompted to write a reflection (see Appendix 2). The phase-2 reflection prompt asked students to explain how their article's ideas would influence their preparations for the upcoming exam (exam 2). In the same vein, the phase-3 prompt asked students to explain how the article's ideas would influence their study strategies for the final exam, reminding them of its comprehensive format. This type of reflection or application task, where participants are asked to espouse the key concepts of their treatment in writing, is common among social-psychological interventions, because endorsing an idea is known to strengthen an individual's belief in it (''saying-is-believing'' effect; Higgins and Rholes, 1978).

Data

Performance data. To test for achievement gaps and mindset effects on chemistry performance, we used students' General Chemistry 1 final-exam score as the dependent variable.§ The exams were *z*-scored within each year to correct for potential cross-year differences in final-exam difficulty and student attributes. Students were grouped according to their randomly assigned intervention conditions (mindset *vs.* control), their self-reported race (URM *vs.* white), and their self-reported sex (female *vs.* male). To increase the sensitivity of our statistical analyses, we accounted for variance (in the dependent variables) that otherwise would be considered error variation. We achieved this by including ACT math scores as an index of students' high-school academic preparation (ACT, 2018). If students reported math scores from the SAT, which is another standardized college entrance exam in the U.S. (The College Board, 2018b), those values were converted to ACT math equivalents using concordance tables (Dorans, 1999). This variable and other measures of mathematics ability have been shown to correlate with general chemistry performance (Tai *et al.*, 2005; Xu *et al.*, 2013). AP scores, which have also been shown to correlate with college chemistry performance (Sadler and Tai, 2007), provided a second index of variation in high-school academic preparation (The College Board, 2018a). These scores may index students' STEM content knowledge, or they may reflect the extent of students' experience with college-level STEM coursework. Either way, many students at our selective institution enter with AP exam scores (*e.g.*, 80% of the current sample) and incoming AP credit. We specifically examined students' performance on four STEM-related AP exams: biology, calculus, chemistry, and physics. We calculated the proportion of exams where they received a score of 4 or 5 (out of 5; ''AP Proportion''), such that students' AP proportion scores increased by 0.25 for each exam score meeting those criteria. For example, if a student earned a 5 on the Biology exam, 3 on Calculus, 4 on Chemistry, and 3 on Physics, then they were assigned an AP proportion score of 0.5. If a student earned 3's on all the exams or did not report any scores, then they received a score of 0. Essentially, this AP proportion score reflects how many of the four AP STEM exams a student excelled at (scores of 4 or 5).

To test the long-term effects of the mindset intervention, we examined participants' subsequent performance in their second-semester course, General Chemistry 2, using exam average (*z*-scored) as the dependent variable. Following the course instructors' procedure, exam average was calculated by averaging together each student's highest 2 (out of 3) unit exam scores and their final exam. This analysis included the same independent variables as above.

Qualitative data. To shed light on students' experiences of the mindset and control interventions, we examined their written responses to the phase 2 and 3 reflection prompts. Response rates during phase 2 were 84.5% and 87.5% in the mindset and control conditions, respectively. During phase 3, 75.8% of mindset participants and 69.9% of controls submitted responses. Collapsing across phases, this provided a total of 888 observations for thematic analysis.

Analysis

Performance data. We conducted a series of between-subjects ANCOVAs to determine the effects of race, sex, and intervention condition on final-exam scores in General Chemistry 1 and exam averages in General Chemistry 2. ACT math scores and AP proportion were used as covariates in all models, and the reported sample means (M) and standard errors (SE) have been adjusted accordingly (see Tables 4 and 5 in Appendix 3 for all unadjusted and adjusted exam means). Significance was evaluated with an alpha criterion of 0.05; partial eta-squared (η^2) provided effect-size estimates (small: 0.01, medium: 0.06, large: 0.14; Richardson, 2011); and *post hoc* comparisons were used to examine significant interactions.

We selected ANCOVAs as our primary statistical modeling technique in order to estimate each main effect and interaction effect, after accounting for the effects of all other terms (including the covariates). In contrast to multiple regression models, ANCOVAs also allow us to report the adjusted mean exam scores for each group of interest, which aid in the interpretation of the model. However, ANCOVAs assume that the independent variables and covariates are independent from one another and do not interact, and the usage of this technique and adjusted means may be inappropriate when those assumptions are not met (*e.g.*, Wicherts, 2005). While the data in Table 1 suggest that the covariates ACT math and AP

§ We used final exam, rather than exam average, because the intervention was not complete until the final exam. Such scores therefore allowed us to capture the ultimate effect of our multi-phase intervention. Moreover, because the final exam was cumulative, it still reflects students' overall learning and performance outcomes. Although final course grade also provides a comprehensive view of students' achievement, it incorporates a wide range of assignments, including homework and weekly quizzes, which might obscure students' performance on summative assessments.

proportion are not confounded with intervention condition, they may correlate with the other independent variables, sex and race. Indeed, visual inspection of Table 1 indicates that academic preparation varies according to race, such that white students enter our institution with stronger standardized test scores than their underrepresented peers, and *t*-tests confirm that these differences are significant (ACT math: $t = 6.25$, $p < 0.05$; AP proportion: $t = 6.93$, $p < 0.05$). Given this concern about violating the assumptions of our statistical model, we ran a series of multiple regression analyses (with dichotomous variables represented by standard dummy-coding) to further explore the data (see Appendix 4). The regression results revealed similar patterns to those observed in the ANCOVA results, though some of the key findings failed to reach significance. We highlight those discrepancies where appropriate in the Results sections below.

Qualitative data. We used content analysis to identify and categorize the emergent themes in students' free responses. Although most qualitative research inherently involves some form of content analysis, we are referring to a specific analytic technique that is more quantitative in nature, focused on capturing the variety and frequency of themes in a body of data (Merriam, 2009). Following Merriam's guidelines, the first step of our content-analysis process involved category construction, where an initial body of data was gradually examined and notated for themes relevant to the research question. One of the co-authors, AH, pseudo-randomly sampled 60 responses each from phases 2 and 3 of the intervention. For each observation, she paraphrased and recorded the most common study-related themes, adding them to a master list. As AH progressed though the sample, she eventually reached a point of saturation, where few new codes were being generated. At this point, the master code list included over 50 distinct codes per phase. Such a large collection of codes would be difficult for multiple raters to implement quickly or reliably; more importantly, it would fall short of distilling only the key emergent themes from our body of data.

AH therefore moved on to the second step of our content analysis, which involved iteratively sorting, refining, and grouping the codes into broader categories. AH made a first attempt at consolidating the master code list, combining and rephrasing codes that were redundant, while also removing those that were overly specific or rarely represented. She then met with the research team, presenting them with the refined code list and representative data points. Over several lab meetings, the group discussed and reached consensus on the phrasing of individual codes and the way they were grouped together. Next, AH applied the revised coding scheme to the remaining observations from her initial random sample, testing whether the scheme sufficiently captured all the study-related themes of interest. When she encountered data points that were difficult to categorize, AH once again met with the research team, who collaborated either to classify those data or to further revise the master code list. These two steps, category construction and refinement, lasted approximately 6 months and resulted in a total of 21 codes belonging to 8 categories (see Table 8 in Appendix 5 for the list of codes, their descriptions, and illustrative examples).

During the third step of our content-analysis process, we sought to establish the validity of our codes by calculating their inter-rater reliability. A group of three raters—AH, CH, and AF—all independently coded a random sample of responses from both intervention conditions and phases, marking the presence (1) or absence (0) of each qualitative code. This sample comprised 11% of all available responses. It is standard practice for multiple raters to examine and calibrate their coding on only a subset of observations in this way (Potter and Levine-Donnerstein, 1999; Krippendorff, 2004). We then used the online statistical calculator ReCal OIR (Freelon, 2013) to derive Krippendorff's alpha statistic, which assesses the agreement among raters' coding while accounting for chance agreement. Analysis confirmed that the raters attained a satisfactory level of consensus, reflected by an average agreement rate of 95% and a Krippendorff's alpha of 0.80, which is considered the threshold for reliability (Krippendorff, 2013). After this validation process, the raters discussed and resolved the points of disagreement in their coding thus far. Finally, AH completed approximately 50% of the remaining qualitative coding, and CH and AF completed about 25% each.

With the finalized coding scheme in hand, we performed the final step of our content analysis, calculating the frequency of each code within each experimental condition using the following procedure. For every participant, AF combined their responses from phases 2 and 3. If a qualitative code had been marked present in either phase of the intervention, then the aggregate response was assigned a value of 1 for that code. If the target code had been marked absent in both phases, then the aggregate was assigned a value of 0. This binary coding, which was collapsed across phases, allowed us to determine the proportion of participants in each condition that mentioned a target theme at some point during the intervention (Table 3). Within intervention conditions, chi-squared tests were conducted to assess the significance of differences across racial groups (Table 3).

