JOURNAL OF CHEMICAL EDUCATION

pubs.acs.org/jchemeduc Article

Short-Term and Long-Term Effects of POGIL in a Large-Enrollment General Chemistry Course

Paulette Vincent-Ruz,* Tara Meyer, Sean G. Roe, and Christian D. Schunn

Cite This: https://dx.doi.org/10.1021/acs.jchemed.9b01052 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Process oriented guided inquiry learning (POGIL) is a specific type of active learning centered on a learning cycle where students first explore a concept through scientific models, followed by a concept invention, and finally a concept application phase. In spite of POGIL's research-based design and the many studies showing it increases learning outcomes, there is still a critical gap in the knowledge behind the effect of POGIL in attitudinal factors and the mechanisms behind it. The current study seeks to build an understanding of the mechanistic ways in which POGIL works and its effect on students' attitudes and learning. The sample consisted of students who enrolled in General Chemistry I in the Fall semester across 7 sections (classrooms) at an R1 (large research) university in the eastern part of the US. Four sections used POGIL instruction (N = 809) while the other three used Traditional teaching methods (N = 543). Statistical models using multilevel statistics show students in the POGIL condition had higher chemistry identity, competency beliefs, and chemistry grades. Furthermore, performance in General Chemistry I appeared to be a core mediator of all the observed differences in General Chemistry II where students in the POGIL condition still performed better and had higher chemistry-related attitudes.

KEYWORDS: *Upper-Division Undergraduate, Chemical Education Research, Inquiry-Based/Discovery Learning*

FEATURE: Chemical Education Research

INTRODUCTION

Decades of science education scholarship have called for greater use of student-centered approaches, particularly active learning that requires students to act like scientists.[1,2] However, such active learning is hard to implement in introductory college science classes because of the large amount of content that must be covered juxtaposed against the time required for doing meaningful scientific inquiry, especially during lectures.[3,4] Further, students often perceive that they learn less from active learning than from passive learning, creating resistance to change.[5] In addition, while there is much research supporting the use of active learning in general, some ways of implementing active learning are less effective,[4] requiring research on more specific forms of active learning. Finally, implementation of pedagogies that rely on teamwork, a common element of active learning, may be difficult to implement,[6] particularly in introductory classes where students may be less skilled in teamwork.[3]

Responding to this challenge, many chemistry instructors have begun to adopt POGIL (process oriented guided inquiry learning) because it provides structured guidance that allows student to engage in inquiry learning in an efficient way.[3] As reviewed in the next section, many studies have found positive effects with POGIL. However, though other types of active learning have shown positive long-term effects,[7] POGIL in particular has little research examining this type of impact (e.g., performance in later courses or sustained attitudes toward chemistry). This is especially important because students will often transition back into later chemistry courses that are using Traditional lecture-based teaching methods.

Why might POGIL have longer-term effects? Active learning interventions have regularly shown long-term effects.[4] For POGIL in particular, some researchers have argued that students are learning the chemistry content in a deeper way.[8] We argue that, by engaging in a POGIL semester-long experience, students will also go through a positive shift in their chemistry-related attitudes. Chemistry-related attitudes are connected to performance.[9] Therefore, we would argue that students experiencing POGIL will have higher performance in both the POGIL course itself and in later courses. In addition, it is also possible that attitude effects will primarily influence decisions to persist in chemistry and that later performance effects are caused by deeper mastery of chemistry content. Finally, Lo and Mendez's review of learning effects in POGIL implementations revealed a lack of large-scale quantitative studies showing more macrolevel evidence of POGIL effectiveness.[10] Research is therefore needed to understand the mechanisms by which POGIL improves learning in order to optimize its implementation or prevent lethal mutations. These considerations led to our research questions.