Results and discussion

We first confirmed that our covariates were correlated with General Chemistry 1 performance. Replicating previous research (Tai *et al.*, 2005; Xu *et al.*, 2013), we found a moderate, positive correlation between ACT math and final-exam scores, $r(563) = 0.38$, $p < 0.05$, which demonstrates that students with stronger math abilities tend to perform better in general chemistry. There was a similar correlation between AP proportion and final-exam scores, $r(563) = 0.39$, $p < 0.05$, which replicates the finding that students with higher scores on STEM-related AP exams typically perform better in general chemistry (Sadler and Tai, 2007). Based on these findings, we retained both ACT math and AP proportion as covariates in all models, to account for variation in students' academic

Table 3 Ranking of qualitative codes from most to least frequent in each condition, with racial group comparisons

	Mindset					Control				
Rank in condition	Code	All (275)	URM (65)	White (210)	Chi. sq.[a]	Code	All (290)	URM (71)	White (219)	Chi. sq.[a]
1	Generative Practice Problems	79.6	75.4	81.0	0.95	Management	73.5	76.1	72.6	0.33
2	Brain Active	58.6	49.2	61.4	3.04	Health Plan	60.7	60.6	60.7	<0.001
3	Repetitive Practice Problems	50.6	50.8	50.5	0.002	Group Plan	52.1	47.9	53.4	0.66
4	Helpful	48.0	40.0	50.5	2.18	Other Generative Techniques	43.8	43.7	43.8	<0.001
5	Other Generative Techniques	36.0	40.0	34.8	0.59	Helpful	39.3	36.6	40.2	0.29
6	Other Repetitive Techniques	33.5	32.3	33.8	0.05	Other Repetitive Techniques	37.2	35.2	37.9	0.17
7	Management	25.8	27.7	25.2	0.16	Resource	31.0	29.6	31.5	0.09
8	Resilience	16.0	16.9	15.7	0.05	Health Do	26.9	25.4	27.4	0.11
9	Resource	8.7	16.9	6.2	7.18*	Repetitive Practice Problems	25.5	23.9	26.0	0.12
10	Group Plan	7.6	9.2	7.1	0.31	Group Do	21.7	22.5	21.5	0.04
11	Confidence	7.6	6.2	8.1	0.26	Generative Practice Problems	20.3	16.9	21.5	0.69
12	Brain Passive	4.7	6.2	4.3	0.38	In Class Generative	14.1	16.9	13.2	0.59
13	Already Have Strategy	4.7	4.6	4.8	0.002	Already Have Strategy	6.2	2.8	7.3	1.86
14	Health Plan	3.6	1.5	4.3	1.07	Resilience	4.5	8.5	3.2	3.46
15	Not Helpful	2.2	1.5	2.4	0.17	Confidence	4.5	1.4	5.5	2.08
16	Group Do	1.1	1.5	1.0	0.16	Not Helpful	2.8	2.8	2.7	0.001
17	Health Do	1.1	0.0	1.4	0.94	Already Knew Info	2.1	0.0	2.7	1.99
18	Already Knew Info	0.7	1.5	0.5	0.78	In Class Connect Concepts	1.7	0.0	2.3	1.65
19	In Class Connect Concepts	0.0	0.0	0.0	—	In Class Repetitive	1.0	2.8	0.5	2.92
20	In Class Generative	0.0	0.0	0.0	—	Brain Active	0.0	0.0	0.0	—
21	In Class Repetitive	0.0	0.0	0.0	—	Brain Passive	0.0	0.0	0.0	—

Note: frequency reflects the proportion of participants assigned the target code, when their phase 2 and 3 intervention responses are combined. *n*s are in parentheses. [a] Pearson's chi-squared tests compared frequency among URM *vs.* white students within each condition, with *p*-values computed by Monte Carlo simulation with 2000 replicates, $^{*}p < 0.05$.

preparation and thereby increase our power to detect any intervention effects.

Research question 1: demographic effects on control performance

We tested for racial and sex-based disparities in General Chemistry 1 by examining the performance of control participants only. As predicted, a 2 (race) × 2 (sex) ANCOVA revealed a significant effect of race, such that white students (adjusted mean (M) = 65.1, standard error (SE) = 0.09) received higher final-exam scores than minority students (M = 59.2, SE = 1.7) after academic preparation is taken into account statistically, $F(1, 284) = 8.08$, $p < 0.05$, partial $\eta^2 = 0.03$. Contrary to our predictions, there was no significant effect of sex, indicating that male (M = 62.5, SE = 1.5) and female students (M = 61.8, SE = 1.2) earned comparable final-exam scores, $F < 1$, $p > 0.05$. There was also no significant interaction between race and sex, $F < 1$, $p > 0.05$.

Research question 2: a selective mindset benefit

Accounting for academic preparation, we conducted a 2 (race) × 2 (sex) × 2 (intervention) ANCOVA on the complete sample of performance data. The model revealed a small but significant main effect of race, such that white students (M = 65.0, SE = 0.6) earned higher adjusted final-exam scores than underrepresented minority students (M = 61.8, SE = 1.2), $F(1, 555) = 4.48$, $p < 0.05$, partial $\eta^2 = 0.01$. There was also a significant main effect of intervention condition, indicating that students in the mindset condition (M = 64.7, SE = 0.9) performed better than students in the control condition (M = 62.1, SE = 0.9), $F(1, 555) = 5.35$, $p < 0.05$, partial $\eta^2 = 0.01$. Critically, a significant two-way interaction emerged between race and intervention, $F(1, 555) = 5.15$, $p < 0.05$, partial $\eta^2 = 0.01$ (see Fig. 1).

Post hoc Tukey tests confirmed the presence of a mindset effect among underrepresented minority students, whose adjusted final-exam scores were more than 5 percentage points higher in the mindset condition (M = 64.6, SE = 1.6) compared to the control (M = 59.1, SE = 1.6), $t(555) = 2.62$, $p < 0.05$. In contrast, there was no mindset effect among white students, who performed similarly in the mindset (M = 64.8, SE = 0.9) and control conditions (M = 65.1, SE = 0.9), $t(555) = 0.04$, $p > 0.05$. Comparison of minority and white students in the mindset condition showed that after adjusting for academic preparation, all mindset participants performed equivalently regardless of race, $t(555) = 0.03$, $p > 0.05$. While the raw means in Fig. 1A indicate a residual difference between minority and white students' actual course performance, the adjusted means in Fig. 1B illustrate how the mindset intervention neutralized the racial achievement gap, once academic preparation is taken into account. Thus, the two-way interaction of intervention and race confirmed our expectations that the mindset intervention would selectively benefit minority students.

As above, there was no significant effect of sex on students' General Chemistry 1 performance on the final exam, $F < 1$, $p > 0.05$. The two-way interaction of sex and intervention also proved non-significant, $F < 1$, $p > 0.05$, indicating that the mindset intervention had a similar impact on female and male students. Finally, there was no significant three-way interaction between sex, race, and intervention, $F < 1$, $p > 0.05$. Given that our sample size may be insufficient to detect such a complex interaction, we refrain from interpreting this null result.

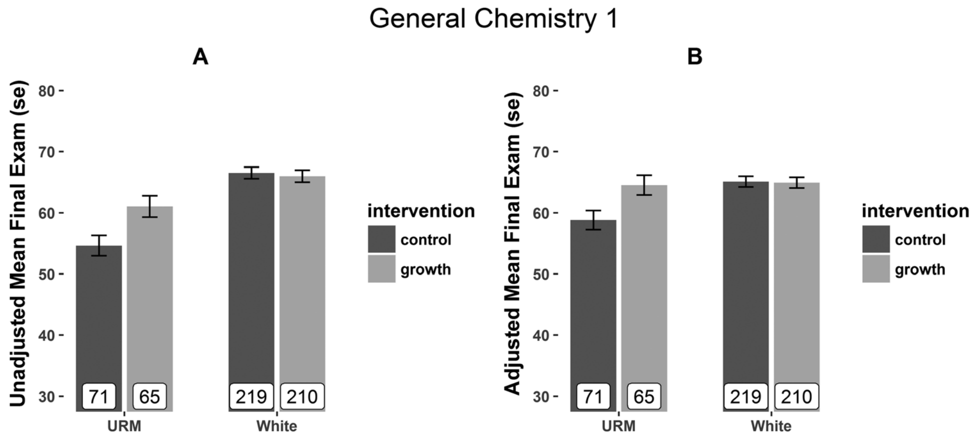

Fig. 1 Unadjusted (A) and adjusted (B) mean final-exam scores (standard error bars) from General Chemistry 1 by condition and racial group, collapsing across sex.

Regression results. When the General Chemistry 1 final-exam scores were analyzed using a treatment-coded multiple regression model (Table 6, Appendix 4), the results were less robust. Critically, the interaction of intervention and race failed to reach significance (unstandardized coefficient (b) = −0.37, $p > 0.05$). Based on this finding, we cannot eliminate the possibility that the mindset intervention similarly affected (or did not affect) underrepresented minority and white students. Accordingly, we constructed follow-up regression models within each racial group, to examine the overall effect of intervention condition when it does not interact with other variables (sex was included as a first-order effect only). These models revealed a significant mindset effect among minority students (b = 0.41, $p < 0.05$), but no effect among white students (b = 0.01, $p > 0.05$; Appendix 4), as depicted in Fig. 1. Thus, the regression and ANCOVA analyses produced a similar pattern of results, providing support for the interpretation that the mindset intervention selectively benefitted underrepresented minority students in General Chemistry 1.

Research question 3: transfer to General Chemistry 2

Next, we examined students' performance in the second-semester course, General Chemistry 2, to test for long-term effects of the mindset intervention. Consistent with the results above, we observed moderate, positive correlations between ACT math and exam average, $r(504) = 0.36$, $p < 0.05$, and between AP proportion and exam average, $r(504) = 0.42$, $p < 0.05$.

We then conducted a 2 (race) × 2 (sex) ANCOVA on the exam averages of control participants only, to test for achievement gaps in the course. The results confirmed that parallel to General Chemistry 1, underrepresented minority students in General Chemistry 2 received lower adjusted exam averages (M = 59.9, SE = 1.8) than white students (M = 65.8, SE = 1.0), $F(1, 254) = 9.26$, $p < 0.05$, partial $\eta^2 = 0.04$. Also parallel to the first-semester course, students' sex had no significant effect on their exam performance in General Chemistry 2, $F < 1$, $p > 0.05$. Based on these findings, we might expect the selective mindset benefit among minority students to transfer straightforwardly to General Chemistry 2.

We tested this prediction using a 2 (intervention) × 2 (race) × 2 (sex) ANCOVA on all participants' exam averages in General Chemistry 2. We found an overall effect of race, such that exam averages were lower among minority students (M = 62.6, SE = 1.3) than among white students (M = 66.4, SE = 0.7) after adjusting for differences in academic preparation, $F(1, 496) = 7.98$, $p < 0.05$, partial $\eta^2 = 0.02$. As expected from the analysis of control participants only, we found no evidence of a sex-based achievement gap, $F < 1$, $p > 0.05$. Crucially, the results revealed a main effect of intervention condition, indicating higher adjusted exam averages among mindset participants (M = 66.2, SE = 1.0) compared to controls (M = 62.8, SE = 1.0), $F(1, 496) = 4.11$, $p < 0.05$, partial $\eta^2 = 0.01$. Although the two-way interaction of intervention and race failed to reach significance, $F(1, 496) = 1.82$, $p > 0.05$, partial $\eta^2 = 0.00$, the means in Fig. 2 follow the same pattern observed in General Chemistry 1, suggesting a larger mindset effect among minority students. The overall effect of mindset did not differ by sex, nor was there a significant three-way interaction between intervention, sex, and race, $F\text{s} < 1$, $p\text{s} > 0.05$.

Regression results. Similar to the General Chemistry 1 data, the multiple regression models of students' General Chemistry 2 exam averages produced less robust results (Table 7, Appendix 4). As in the ANCOVA, the two-way interaction of intervention and race failed to reach significance (b = −0.30, $p > 0.05$). Unlike the ANCOVA, the first-order effect of intervention also failed to reach significance (b = 0.27, $p > 0.05$). Importantly, this first-order effect does not estimate the overall effect of the mindset intervention among all participants; instead, it captures the (non-significant) mindset effect among the reference group only, which comprises female minority students under our coding scheme. We therefore constructed follow-up regression models to examine the overall intervention effect within each racial group, when only first-order effects were included. These models revealed

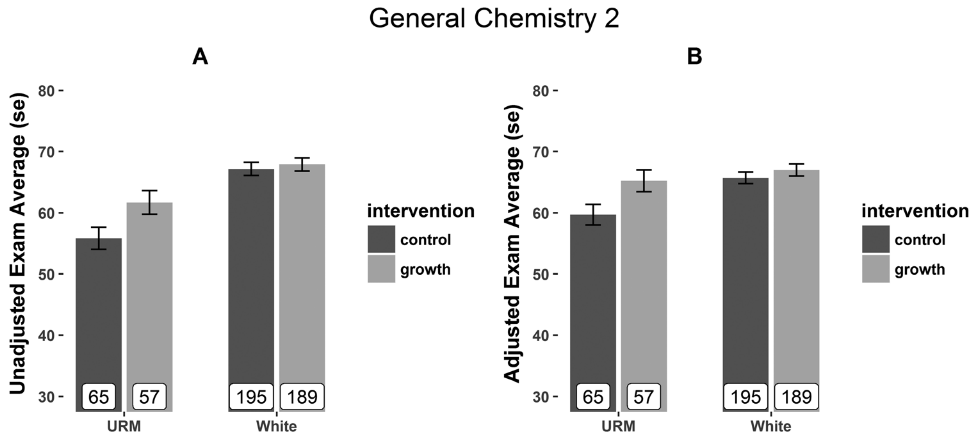

Fig. 2 Unadjusted (A) and adjusted (B) mean exam average (standard error bars) from General Chemistry 2 by condition and racial group, collapsing across sex. Exam average combines students' top 2 (of 3) unit exam scores and their final exam.

no significant mindset effects, despite a positive numerical relationship between the mindset treatment and performance, especially among the minority students (URMs: $b = 0.26$, $p > 0.05$; whites: $b = 0.04$, $p > 0.05$). Because the regression and ANCOVA results are equivocal, we refrain from drawing strong conclusions about the long-term effects of the mindset intervention on students' subsequent chemistry performance.

Research question 4: emergent themes

In the remainder of this section, we use students' written responses during the intervention as a window on their experiences of the mindset and control materials. As shown in Table 3, the data confirmed that mindset and control participants tended to focus on different themes during the intervention. Mindset participants reflected heavily on growing their brains and intelligence through generative study, while control participants reflected more on general student skills and time management. For example, the most frequent codes in the mindset condition were Generative Practice Problems (79.6%), Brain Active (58.6%), and Repetitive Practice Problems (50.6%); whereas the most frequent codes in the control condition were Management (73.5%), Health Plan (60.7%), and Group Plan (52.1%); see Appendix 5 for the coding manual. These results show that the mindset and control materials evoked different responses from students, verifying the effectiveness of our experimental manipulation. We explore two additional findings from our thematic analysis below.

Finding 1: generative practice is critical to growth. The reflection data suggest that while flexible intelligence is the core concept of mindset interventions, information about effective study techniques may contribute to their effects. As noted above, the code Generative Practice Problems (79.6%) was assigned to mindset responses more frequently than the code Brain Active (58.6%), making it the primary theme to emerge from mindset participants' reflections. This pattern is not necessarily surprising, because the reflection prompt directly asked students to explain how the article's ideas might influence their study habits. In addition, the mindset article explicitly informed students that brain growth requires "good strategies," stating that "You actually have to practice the right way—and usually that means the hard way—to get better at something." Nonetheless, this finding demonstrates the centrality of the generative-study message in participants' experience of the growth-mindset intervention, and it strengthens the argument that such interventions should encourage effective study strategies and not just hard work (Yeager and Dweck, 2012; Yeager *et al.*, 2016a).

The following reflection excerpt illustrates how a strategically-timed mindset intervention like ours, which was synced to the exam schedule, can prompt students to re-assess their study strategies at critical points in the semester:

"For the previous exam, I worked through problems that were not that much of a challenge for me. This resulted in a false sense of confidence and mastery of the topics. For the upcoming exam, I plan on challenging myself by attempting difficult problems and concepts, to simulate the harder and more conceptual problems on the exam. These harder problems will allow me to 'strengthen' my brain through more vigorous and stimulating exercises."—URM female, mindset condition

At least for some participants, the mindset effect seems to depend on not only a belief in flexible intelligence, but also an understanding of how to capitalize on that flexibility. In the example above, the student's ultimate goal is to strengthen her brain and chemistry skills so that she can earn a higher exam grade. However, most of her reflection describes her gradual discovery of the best way to achieve that goal—namely, generative practice problems. This response exemplifies how a growth mindset can help students respond resiliently and productively to failure (Yeager and Dweck, 2012), shifting their focus from academic performance itself to the learning processes that underlie performance.

Other examples similarly attest to the connection between a growth mindset and engagement with challenging practice:

"I feel that challenging myself with hard problems, while it may be frustrating at times, will ultimately make me a better student because I will be more prepared for whatever is thrown at me on the actual exam."—White male, mindset condition

This response succinctly conveys how mindset interventions can benefit students by changing the meaning they ascribe to effort expenditures, making them more willing to engage in challenging practice. Based on the prominence of the Generative Practice Problem code in the mindset condition, as well as the illustrative quotations above, we argue that discussion of effective study is an important component of successful mindset interventions.

Finding 2: a strong, positive control condition. Thus far, students' written responses have provided a manipulation check and revealed the distinctive themes that might underlie the mindset effect on students' performance. However, the responses also indicate that some themes were shared across both experimental conditions, albeit to different extents. For instance, the mindset intervention prompted some reflection on the control theme of time management, such that 25.8% of mindset responses were assigned the code Management. We speculate that this pattern arose because the mindset article implicitly addressed the issue of time management through its emphasis on consistent, incremental practice. The following excerpt supports that interpretation:

"I think that continuous exercise of my chemistry knowledge by doing problems every day, instead of just in the day before the test, will help improve my performance on the exam."—URM male, mindset condition

When this student mentions "continuous exercise" of his knowledge, he is alluding to an analogy presented in the mindset article, which stated that regular mental exercise can strengthen the brain just like regular training strengthens a weight lifter's muscles. Thus, despite the mindset article's focus on increasing brain connections and intelligence, it also conveyed the importance of time management by encouraging incremental study rather than cramming.

The overlap between conditions ran in the opposite direction as well, with the control article prompting some reflection on mindset themes. In particular, the control intervention led students to discuss the key mindset theme of generative study, though for potentially different reasons. While the mindset intervention recommended generative practice as a means of growing your brain and intelligence, the control intervention took a more pragmatic approach. Specifically, it argued that active course participation throughout the term would save time and trouble leading up to the exam. The following reflection excerpts show how this practical reasoning might have influenced control participants:

"I also plan on asking myself questions during class, because right now it is more as though I am simply writing down what they say."—URM male, control condition

"I am trying to understand the material, not memorize a step-by-step process, so I can apply the knowledge to any problem I am given."—White female, control condition

According to these responses, the control article led some students to reassess their approach to General Chemistry 1, shifting away from memorization towards more integrative and abstract learning. The data in Table 3 complement these excerpts, demonstrating that 20.3% of control responses received the code Generative Practice Problems, and 43.8% were marked for Other Generative Techniques. Given that abstraction-based learning has been shown to produce better performance in chemistry courses than rote learning (Frey *et al.*, 2017), the reflection data suggest that we succeeded at designing a strong, positive control condition that might support students' performance in General Chemistry 1.

General discussion

The overarching goal of this investigation was to experimentally test whether social-psychological interventions like the growth-mindset intervention can improve the first-year college-chemistry achievement of women and minorities, who remain nationally underrepresented in this and other STEM fields in the U.S. (National Science Foundation, 2017). First, we examined the final-exam scores of control participants in General Chemistry 1, predicting that such diagnostic course assessments would trigger feelings of stereotype threat and poorer performance among underrepresented groups (Steele and Aronson, 1995; Good *et al.*, 2008). We observed a racial disparity, such that minority students received lower scores than their white peers, which remained reliable even when measures of high-school academic preparation (ACT Math, AP Proportion) were taken into account. However, our sample revealed no significant effect of sex of students' final-exam scores. The absence of a sex-based disparity among controls may reflect the equal representation of both sexes in this course; if anything, females are slightly overrepresented. This interpretation parallels previous research on gender-based gaps, which has shown that gender gaps in STEM achievement tend to be concentrated in male-dominated contexts (Walton *et al.*, 2015). In general, these results support the idea that situational cues like physical under-representation play a key role in determining whether students experience stereotype threat in an educational context (Murphy *et al.*, 2007).

To rectify potential achievement gaps, we administered a three-part, online growth-mindset intervention using a random assignment experimental design. Although previous chemical-education research has shown the impact of social-psychological factors on students' success in general chemistry (Ferrell and Barbera, 2015; Chan and Bauer, 2016; Ferrell *et al.*, 2016), ours is the first study to directly manipulate a psychological process in order to facilitate students' transition into college-level chemistry. We predicted that our course-based mindset intervention would selectively benefit females and underrepresented minority students, given the additional psychological stressors they might face in quantitative achievement domains like chemistry. However, results from both the

ANCOVA and regression analyses showed a selective mindset effect among minority students only. After adjusting for academic preparation, the mindset intervention eliminated the racial achievement gap in General Chemistry 1: while minority students earned lower final-exam scores than white students in the control condition, both groups achieved comparable scores in the mindset condition. Although the absence of a mindset effect among females contradicts our predictions, it makes sense in light of the finding that females performed at the same level as males in General Chemistry 1. Such results are consistent with previous research portraying growth-mindset interventions as a method for mitigating stereotype threat and other disadvantaging forces (Aronson *et al.*, 2002; Good *et al.*, 2003; Yeager *et al.*, 2016b). If no threat or achievement gap is apparent, then the mindset treatment may have no effect. In other words, mindset interventions are not necessarily a panacea that will boost all students' outcomes, but a strategy specifically for helping underrepresented or marginalized students reach their potential. This conclusion aligns with a recent meta-analysis of mindset interventions: a broad survey of the literature revealed generally weak mindset-intervention effects, while also supporting the claim that such interventions may provide a significant benefit to targeted groups of underserved or at-risk students (Sisk *et al.*, 2018).