Received: December 1, 2019
Revised: March 30, 2020

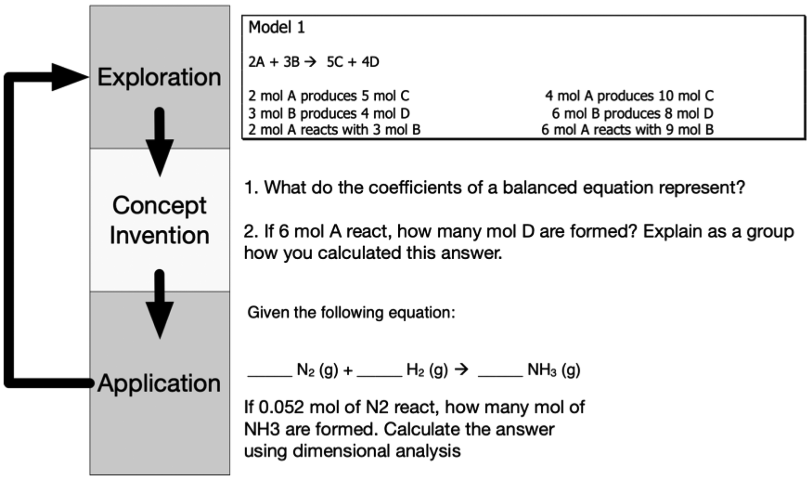

Figure 1. Example of a POGIL activity used in chemistry class and its mapping to each of the steps of the learning cycle.

- RQ1: What are the effects of POGIL on immediate outcomes (attitudinal, knowledge, and retention) within large lecture classrooms?
- RQ2: What are the effects on later coursework?
- RQ3: Are the effects on later coursework a result of increased chemistry knowledge or higher attitudes?

In the next section, we give a short overview of POGIL and prior research on POGIL's effects on student learning, and then, we present a new framework for how POGIL may be able to support the development of positive chemistry-related attitudes since there is relatively little research on that topic.

■ POGIL THEORY AND PRIOR RESEARCH

POGIL is an active learning tool that has two conceptual elements: process oriented and guided inquiry. Process oriented refers to explicit attention to student-centered and teamwork aspects of the approach that prioritize critical thinking and building ideas across different students.[3,4,6,11−13] For example, students are asked to collaborate in groups, which forces them to increase their oral communication skills as well as build on each other's ideas. Guided inquiry refers to the learning cycle in which students are asked in every POGIL activity to infer chemistry concepts from data in a highly structured approach.

More specifically, the POGIL learning cycle always has three steps (Figure 1): (1) an exploration phase where students are given a model (e.g., graph, data table, diagram) and given scaffolding questions to extract information from the model, (2) a concept invention phase where students must synthesize information from the previous section and make a generalization,[14,15] and (3) an application phase in which students apply the concept to exercises that are close-ended questions.[16,17] The combination of the learning cycle and the teamwork leads to students learning each key chemistry concept.

Many studies have examined the effects of the POGIL approach on students passing the class in which the approach was used, including general chemistry, organic chemistry, biochemistry, analytical chemistry, and computational thinking classes. A recent meta-analysis of 21 studies found significant effects of POGIL on reducing failure rates, on average reducing the risk of failing the class by 38%.[18] The typical class size in these studies was small (less than 50 students) or medium (51−100 students), with far fewer studies of POGIL being used in large classrooms.[19] Furthermore, broader research studies on active learning, and also research specifically on POGIL, have found that student success is dependent on factors that are challenging in large lectures: instructor facilitation strategies,[20] the way students engage in scientific discourse,[21,22] and prior knowledge or skills.[23] Therefore, the paucity of studies in large lecture courses is unfortunately because the implementational logistics may be too complicated to do so with fidelity and be effective in such situations.[6,24]

In addition, little work has examined the long-term effects of POGIL (e.g., retention to the next course or performance in later courses). Further, despite a frequently hypothesized connection of active learning to attitudinal development along with content learning,[1] only a small number of POGIL studies have focused on motivational outcomes such as chemistry competency beliefs,[20] and no specific mechanistic explanation has been provided for why POGIL should affect attitudes nor which attitudinal variables in particular should be affected (e.g., competency beliefs, sense of belonging, interest, and identity). On the basis of a review of the motivation literature, we propose the following framework for understanding how each step of the learning cycle may be supporting the development of different kinds of chemistry-related attitudes, which then motivates the inclusion of specific attitudinal measures in the current study.