Our third research question examined the longevity of potential mindset effects. Based on previous studies (Aronson *et al.*, 2002; Yeager *et al.*, 2016b), we expected that any mindset effects observed in General Chemistry 1 might carry over to the second semester course, General Chemistry 2. Our results did not conclusively support or contradict that prediction, because their significance depended upon the statistical modeling technique being used. While the ANCOVA showed a significant overall effect of intervention condition on students' General Chemistry 2 scores, such that mindset participants outperformed control participants, that finding did not replicate in the regression analyses. Thus, more evidence is needed to determine whether brief social-psychological interventions are likely to have a lasting effect on underrepresented students' transition into and navigation through college chemistry.

Finally, thematic analysis of students' reflections during the intervention deepened our understanding of the performance results. A scheme of 21 qualitative codes emerged, and their distributions in the mindset and control conditions confirmed that students reflected heavily on the messages of their assigned articles. A substantial portion of mindset responses focused on strengthening chemistry intelligence, especially through generative practice problems, while many control responses focused on time management and work-life balance. The mindset participants' shared focus on brain growth and generative study suggests that multiple components of the mindset intervention may contribute to its performance effects. We elaborate on this idea and integrate it with the mindset literature below.

At the same time, the reflection data revealed some points of overlap between conditions, such that mindset participants discussed time management and the control participants considered generative study techniques. Such evidence reinforces our claim that the Transition Tips control article provided a strong control for the mindset intervention. Specifically, the prevalence of the codes Generative Practice Problems and Other Generative Techniques among control responses indicates that the control article encouraged students to endorse more effective, retrieval-based study strategies (Roediger and Butler, 2011), similar to the mindset intervention. Based on such data, we infer that the Transition Tips might have boosted control participants' performance, thereby reducing the performance difference between conditions. Accordingly, our experiment may have underestimated the potential mindset effects that would obtain when the intervention is implemented in chemistry courses that focus on teaching chemistry content, without providing tips for how to study or learn.

Designing a mindset intervention

We incorporated several novel features into our experimental design, which may guide other researchers and educators interested in implementing mindset interventions. For instance, the reflection prompts in our mindset intervention targeted students' domain-specific views of chemistry intelligence, as opposed to their beliefs about general intelligence. Specifically, they prompted students to discuss how the mindset (or control) concepts would influence their studies for upcoming exams in General Chemistry 1. We adopted this approach because of recent evidence showing that domain-specific views about mathematics intelligence influenced students' math grades (Good *et al.*, 2012; Rattan *et al.*, 2012). Our results validate this design, demonstrating that a domain-specific growth-mindset intervention can mitigate achievement gaps in the targeted subject. Because the domain-specificity of our intervention hinges primarily on the reflection writing prompts, it could be easily adapted to other fields (see Appendix 2). Indeed, we plan to integrate a physics-specific version of this growth-mindset intervention into Introductory Physics at our institution in the coming academic year. Despite the success of our domain-specific design, several issues remain open for future investigation. As noted by Dweck and colleagues (Yeager and Dweck, 2012), it remains an empirical question whether growth-mindset interventions manipulating general *versus* domain-specific mindsets about intelligence are more advantageous to students' performance in the targeted field. An experimental study directly comparing the two approaches would add nuance to the mindset literature and provide important evidence regarding the most effective pedagogical practice.

A distinct but related question explores whether general *versus* domain-specific mindset interventions differ in the breadth or persistency of their effects. In terms of breadth, one might reasonably predict that a general mindset intervention would have wider-ranging effects (*i.e.*, effects across various domains) than a domain-specific intervention, but there is little evidence addressing this question. The current study did not test for effects of the chemistry-specific mindset intervention in other STEM subjects, and previous mindset studies at the collegiate level examined composite outcome variables like cumulative GPA and full-time enrollment rates (Aronson *et al.*, 2002; Yeager *et al.*, 2016b), rather than testing for effects in different domains. In terms of persistency, previous work using a general intervention (Yeager *et al.*, 2016b) has shown mindset effects that extend from the beginning to the end of an academic year. Longitudinal studies that examine mindset participants' academic performance across multiple years or explores longer-term outcomes like enrollment in graduate school would shed light on the limits of mindset effects. Such information is critical for educators and policymakers who are considering social-psychological interventions as a measure for counteracting the underrepresentation of women and minorities in STEM.

Another unique aspect of our mindset design was the syncing of our multi-phase intervention with students' exam schedule. Students were exposed to the intervention materials shortly before the first unit exam (phase 1); their first reflection occurred right before the second unit exam (phase 2); and their final reflection preceded the cumulative final exam (phase 3). We developed this procedure to engage students with the mindset concept at critical points in the semester, *i.e.*, while preparing for high-stakes assessments, in the hopes that a growth mindset would encourage them to adopt more challenging and effective study strategies. Recent evidence supports the idea that pre-exam reflections can improve students' study techniques and subsequent course performance. For example, Chen and colleagues (2017) found that students who reflected on academic resource usage prior to each exam not only used those resources more effectively, but also earned higher course grades compared to controls. Further investigation is needed to determine whether our strategically-timed, multi-phase intervention has differential effects than previous procedures. Nonetheless, the positive performance results and students' written reflections provide support for our approach, and we suggest that this design may help educators to integrate a mindset intervention into their course structure.

Multiple sources of mindset effects

Yet another distinctive feature of this study is our qualitative analysis of students' written reflections during the intervention. The reflection data revealed that mindset participants discussed a range of different themes, with Generative Practice Problems being the most prominent. Based on these findings, we argued that mindset effects require not only a belief in flexible intelligence and brain growth, *i.e.*, a growth mindset *per se*, but also an awareness of how to exercise that potential through effective study (*Finding 1*). While this hypothesis complicates the explanation for mindset effects and deserves further scrutiny, we contend that it fits well with the existing literature on growth mindsets (see also Sisk *et al.*, 2018).

Since the seminal work of Dweck and Leggett (1988), students' mindsets about intelligence have been described as the foundation of a cognition-affect-behavior chain that impacts their performance in academics and other domains. When compared to students with fixed mindsets, students with growth mindsets are more likely to (a) set mastery-oriented learning goals, (b) view challenges in a positive light, and (c) develop adaptive strategies for overcoming obstacles. Therefore, the observation that mindset participants reflected on a range of themes, including their learning goals (*e.g.*, strengthening their brains) and their strategies for improvement (*e.g.*, challenging practice), seems to indicate that different participants focused on different stages of the cognition-affect-behavior chain that reflects a growth mindset.

This interpretation integrates the current study with prior research, and it raises some new questions as well. For instance, although our data demonstrate that a performance benefit can arise even when participants reflect on various aspects of the mindset framework (*i.e.*, implicit beliefs *vs.* course-related behaviors), are there any differences in the effects of these distinct reflection foci? Previous work has argued the importance of targeting mindset interventions at the earliest stage of the cognition-affect-behavior chain, because a change to students' mindsets can have recursive, downstream effects on their affects and behaviors (Yeager and Walton, 2011; Walton, 2014). However, recent evidence shows that reflection on later stages of the cycle, *e.g.*, reflection on the behavior of academic-resource use (Chen *et al.*, 2017), can also boost performance. We hypothesize that differences may emerge between these two approaches, in terms of the longevity or generalization of their effects. Comparison of these intervention techniques may be a fruitful avenue for future research.

Limitations

It is important to acknowledge the limitations of the current investigation. Like any experiment conducted in a classroom setting, our results may have been influenced by the extremely complex course landscape in General Chemistry 1. At our institution, this course involves multiple instructors, mandatory recitations and labs, and optional supplements like PLTL and instructor-led help sessions. All of these factors may have impacted students' performance on the final exam, which was our dependent measure. For this reason, random assignment of participants to the mindset and control conditions was the foundation of our design. This method is the gold standard for ruling out confounding variables. Our random assignment procedure deviated somewhat from the standard approach, because we excluded participants from the experimental sample after all students enrolled in General Chemistry were randomly assigned to a condition. However, tests confirmed that the two experimental groups were comparable in all regards, at least for the variables we assessed. Further, the instructors were blind to the assignment of students to their experimental groups, and students' decisions to participate in the study by providing consent happened in the absence of knowledge about their assignments. Consequently, we maintain that the use of random assignment greatly strengthens our claim that the mindset intervention caused the selective performance improvement that we observed among underrepresented minority students.

Although our results support a causal link between our experimental manipulation and performance, we cannot pinpoint with certainty the mechanism(s) underlying that effect. We have argued throughout the Results and discussion sections that multiple mechanisms may come into play, because students' mindsets about intelligence are interconnected with their affect and behaviors related to a course (Dweck and Leggett, 1988). One possibility is that simply adopting a growth mindset is sufficient to induce a performance effect, though the results of Sisk *et al.*'s (2018) recent meta-analysis do not support this conclusion. Alternatively, adoption of a growth mindset may be insufficient—a concomitant reduction of anxiety or stereotype threat may be needed to boost the exam scores of disadvantaged students. In other words, adoption of a growth mindset may have an indirect effect on performance, mediated by accompanying changes in affect. Yet another possibility (though not mutually exclusive) is that the success of mindset interventions may hinge on their stimulating more effective and appropriate strategies for learning the target materials. These forces could vary across participants, as suggested in the previous section, or they could co-occur within individual students. The current study is limited in the sense that it does not adjudicate amongst these mechanistic explanations.