- Exploration: A key aspect of the exploration phase is teamwork. In this stage, students brainstorm and compare their reasoning behind their model interpretation. Productive interactions with others can build a sense of belonging in the classroom, which can have an effect on a student's feeling of belonging in chemistry more generally, which ultimately should affect their chemistry identity.[25] Furthermore, students are exposed to a realistic aspect of the scientific endeavor; observing the class culture developed around exploration and productive struggle can help students realize they are not the only ones that find the concept difficult (i.e., influence their chemistry competency beliefs).
- Concept Invention: Open-ended problems can increase interest but can also lead students to experience a struggle, which can be anxiety inducing and have negative effects on students' motivation.[22] Such negative effects can be buffered with adequate instructional support like the scaffolding provided in POGIL activities, which can ultimately have the effect of increasing their chemistry competency beliefs.[19,23,24]

https://dx.doi.org/10.1021/acs.jchemed.9b01052
J. Chem. Educ. XXXX, XXX, XXX−XXX

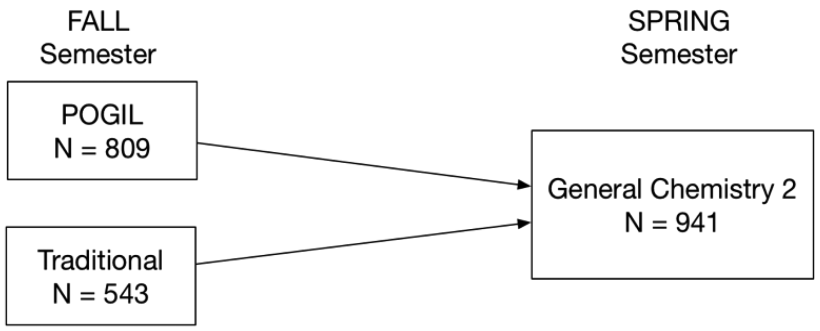

Figure 2. Student distribution across conditions in General Chemistry I, as well as number of students retained to General Chemistry II

- Application: The application phase provides students with the opportunity to transfer theoretical concepts through practice with close-ended problems.[17] It is in this step where concept understanding is cemented from a general and intuitive idea to the proper name, definition, and use of the concept in chemistry. Practice with closed-form problems can further lead to increases in competency beliefs because of the lower struggle with such problems.

■ METHODS

Throughout the rest of the paper, we will capitalize names when referring to specific measures/variables in methods and analyses rather than constructs, similar to distinguishing a specific course that was studied versus the general course on that topic.

Student Participants

The sample consists of students who enrolled in General Chemistry I in the Fall semester across 7 sections (classrooms) at a primarily research university in the eastern part of the US. Four sections used POGIL instruction ($N = 981$), and classroom size ranged from 218 students to 258 students. The other four sections used Traditional teaching methods (called the Traditional sections; $N = 620$), and classroom size ranged from 95 to 258 students (only one section had <200 students). Each section was taught by one instructor; both approaches involved a mixture of more and less experienced instructors, and both involved tenure- and teaching-stream instructors. There was at least one POGIL and one Traditional section taught at roughly the same time throughout the day. Students did not know to which type of classroom they had enrolled until the first day of classes. The sample for analysis was restricted to freshmen (84%) giving a total of 1,352 students (Figure 2). Of those students, only 941 enrolled in General Chemistry II in the following semester (69% overall retention rate). Table 1 shows Demographics by group at the start of General Chemistry I and General Chemistry II. There were no important demographic differences between conditions.

Instructors

To ensure that the teachers were roughly comparable in teaching quality, students' grades in General Chemistry I (all traditionally taught) in the prior year were examined within the teachers who had taught this course the prior year. In fact, students in the classes taught by the teachers who would implement POGIL the next year had slightly lower grades than those who would implement Traditional instruction again the next year ($p < 0.0001$, $d = 0.27$); i.e., there was a slight bias in favor of the Traditional group in terms of prior teaching outcomes.

Table 1. Demographic Distributions within Each Course as a Function of Instructional Condition in the General Chemistry I

	Students, %, By Course			
	General Chemistry I (N = 1,352)		General Chemistry II (N = 941)	
Variable[a]	POGIL	Traditional	POGIL	Traditional
Woman	67	64	68	64
First generation	9	8	8	8
Pacific Islander	0.2	1	0.3	2
Native American	0.8	0.3	0.7	0.3
Latinx	5	7	4	6
Black	8	7	7	5
Asian	26	26	30	32
White	68	69	64	66

[a]Race/ethnicity percentage add up to more than 100% due to students identifying as more than one race/ethnicity. 0.8% of respondents identified as transgender or genderqueer in the survey. Although nonbinary gender identity is an important element of attrition in STEM, given the statistical nature of the analysis in this study, trans and nonbinary students could not be meaningfully included.

Intervention Implementation

An intervention is defined as "a specified set of activities designed to put into practice an activity of known dimensions".[26] The extent to which delivery of an intervention adheres to the model originally developed is called the fidelity of implementation (FIO). However, DBER interventions in the published literature often fail to describe these interventions with enough detail so that others can judge both the results and reproduce them in different environments. This is especially important when multiple instructors are involved.[27] The US National Science Foundation has called for the measurement of FOI when conducting impact studies and analyzing relationships between variations in FOI and intervention outcomes.[28] Stains and Vickrey[27] developed a set of guidelines to measure FOI within DBER settings. Using those guidelines, we outline our implementation of the POGIL intervention.

- POGIL implementation in the classroom: The intervention used hybrid POGIL, defined as a combination of the POGIL method (~1/2 of weekly class time) and a more conventional lecture format (sometimes by having half lecture and half POGIL the same day, and sometimes full POGIL sessions), following the model by Perry and Wright.[29] Teams were created simply on the basis of where students chose to sit on the first day of class. The POGIL routines established in class were adapted from

https://dx.doi.org/10.1021/acs.jchemed.9b01052
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 2. Means, Standard Deviations, and Pearson Correlations for Outcome Measures at Time 1, Beginning of General Chemistry II, and Time 2, End of General Chemistry II

Outcome Measure	Mean (SD), N = 941[a]	Correlations[b]						
		GenChem1	GenChem2	I1	I2	CB1	CB2	F1
GenChem1 grade	2.6 (1.0)							
GenChem2 grade	2.8 (0.9)	0.72[b]						
Identity1 (I1)	2.6 (0.7)	0.32	0.19					
Identity2 (I2)	2.7 (0.7)	0.50	0.48	0.81[c]				
Competency Beliefs1 (CB1)	2.8 (0.5)	0.39	0.31	0.64	0.55			
Competency Beliefs2 (CB2)	2.7 (0.5)	0.49	0.46	0.49	0.64	0.66[c]		
Fascination1 (F1)	2.8 (0.6)	0.18	0.17	0.60	0.47	0.44	0.42	
Fascination2 (F2)	2.8 (0.6)	0.22	0.21	0.58	0.58	0.41	0.50	0.69[c]

[a]Sample size reported for the pairwise-correlations between Time 1 and Time 2. [b]All correlations are statistically significant at the $p < 0.001$ level. [c]Correlation measures construct stability/malleability.

the work of Vishnumolakala and collaborators to fit our institutional constraints.[30] The class started with a quick introduction to the topic and learning objectives. Students then started the POGIL activity and stopped at the end of the "Exploration" section to do a quick clicker quiz, where the instructor provided feedback. Then, students continued the activity and stopped at the end of the "Concept Invention" section to do a quick clicker quiz, where the instructor again provided feedback. The class would then move to the "Concept Application" section for the rest of class.

- Class facilitation: POGIL pedagogy is highly dependent on the facilitator's role in the classroom.[3,21] In a large-enrollment class, it is not possible for a single teacher to monitor approximately 80 teams in a single class (N = 264). A group of undergraduate and graduate TAs were trained specifically to be in-class facilitators of POGIL activities. This training consisted of a one-hour-long session on POGIL at the beginning of the semester that was led by a certified POGIL instructor that has been training POGIL teachers for around 10 years. There were also weekly meetings with instructors to talk about the next POGIL activity and issues that were raised the prior week. TAs were required to solve the POGIL activity for that week before each weekly training sessions. There were 8 TA facilitators present in a given classroom (1 graduate student TA and 7 undergraduate student TAs), and each TA facilitator was in charge of approximately 11 teams (~35 students).
- Curriculum design and implementation: The syllabus, clicker questions, exams, and homework were developed by the four chemistry instructors and the research team. The 19 POGIL activities were selected using the Wiley Custom Select POGIL activity collection, and all students in the POGIL sections used the same selected POGIL activities in class.