Another issue is the potentially limited generalizability of this study. On one hand, our results expand the evidence that mindset interventions can support students transitioning into college, especially if those students belong to underrepresented groups that may be subject to stereotyping. On the other hand, these results were generated within a very specific educational context, *i.e.*, General Chemistry 1 at a private university, and using a very specific mindset intervention, *i.e.*, a multi-phase, course-based, domain-specific treatment. Moreover, we found that our experimental sample differed from some of the other students enrolled in the General Chemistry course sequence. Specifically, students who consented but never actively participated in our study were more likely than sample participants to be male, and they were less likely to participate in the Departmental resource of PLTL. We might expect that underrepresented students in this group of non-participants would indeed benefit from a growth mindset intervention; for example, it might induce them to seek more guidance from peers and instructors regarding effective study strategies. Nonetheless, it remains an empirical question whether our results would extend beyond our sample. As a result, we must exercise caution in concluding that mindset and other social-psychological interventions will offer benefits to underrepresented or at-risk students across the board. Indeed, one recent study challenges that idea, demonstrating that the incentive system used in a classroom influences the effects of a growth mindset intervention (Chao *et al.*, 2017). That study also explores the broader question of whether the mindset literature to-date, which predominantly investigates mindset interventions in Western educational contexts, extends across cultures and nations. Such work illustrates the importance of assessing one's local educational context before implementing a mindset intervention, rather than treating it like a "magic bullet" (Yeager and Walton, 2011).

Finally, we recognize that our study oversimplifies the issue of underrepresentation in chemistry and STEM. First, the demographic variables we utilized are coarse-grained, potentially masking important facets of students' identities and therefore variation in their experiences of introductory STEM classrooms. For example, we have collapsed across diverse ethnic backgrounds by focusing on racial groups. The usage of a binary sex variable was beyond our control, but it also has limitations. This approach was not inclusive of intersex individuals. It runs the risk of conflating biological sex with the social construct of gender identity; the latter construct has been emphasized in previous work. We hope that by consistently using distinct terms for these concepts, we have highlighted the similarities and differences between the current study and previous research. Second, by focusing narrowly on phenotypical attributes like sex and race, we have painted an incomplete picture of the identity threats that students' may face. As noted in the introduction, many aspects of a students' social identities can influence their acceptance and success in the scientific community (Carlone and Johnson, 2007) and their feelings of cultural alignment with that community (Brown *et al.*, 2016). This may include socioeconomic status and first *versus* continuing generation status, among other things, which are known to influence students' wellbeing and achievement in college (Jury *et al.*, 2015, 2017). Incorporation of these variables into our classroom study was not feasible, because it would have created even smaller comparison groups and reduced our power to detect statistical differences. Nonetheless, we wish to highlight the complexity of fostering diversity and inclusion in STEM, and we encourage educators to think broadly about which students might be disadvantaged at their institution and might benefit from a growth-mindset or other social-psychological intervention.

Conclusions

In a randomized classroom experiment, we demonstrated that a course-based, chemistry-specific, growth-mindset intervention can improve the General Chemistry performance of first-year college students in the U.S. who come from backgrounds underrepresented in STEM. Because we observed a selective benefit among underrepresented minority students but not females, who were well-represented in General Chemistry 1, our results support the idea that mindset interventions will boost performance specifically among students who are disadvantaged in their educational context. We therefore encourage researchers and educators to examine historical data in order to determine whether such interventions are appropriate for their target course. If patterns of identity-based underachievement are apparent, then strategic incorporation of a domain-specific growth-mindset intervention into the course structure might ease the college transition for at-risk students. We view such techniques as complementary to other strategies that have emerged from the chemical education literature, which support students' early achievement in general chemistry *via* supplemental academic training and more interactive courses. In a sense, tools like the mindset intervention offer a relatively low-cost, easy-to-implement strategy for filling in any cracks in the current support systems available to chemistry students. By providing students with a suite of resources that address not only their academic skills, but also their motivation for and perceptions of learning, chemical educators can foster success among diverse students and increases their chances of persisting in the field.

Conflicts of interest

There are no conflicts to declare.

Appendix 1: intervention stimuli

For the growth-mindset intervention, we utilized the same article as Yeager and colleagues (2016b), who recently tested a growth-mindset intervention among other populations of first-year college students (*i.e.*, recent charter high-school graduates and public university students, as opposed to private university students). The article is entitled ''You Can Grow Your Brain,'' and it is closely modeled on the article ''You Can Grow Your Intelligence'' (Blackwell *et al.*, 2007), which has been utilized in other growth-mindset studies. For the control condition, we designed the ''Transition Tips'' article presented below. With this article, we sought to establish a control that might genuinely help first-year students with the academic transition to college-level STEM courses. We also aimed to match the mindset materials in length and substance, so that all experimental participants performed a similar amount of work to earn the extra credit points for study completion. Participants read their assigned articles in full during phase 1 of the intervention. During phases 2 and 3, they were provided with a summary of their article's key points, as well as link to the original article, before reflecting on how the materials would influence their exam study. Both the key points and the reflection prompts are included below.

Transition tips article

''College courses can cover difficult material and require heavy workloads, but the greatest obstacles to college success are unrelated to the content of any particular course. Students must navigate the freedoms and responsibilities that accompany the college lifestyle and successfully budget their time, budget their money, handle their responsibilities, and take care of their physical and mental health. All of these things are easier said than done, but keeping a few important points in mind can make you a more successful college student.

Get organized

Staying on top of everything is a significant challenge. Some courses have frequent deadlines, and keeping track of these deadlines for multiple courses can be challenging. Conversely, some courses have very few deadlines and base grades on only a few big exams or projects. These courses present different but equally formidable challenges, as the temptation to coast during some weeks can create nearly unmanageable weeks later in the semester when you have to play catch-up with the course material, deal with multiple exams or large projects, and also completing your normal weekly requirements.

Make things easier on yourself by making a plan early in the semester before the workload cranks up and sticking to it the best you can. Become an expert on course requirements, due dates, and exam dates. Think about each assignment and exam, estimate the time you'll need, and realize that your initial estimates will probably underestimate the actual time. Due dates and exams can come in bunches, so look for trouble spots and make plans to deal with them (*e.g.*, finish a paper a week early, so that you can spend more time the following week studying for upcoming exams). Make a general plan for how you'll navigate the semester and then make weekly plans for what you'll accomplish each week to meet your semester-long goals. Make sure your plans include spending time on each course, even if some courses have no deadlines. You can focus the majority of your effort on certain courses in a given week, but make never to neglect any of your courses entirely. Even a small amount of effort, such as doing assigned reading and attending class, can keep you up to speed and save you the unnecessary effort required to ''catch up'' in subsequent weeks.

Maintain your health and balance

Although academics should be a focal point of college, sustained success also requires that you attend to other important aspects of your well-being. No matter how busy you get, never neglect to do the following things.

Be social. Even in your busiest of weeks, make sure you take study breaks to do something social, even if it's just to have a 15-minute coffee break with your friend.

Take some time for yourself. Alone time can be scarce on a college campus, but every week, no matter how busy you are, find a little time to do something non-academic by yourself that doesn't involve a TV, computer or phone. Whether it's walking through the park or working on a hobby, your mental health will benefit.

Eat healthy and exercise. A certain amount of pizza and ice cream may be unavoidable, but a diet without fruits, vegetables, and healthy proteins will sap your energy, increase the chance of illness, and potentially impair your academic performance. Allow yourself treats, but eat a balanced diet, and don't go to class on an empty stomach. Also, make sure to exercise. Even if you don't like the gym, find ways to stay active, whether it's walking, playing intramural sports, or using a bike as your primary form of transportation. Staying physically active will increase your health and energy, and it may help keep your mind sharp as well.

Sleep. Sleep not only impacts energy levels and physical health, but it is also a key ingredient to building lasting memories. In order for your body and mind to work at their full potential, you must get sufficient sleep. This is easier said than done, but following the advice in the 'Get Organized' section will reduce your need for late-night study sessions and make 7–8 hours of sleep per night more attainable.

Be a consistently active participant in your courses

Save yourself later headaches and study time by actively participating in your courses on a consistent basis. At the very least, attend every class period. Even that little bit will help keep you engaged with the course material and save you later study time. Make class time even more useful by reading the material before class and coming armed with a few questions or insights you pulled from the material. This will make you more comfortable participating in class and keep you engaged, which in turn will help you learn and remember more from each class period. Always keep in mind that the learning you do in class can save you time studying for the exam.

Make sure you complete all homework, when it is due, even if it is worth minimal points. And do not cut any corners on your homework. If an instructor has assigned homework, he or she thinks it is important that you practice with this material, and you should put the effort into trying to learn what the instructor wants you to learn from it. If an instructor or TA provides feedback on homework, study and understand this feedback, rather than just looking at your grade. This feedback is another route to learning, and the learning you do while completing the homework and looking over the feedback can ultimately save you later study time.

Use available resources

Other students, TAs, and instructors can be great resources. Some courses require group projects, but if not, try to make friends with others in class and form study groups. At the very least, those students can help catch you up if you have to miss class. Further, a study group that meets consistently (*e.g.*, once a week) keeps you accountable and more likely to keep up with the material. Also, members of a group can help teach each other. Group homework sessions can expose students to different perspectives or strategies they can add to their toolkits. Also, if the material contains several difficult concepts, each group member can focus effort on understanding one of the concepts, and then the members can teach each other.

Also keep in mind that the instructor can be a great resource. Don't just interact with the instructor in class, but also find an excuse to go to office hours/help sessions and introduce yourself personally to the instructor, as early in the semester as possible. That way, if you start to struggle or otherwise want to talk to the instructor about something, you'll be more comfortable contacting him or her. If you are doing everything you can but still feel that you are struggling in a course, contact the instructor as soon as you can. Instructors likely have too many students to keep track of which students might need help, but they likely are willing to help students who ask. Asking for help early in a semester can save unnecessary weeks or even months of stress and poor performance.

At times, college can be difficult for almost everybody. While no amount of tips or tricks can make college easy, the advice provided here can be implemented by anyone and, if followed, can make you a more successful student and allow you to experience more of the learning, growth, and fun that the college experience has to offer.''

Appendix 2: key points and reflection prompts from intervention phases 2 and 3

Growth mindset key points

''Recall the article that you read previously about learning and the brain. Here are some of the key points:

- The brain is like a muscle because you can strengthen it through exercise.
- Studies have shown that animals who live in challenging environments have more and stronger brain connections than animals who live in less challenging environments.
- Studies with humans have shown that practicing a skill (such as juggling) not only leads to better performance but also growth in associated brain areas.
- To maximize brain growth, people should focus their effort on solving new and challenging problems.''

Control key points

''Recall the article you read previously that provided tips for college. Here are some of the key points:

- You should make a general semester-long plan as well as weekly plans and make sure your weekly plans do not neglect any of your classes.
- It is important to be social and stay healthy by exercising, eating a balanced diet, and getting adequate sleep.
- While in class you should think through the material and ask yourself questions rather than just taking down word-for-word notes.
- Finding a study group is one of the most helpful things you can do in a class.''

Phase 2 reflection prompt

With this article in mind, explain in a brief paragraph how these ideas will influence the way you'll prepare for the upcoming General Chemistry 1 exam.

Phase 3 reflection prompt

With this article in mind, explain in a brief paragraph how these ideas will influence the way you'll prepare for the General Chemistry 1 final exam, which is a comprehensive exam.

Appendix 3: descriptive statistics by race and sex

Table 4 Final-exam scores for the sample of General Chemistry 1 participants

Condition	Race	Sex	n	Unadjusted mean (SE)	Adjusted mean[a] (SE)
Mindset	URM	F	36	60.4 (2.4)	64.3 (2.2)
		M	29	61.9 (2.6)	64.9 (2.4)
	White	F	119	66.1 (1.3)	65.9 (1.2)
		M	91	65.8 (1.5)	63.8 (1.4)
Control	URM	F	47	52.9 (2.1)	58.2 (1.9)
		M	24	58.0 (2.9)	59.9 (2.6)
	White	F	120	66.1 (1.3)	65.0 (1.1)
		M	99	67.1 (1.4)	65.3 (1.3)

Note: SE indicates standard error. [a] Adjusted for the covariates ACT math and AP proportion. For students who reported SAT instead of ACT math scores, concordance tables (Dorans, 1999) were used to convert those data. AP proportion reflects students' performance on the advanced placement exams for 4 STEM subjects: biology, calculus, chemistry, and physics. For each exam where a student earned a score of 4 or 5 (out of 5), their AP proportion score increased by 0.25.

Table 5 Exam averages for the sample of General Chemistry 2 participants

Condition	Race	Sex	n	Unadjusted mean (SE)	Adjusted mean[a] (SE)
Mindset	URM	F	31	60.3 (2.6)	64.2 (2.4)
		M	26	63.3 (2.9)	66.5 (2.6)
	White	F	109	66.5 (1.4)	66.4 (1.3)
		M	80	69.8 (1.6)	67.8 (1.5)
Control	URM	F	42	54.2 (2.3)	59.1 (2.1)
		M	23	58.8 (3.1)	60.6 (2.8)
	White	F	108	67.4 (1.4)	66.3 (1.3)
		M	87	66.9 (1.6)	67.8 (1.4)

Note: SE indicates standard error. Exam average is calculated by combining students' top 2 (of 3) unit-exam scores and their final-exam score. [a] Adjusted for the covariates ACT math and AP proportion. For students who reported SAT instead of ACT math scores, concordance tables (Dorans, 1999) were used to convert those data. AP proportion reflects students' performance on the advanced placement exams for 4 STEM subjects: biology, calculus, chemistry, and physics. For each exam where a student earned a score of 4 or 5 (out of 5), their AP proportion score increased by 0.25.

Appendix 4: multiple regression results

Table 6 Multiple regression predicting General Chemistry 1 final-exam scores

Dependent variable	Data	Predictors[a]	b	β	SE	t-value	R^2
General Chemistry I final-exam (z-scored)	Sample (n = 565)	Intervention	0.45	0.23	0.19	2.35*	0.24
		Race	0.43	0.19	0.16	2.76*	
		Sex	0.13	0.06	0.22	0.57	
		ACT Math	0.22	0.25	0.02	6.14*	
		AP Proportion	0.98	0.28	0.14	7.01*	
		Intervention × Race	−0.37	−0.18	0.23	−1.65	
		Intervention × Sex	−0.10	−0.04	0.31	−0.34	
		Race × Sex	−0.07	−0.03	0.25	−0.27	
		Intervention × Race × Sex	−0.06	−0.02	0.35	−0.17	
	URMs only (n = 136)	Intervention	0.41	0.19	0.17	2.45*	0.26
		Sex	0.05	0.02	0.17	0.31	
		ACT Math	0.12	0.31	0.03	3.81*	
		AP Proportion	1.03	0.25	0.34	3.01*	
	Whites only (n = 429)	Intervention	0.01	0.01	0.08	0.12	0.17
		Sex	−0.01	−0.01	0.08	−0.17	
		ACT Math	0.10	0.22	0.02	4.66*	
		AP Proportion	0.96	0.29	0.15	6.36*	

Note: b = unstandardized regression coefficient, β = standardized coefficient, SE = standard error of b, and * indicates $p < 0.05$. [a] Categorical predictors were treatment coded, and the reference levels for intervention, race, and sex were the control condition, URMs, and females, respectively.

Table 7 Multiple regression predicting General Chemistry 2 exam averages

Dependent variable	Data	Predictors[a]	b	β	SE	t-value	R^2
General Chemistry II exam average (z-scored)	Sample (n = 506)	Intervention	0.27	0.15	0.19	1.43	0.25
		Race	0.48	0.23	0.15	3.21*	
		Sex	0.13	0.07	0.21	0.61	
		ACT Math	0.08	0.21	0.02	4.89*	
		AP Proportion	0.98	0.31	0.13	7.33*	
		Intervention × Race	−0.30	−0.16	0.22	−1.36	
		Intervention × Sex	0.03	0.01	0.30	0.10	
		Race × Sex	−0.23	−0.12	0.24	−0.97	
		Intervention × Race × Sex	0.14	0.06	0.34	0.41	
	URMs only (n = 122)	Intervention	0.26	0.13	0.16	1.62	0.25
		Sex	0.11	0.05	0.16	0.65	
		ACT Math	0.13	0.36	0.03	4.17*	
		AP Proportion	0.70	0.19	0.32	2.19*	
	Whites only (n = 384)	Intervention	0.04	0.02	0.08	0.51	0.18
		Sex	−0.01	−0.01	0.08	−0.15	
		ACT Math	0.06	0.15	0.02	2.98*	
		AP Proportion	1.06	0.35	0.15	7.28*	

Note: b = unstandardized regression coefficient, β = standardized coefficient, SE = standard error of b, and * indicates $p < 0.05$. [a] Categorical predictors were treatment coded, and the reference levels for intervention, race, and sex were the control condition, URMs, and females, respectively.

Appendix 5: qualitative coding manual

Table 8 Coding manual for thematic analysis of students' free responses during the intervention phases 2 and 3

Category	Code	Description	Example of student response
Brain	Brain Active	The student refers to the brain (or mind) in an active way; the student describes *doing something* to cause brain growth, stretching brain, growing connections, *etc.* • Even if we do not agree that the student's "doing something" technique is beneficial/active, the student is still scored "1" for Brain Active if they mention the brain in the context of *doing something*.	"**I will work challenging problems because that will help my brain get stronger at chemistry**." "While reading this article I learned that **the brain gets stronger when you use it, and the more you challenge your mind to learn, the more your brain cells grow. In other words, the brain grows more when you learn something new and hard. I can apply this information to preparing for the next Chem 111 exam by understanding the material in ways that I haven't before and by practicing challenging problems. Relearning things in new ways and challenging myself should help strengthen my brain and hopefully result in a good test score.**"
	Brain Passive	The student refers to the brain in a passive way; the student describes brain growth, stretching, connection growth, *etc.* as just happening from, for example, being in a challenging environment at Wash U.	"I will exercise my brain lots by re-watching lectures, creating quizlets, retaking all the quizzes and reworking all the homework to strengthen my brain to maximum capacity before the exam. **Chem 111 is a challenging environment, and I will therefore have more and stronger brain connections than humans not in Chem 111.**" * Score "1" for Brain Active and "1" for Brain Passive for this response, as it has aspects of each category.
Study Techniques	Generative Practice Problems	The student seeks out challenging, novel practice problems to complete as part of their exam study. This includes: • Doing the practice test • Doing the quizzes from the other sections • Redoing old problems that the student specifically picks because they are challenging • Reviewing old problems or exams specifically to understand areas of weakness	"I will start studying early to enhance my performance on the exam. I will also go through the notes and repeat exercises many times to strengthen my brain. Not only will I repeat exercises, **I will also try new challenging problems to maximize my brain growth**. These can be the non-graded problems." * Score "1" for Generative Practice Problems and "1" for Repetitive Practice Problems for this response. "**I will prepare for the exam by going through my previous tests and seeing where I went wrong to try and capitalize on those mistakes**"

Table 8 (*continued*)

Category	Code	Description	Example of student response
	Repetitive Practice Problems	The student states that they will do as many problems as possible, redo or review old practice problems (including old exams, quizzes, PLTL or POGIL packets) in a non-targeted fashion, or do easy problems.	"To prepare for the Chem 111 final exam, **I will work many practice problems** and attend all available help sessions. This way, **I am practicing these concepts repeatedly**" "I will make sure to go to every help session, review my mistakes from my previous exams, make a study outline, **take the practice exam multiple times**, and get enough sleep before the exam." * Score "1" for both Generative Practice Problems and "1" for Repetitive Practice Problems
	Other Generative Techniques	The student states that they will engage in other study techniques (*i.e.*, not practicing problems) that are considered generative. Some examples of these techniques include: re-writing class notes, taking notes on lecture videos, or practicing connecting between concepts. • Asking questions <u>does not</u> count unless the student elaborates on how the process helps them.	"**I will try to connect the different ideas throughout the semester and relate them to each other.**" "During my study time for chemistry, **I will rewrite my notes and give myself the opportunity to think through every topic**." "The best thing for me is doing practice problems and **then talking to someone if I get them wrong about what I did that was different.**" <u>Mark "0"</u>: "I have reached out to my friends, and we plan on forming a study group that will meet after every Chem final exam review. By doing so, we can review all of the material together and **ask any questions**."
	Other Repetitive Techniques	The student states that they will engage in other study techniques (*i.e.*, not practicing problems) that are considered repetitive. Some examples of these techniques include: reading class notes, watching lecture videos, memorizing formulae or concepts.	"In order to prepare for my exam **I will be reviewing my notes, reviewing lectures I don't understand,** and looking over past tests." "I will prepare for my final **exam by repeatedly reviewing the topics in videos** or review sessions with other teachers and professors."
Group Work	Group Do	The student is already engaging in group work (already "doing" group work).	"**I will study with my study group**..."
	Group Plan	The student plans to engage in group work.	"Personally, my best study strategies include working through old problems and exams **and then going over other problems and ideas with a study group.** I find it best to bounce ideas off of other people and see the different ways to solve similar problems." * This response is tricky to code for group work as it is a bit unclear if the student has a consistent study group, or if the student just forms a study group right before the exam. For this reason, I would score a "1" for Group Plan" and a "0" for "Group Do."
In Class	In-Class Connect Concepts	The student mentions actively connecting between concepts during class/lecture.	"Particularly when an exam is coming up, **during lecture I like to think of how I can tie each topic back to previous, exam-related topics**."
	In-Class Generative	The student describes engaging in a generative behavior during class. Some examples include: taking paraphrased notes, answering clicker questions, and discussing clicker questions with a partner.	"**While in class I make sure to learn from the clicker questions and examples as much as possible**..."
	In-Class Repetitive	The student describes engaging in a repetitive behavior during class. Common example: taking word-for-word notes.	"**Also, I disagree in not trying to take word for word notes. Notes help regardless: I believe that it's better-in general-to have more than enough than not enough.** Take notes, and think and ask questions."
Student Skills	Resource	The student mentions utilizing a resource that is specifically available at Wash U. This <u>does not</u> include PLTL/POGIL, as most students utilize those resources. This <u>does</u> include, but is not limited to: Residential Peer Mentoring hours, professor/TA-run help sessions, Cornerstone, and review sessions.	"I am going to every PLTL and **going to all the RPM hours** in order to get all my questions answered about any topic I am confused by."

Table 8 (*continued*)

Category	Code	Description	Example of student response
	Management	The student refers to scheduling their time (planning a schedule for: study, sleep, eating, socializing, doing work for other classes) or demonstrates that they are managing their student affairs.	"**I have also made a study plan for the next few days and this weekend so that I can divide all of the studying I need to do instead of cramming.** So far, I have kept with it this week. I am getting a fair amount of sleep and I am eating really healthy this week...." "**In terms of planning, I will start studying more than a week in advance** and review quizzes, PLTLs, POGILs and my notes"
Health	Health Do	The student mentions that they are *already doing* certain activities that benefit their health. These activities include, but are not limited to: sleeping an appropriate amount each night, exercising, maintaining a good diet, and mental-health care.	"**I take breaks from studying by going to the gym, getting meals with friends, and eating healthy snacks while studying, instead of eating processed foods. I also make sure to keep track of my caffeine intake. When it is late and I feel too tired to keep studying, I go to sleep and wake up early to continue studying before my first classes. Keeping myself sane helps me be prepared at any moment to study.**"
	Health Plan	The student mentions that they *plan* to do a certain activity to benefit their health, particularly to benefit their health right before an exam. These activities include, but are not limited to: planning to sleep a certain amount before the exam, planning to exercise before the exam, planning to eat well before the exam, planning to do something to prevent too much stress before the exam.	"**The night before the exam I try to get a good night's sleep.**" "**Furthermore, in order to be more prepared physically for the exam I will be sure to get plenty of sleep the night before and eat well the day of.**"
Self-Perception	Resilience	The student demonstrates that they are willing to keep working through challenges, respond productively to failure, or depend upon themselves to complete work (*i.e.*, not look at the answer sheet until after doing problems).	"**Chemistry has never been a topic that has come easy to me. I spend time with the material but struggle to piece together all of the information. I have to focus my efforts in solving new and challenging problems to maximize my brain growth.** I certainly am in a challenging in environment. **I have to try my best and that is all that matters. I can't beat myself up over not getting A's if I am putting in my best effort.**"
	Confidence	The student makes a statement about confidence.	"...**My confidence for the exam will rise, and maybe I'll learn something I missed.**"
Helpfulness	Already Knew Info	The student states that they already knew the information presented in the article they read.	"**Most of the advice, however, I already try to make habit, so this article serves as motivation to maintain the habits.**"
	Already Have Strategy	The student states that they already have a study strategy, especially if they feel that strategy works best for them.	"**I'm going to stick to the methods I've been doing.**"
	Helpful	The student either states or demonstrates that the article was directly helpful to them. The student states that the article will influence their studying or student practices.	"**These ideas will influence the way I will prepare for the upcoming Chem 111 exam by helping me improve my study habits.**"
	Not Helpful	The student either states or demonstrates (through their response) that the article was not helpful to them.	"**Actually, no, it didn't. It's Sunday night and I'm only starting now. My methods have always worked and will continue to work.**"

Note: "Chem 111" is the course number for General Chemistry 1. The descriptions for some codes are timing and context dependent. For example, the code Generative Practice Problems was applied if a student mentioned the practice test for exam 2 prior to that exam (*i.e.*, during phase 2), because its questions were necessarily novel and challenging at that point in time. However, prior to the final exam (*i.e.*, during phase 3), responses mentioning the practice tests for exams 1–3 were assigned the code Repetitive Practice Problems, because there was a high likelihood that students had already seen or worked those problems. A context-dependent example is the code Resource. At our institution, responses mentioning Peer-Led Team Learning (PLTL) were not marked for Resource usage, because the majority of students participate in PLTL as part of the General Chemistry course sequence. If this were an activity that a minority of students opted into, then it would qualify as a supplemental Resource.

Acknowledgements

This research was funded by an Undergraduate STEM Education Initiative grant from the Association of American Universities. We thank the General Chemistry instructors (Megan Daschbach, Jia Luo, and Gabriela Szteinberg) for their support and assistance in administering this study. We also thank Claudia Hendrick for her contributions to the qualitative coding.

References

ACT, (2018), About ACT, viewed February 2018, https://www.act.org.

Aronson J., Fried C. B. and Good C., (2002), Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence, *J. Exp. Soc. Psychol.*, **38**(2), 113–125.

Báez-Galib R., Colón-Cruz H., Resto W. and Rubin M. R., (2005), Chem-2-Chem: a one-to-one supportive learning environment for chemistry, *J. Chem. Educ.*, **82**(12), 1859–1863.

Bentley A. B. and Gellene G. I., (2005), A six-year study of the effects of a remedial course in the chemistry curriculum, *J. Chem. Educ.*, **82**(1), 125–130.

Blackwell L. S., Trzesniewski K. H. and Dweck C. S., (2007), Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention, *Child Dev.*, **78**(1), 246–263.

Botch B., Day R., Vining W., Stewart B., Hart D., Rath K. and Peterfreund A., (2007), Effects on student achievement in general chemistry following participation in an online preparatory course. Chemprep, a voluntary, self-paced, online introduction to chemistry, *J. Chem. Educ.*, **84**(3), 547–553.

Brown B. A., Henderson J. B., Gray S., Donovan B., Sullivan S., Patterson A. and Waggstaff W., (2016), From description to explanation: an empirical exploration of the African-American pipeline problem in STEM, *J. Res. Sci. Teach.*, **53**(1), 146–177.

Bunce D. M., Flens E. A. and Neiles K. Y., (2010), How long can students pay attention in class? A study of student attention decline using clickers, *J. Chem. Educ.*, **87**(12), 1438–1443.

Burnette J. L., O'Boyle E. H., VanEpps E. M., Pollack J. M. and Finkel E. J., (2013), Mind-sets matter: a meta-analytic review of implicit theories and self-regulation, *Psychol. Bull.*, **139**(3), 655–701.

Burns K. C. and Isbell L. M., (2007), Promoting malleability is not one size fits all: Priming implicit theories of intelligence as a function of self-theories, *Self and Identity*, **6**(1), 51–63.

Carlone H. B. and Johnson A., (2007), Understanding the science experiences of successful women of color: Science identity as an analytic lens, *J. Res. Sci. Teach.*, **44**(8), 1187–1218.

Chan J. Y. and Bauer C. F., (2016), Learning and studying strategies used by general chemistry students with different affective characteristics, *Chem. Educ. Res. Pract.*, **17**(4), 675–684.

Chang M. J., Sharkness J., Hurtado S. and Newman C. B., (2014), What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups, *J. Res. Sci. Teach.*, **51**(5), 555–580.

Chao M. M., Visaria S., Mukhopadhyay A. and Dehejia R., (2017), Do rewards reinforce the growth mindset? Joint effects of the growth mindset and incentive schemes in a field intervention, *J. Exp. Psychol.: Gen.*, **146**(10), 1402–1419.

Chen P., Chavez O., Ong D. C. and Gunderson B., (2017), Strategic resource use for learning: a self-administered intervention that guides self-reflection on effective resource use enhances academic performance, *Psychol. Sci.*, **28**(6), 774–785.

Cooper M. M., Cox Jr, C. T., Nammouz M., Case E. and Stevens R., (2008), An assessment of the effect of collaborative groups on students' problem-solving strategies and abilities, *J. Chem. Educ.*, **85**(6), 866.

Dika S. L. and D'Amico M. M., (2016), Early experiences and integration in the persistence of first-generation college students in STEM and non-STEM majors, *J. Res. Sci. Teach.*, **53**(3), 368–383.

Dorans N. J., (1999), Correspondences between ACT™ and SAT® I scores, *College Board Research Report 99-1*, New York: The College Board.

Dweck C. S., (2006), *Mindset: the new psychology of success*, New York: Random House Incorporated.

Dweck C. S. and Leggett E. L., (1988), A social-cognitive approach to motivation and personality, *Psychol. Rev.*, **95**(2), 256.

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

Ferrell B. and Barbera J., (2015), Analysis of students' self-efficacy, interest, and effort beliefs in general chemistry, *Chem. Educ. Res. Pract.*, **17**(2), 318–337.

Ferrell B., Phillips M. M. and Barbera J., (2016), Connecting achievement motivation to performance in general chemistry, *Chem. Educ. Res. Pract.*, **17**(4), 1054–1066.

Freelon D., (2013), ReCal OIR: ordinal, interval, and ratio intercoder reliability as a web service, *Int. J. Internet Sci.*, **8**(1), 10–16.

Freeman S., Eddy S. L., McDonough M., Smith M. K., Okoroafor N., Jordt H. and Wenderoth M. P., (2014), Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U. S. A.*, **111**(23), 8410–8415.

Frey R. F., Cahill M. J. and McDaniel M. A., (2017), Students' concept-building approaches: a novel predictor of success in chemistry courses. *J. Chem. Educ.*, **94**(9), 1185–1194.

Gafney L. and Varma-Nelson P., (2008), *Peer-led team learning: evaluation, dissemination, and institutionalization of a college level initiative*, Berlin: Springer Science & Business Media, vol. 16.

Good C., Aronson J. and Harder J. A., (2008), Problems in the pipeline: stereotype threat and women's achievement in high-level math courses, *J. App. Dev. Psychol.*, **29**(1), 17–28.

Good C., Aronson J. and Inzlicht M., (2003), Improving adolescents' standardized test performance: an intervention to reduce the effects of stereotype threat, *J. App. Dev. Psychol.*, **24**(6), 645–662.

Good C., Rattan A. and Dweck C. S., (2012), Why do women opt out? Sense of belonging and women's representation in mathematics, *J. Pers. Soc. Psychol.*, **102**(4), 700–717.

Gregorius R. M., (2017), Performance of underprepared students in traditional versus animation-based flipped-classroom settings, *Chem. Educ. Res. Pract.*, **18**(4), 841–848.

Higgins E. T. and Rholes W. S., (1978), "Saying is believing": effects of message modification on memory and liking for the person described, *J. Exp. Soc. Psychol.*, **14**(4), 363–378.

Higher Education Research Institute, (2010), *Degrees of success: Bachelor's degree completion rate among initial STEM majors*, Higher Education Research Institute at UCLA, Los Angeles.

Hockings S. C., DeAngelis K. J. and Frey R. F., (2008), Peer-led team learning in general chemistry: implementation and evaluation, *J. Chem. Educ.*, **85**(7), 990–996.

Jordt H., Eddy S. L., Brazil R., Lau I., Mann C., Brownell S. E., King K. and Freeman S., (2017), Values Affirmation Intervention

Reduces Achievement Gap between Underrepresented Minority and White Students in Introductory Biology Classes, CBE-Life *Sci. Educ.*, **16**(3), ar41.

Jury M., Smeding A. and Darnon C., (2015), First-generation students' underperformance at university: the impact of the function of selection, *Front. Psychol.*, **6**, 710.

Jury M., Smeding A., Stephens N. M., Nelson J. E., Aelenei C. and Darnon C., (2017), The Experience of low-SES students in higher education: psychological barriers to success and interventions to reduce social-class inequality, *J. Soc. Issues*, **73**(1), 23–41.

Kennepohl D., Guay M. and Thomas V., (2010), Using an online, self-diagnostic test for introductory general chemistry at an open university, *J. Chem. Educ.*, **87**(11), 1273–1277.

Kennett D. J. and Keefer K., (2006), Impact of learned resourcefulness and theories of intelligence on academic achievement of university students: an integrated approach, *Educ. Psychol.*, **26**(3), 441–457.

Krippendorff K., (2004), Reliability in content analysis: some common misconceptions and recommendations, *Hum. Comm. Res.*, **30**(3), 411–433.

Krippendorff K., (2013), Content Analysis: An Introduction to Its Methodology, 3rd edn, Thousand Oaks, CA: Sage.

Lewis S. E. and Lewis J. E., (2007), Predicting at-risk students in general chemistry: comparing formal thought to a general achievement measure, *Chem. Educ. Res. Pract.*, **8**(1), 32–51.

MacArthur J. R. and Jones L. L., (2008), A review of literature reports of clickers applicable to college chemistry classrooms, *Chem. Educ. Res. Pract.*, **9**(3), 187–195.

Merriam S. B., (2009), *Qualitative research: a guide to design and interpretation*, San Francisco: Jossey-Bass.

Mills P., Sweeney W. and Bonner S. M., (2009), Using the first exam for student placement in beginning chemistry courses, *J. Chem. Educ.*, **86**(6), 738–743.

Murnane R. J. and Willett J.B., (2010), *Methods matter: improving causal inference in educational and social science research*, Oxford: Oxford University Press.

Murphy M. C., Steele C. M. and Gross J. J., (2007), Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychol. Sci.*, **18**(10), 879–885.

National Science Foundation, National Center for Science and Engineering Statistics, (2017), Women, minorities, and persons with disabilities in science and engineering: 2017, Special Report NSF 17-310, Arlington, VA, available at www.nsf.gov/statistics/wmpd/.

Paunesku D., Walton G. M., Romero C., Smith E. N., Yeager D. S. and Dweck C. S., (2015), Mind-set interventions are a scalable treatment for academic underachievement, *Psychol. Sci.*, **26**(6), 784–793.

Pienta N. J., (2003), A placement examination and mathematics tutorial for general chemistry, *J. Chem. Educ.*, **80**(11), 1244–1246.

Potgieter M., Ackermann M. and Fletcher L., (2010), Inaccuracy of self-evaluation as additional variable for prediction of students at risk of failing first-year chemistry, *Chem. Educ. Res. Pract.*, **11**(1), 17–24.

Potter W. J. and Levine-Donnerstein D., (1999), Rethinking validity and reliability in content analysis, *J. Appl. Commun. Res.*, **27**(3), 258–284.

Prince M., (2004), Does active learning work? A review of the research, *J. Eng. Educ.* **93**(3), 223–231.

Rath K. A., Peterfreund A., Bayliss F., Runquist E. and Simonis U., (2012), Impact of supplemental instruction in entry-level chemistry courses at a midsized public university, *J. Chem. Educ.*, **89**(4), 449–455.

Rattan A., Good C. and Dweck C. S., (2012), ''It's ok—Not everyone can be good at math'': instructors with an entity theory comfort (and demotivate) students, *J. Exp. Soc. Psychol.*, **48**(3), 731–737.

Richardson J.T., (2011), Eta squared and partial eta squared as measures of effect size in educational research, *Educ. Res. Rev.*, **6**(2), 135–147.

Roediger H. L. and Butler A. C., (2011), The critical role of retrieval practice in long-term retention, *Trends. Cogn. Sci.*, **15**(1), 20–27.

Sadler P. M. and Tai R. H., (2007), Advanced placement exam scores as a predictor of performance in introductory college biology, chemistry and physics Courses, *Sci. Educ.*, **16**(2), 1–19.

Shields S. P., Hogrebe M. C., Spees W. M., Handlin L. B., Noelken G. P., Riley J. M. and Frey R. F., (2012), A Transition Program for underprepared students in general chemistry: diagnosis, implementation, and evaluation, *J. Chem. Educ.*, **89**(8), 995–1000.

Sisk V. F., Burgoyne A. P., Sun J., Butler J. L. and Macnamara B. N., (2018), To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses, *Psychol. Sci.*, **29**(4), 549–571.

Smith J. L., Lewis K. L., Hawthorne L. and Hodges S. D., (2013), When trying hard isn't natural women's belonging with and motivation for male-dominated stem fields as a function of effort expenditure concerns, *Pers. Soc. Psychol. Bull.*, **39**(2), 131–143.

Spencer S. J., Steele C. M. and Quinn D. M., (1999), Stereotype threat and women's math performance, *J. Exp. Soc. Psychol.*, **35**(1), 4–28.

Steele C. M., (1997), A threat in the air: How stereotypes shape intellectual identity and performance, *Am. Psychol.*, **52**(6), 613–629.

Steele C. M. and Aronson J., (1995), Stereotype threat and the intellectual test performance of African Americans, *J. Pers. Soc. Psychol.*, **69**(5), 797–811.

Stoet G. and Geary D. C., (2018), The gender-equality paradox in science, technology, engineering, and mathematics education, *Psychol. Sci.*, **29**(4), 581–593.

Tai R. H., Sadler P. M. and Loehr J. F., (2005), Factors influencing success in introductory college chemistry, *J. Res. Sci. Teach.*, **42**(9), 987–1012.

The College Board, (2018a), Discover AP, visited February 2018, https://ap.collegeboard.org.

The College Board (2018b), SAT Suite of Assessments, visited February 2018, https://collegereadiness.collegeboard.org/sat.

Venezia A. and Kirst M. W., (2005), Inequitable opportunities: How current education systems and policies undermine the chances for student persistence and success in college, *Educ. Policy*, **19**(2), 283–307.

Wagner E. P., Sasser H. and DiBiase W. J., (2002), Predicting students at risk in general chemistry using pre-semester assessments and demographic information, *J. Chem. Educ.*, **79**(6), 749–755.

Walton G. M., (2014), The new science of wise psychological interventions, *Curr. Dir. Psychol. Sci.*, **23**(1), 73–82.

Walton G. M., Logel C., Peach J. M., Spencer S. J. and Zanna M. P., (2015), Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering, *J. Educ. Psychol.*, **107**(2), 468–485.

Wicherts J. M., (2005), Stereotype threat research and the assumptions underlying analysis of covariance, *Am. Psychol.*, **60**(3), 267–269.

Wilson T. D. and Linville P. W., (1982), Improving the academic performance of college freshmen: attribution therapy revisited, *J. Pers. Soc. Psychol.*, **42**(2), 367–376.

Wilson T. D. and Linville P. W., (1985), Improving the performance of college freshmen with attributional techniques, *J. Pers. Soc. Psychol.*, **49**(1), 287–293.

Xu X., Villafane S. M. and Lewis J. E., (2013), College students' attitudes toward chemistry, conceptual knowledge and achievement: structural equation model analysis. *Chem. Educ. Res. Pract.*, **14**(2), 188–200.

Yeager D. S. and Dweck C. S., (2012), Mindsets that promote resilience: when students believe that personal characteristics can be developed, *Educ. Psychol.*, **47**(4), 302–314.

Yeager D. S., Romero C., Paunesku D., Hulleman C. S., Schneider B., Hinojosa C., Lee H. Y., O'Brien J., Flint K., Roberts A., Trott J., Greene D., Walton G. M. and Dweck C.S., (2016a), Using design thinking to improve psychological interventions: the case of the growth mindset during the transition to high school, *J. Educ. Psychol.*, **108**(3), 374–391.

Yeager D. S. and Walton G. M., (2011), Social-psychological interventions in education: they're not magic. *Rev. Educ. Res.*, **81**(2), 267–301.

Yeager D. S., Walton G. M., Brady S. T., Akcinar E. N., Paunesku D., Keane L., Kamentz D., Ritter G., Duckworth A. L., Urstein R., Gomez E. M., Markus H. R., Cohen G. L. and Dweck C. S., (2016b), Teaching a lay theory before college narrows achievement gaps at scale. *Proc. Natl. Acad. Sci. U. S. A.*, **113**(24), E3341–E3348.