Data Collection Procedures

The university provided researchers with enrollment data (for tracking retention) and grades (for tracking performance effects) in both General Chemistry I and II. The university also provided other institutional data on the students to help establish equivalence among conditions: race/ethnicity, first-generation status, Advanced Placement (AP) Chemistry scores, Math and Verbal SAT scores, high school GPA, and grades for all university courses.

All students who continued to General Chemistry II enrolled in traditionally taught sections that each combined some students who had previously experienced POGIL and some students who had previously experienced Traditional instruction, thereby removing any confounds between prior instructional format and current instructor. Within General Chemistry II, students were surveyed regarding their attitudes toward chemistry at the beginning (referred to as Time 1) and end of the semester (referred to as Time 2) to assess immediate and long-lasting difference across conditions.

Academic Controls

To rule out prior differences in who enrolled in the POGIL and traditionally taught sections as an alternative explanation, we focused on dimensions previously found to be especially relevant to general chemistry: standardized math scores,[9] high school grades,[31] and Advanced Placement Chemistry experience and performance.[32] Studies have shown the predictive power of these variables on students' college success, specifically in the early undergraduate years.[32−35] However, it is important to note that such grades and scores also reflect systemic inequities against marginalized students, such as socioeconomic status and opportunity access.[36−39] Inequities can be found in "shadow education" (educational activities that occur outside of formal schooling)[40−42] or school resources like AP course offerings.[37] Therefore, when reporting and interpreting coefficients from these variables, it is important to highlight the underlying complexities in the factors these variables actually capture.

Standardized Math Score

In the United States, many universities require students to provide scores from either SAT or ACT exams; of the component scores within each exam, the mathematics component is most relevant when it comes to STEM courses. We used correspondence tables provided by the testing agencies in order to convert ACT math scores into the SAT math scale[40] and henceforth refer to it as the standardized math score (M = 680, SD = 66, max = 800, national average = 530).

High School GPA

In the United States, high school course grades are reported on a 0−4 scale, but the grades from any Advanced Placement courses (designed to be equivalent to introductory university courses and involve a standardized end-of-course exam) are reported on a 1−5 scale, effectively making GPA a 0 to 5 scale. Overall high school GPA is used as a continuous variable (M = 4.0, SD = 0.4).

https://dx.doi.org/10.1021/acs.jchemed.9b01052
J. Chem. Educ. XXXX, XXX, XXX−XXX

Advanced Placement in Chemistry (AP Chemistry)

AP Chemistry covers general chemistry topics like atomic theory, chemical bonding, phases of matter, solutions, types of reactions, chemical equilibrium, reaction kinetics, electrochemistry, and thermodynamics. 25% of students enrolled in General Chemistry I had AP Chemistry exam scores ($M = 3.0$, SD = 1.0). Although students with high exam scores are eligible to enroll directly in General Chemistry II, some students with such scores choose to enroll in General Chemistry I anyway; they are most commonly students who seek high average science course grades for their applications to medical school. On the basis of the distribution of scores, we created a three-level categorical variable names for the AP Chemistry Condition: (1) "High Score" for students who scored 4−5 on the exam (8%); (2) "Low Score" for students who scored 1−3 on the exam (17%); and (3) "Did not Present" for those who did not have an exam score in our data set (75%), meaning they either did not take the course, took the course but did not take the exam, or took the exam but chose to not share the score with the university because it was a low score.

Outcome Measures

In this study, we use both academic and attitudinal measures as outcome measures. Grades are not necessarily indicators of deep learning.[41,42] However, students internalize grades as a measure of their abilities[43,44] and make course and career decisions based on them.[45,46] Therefore, when evaluating interventions, it is important to pay attention to changes in grades and science-related attitudes.[47−49] Means, standard deviations, and intercorrelations between outcome measures are provided in Table 2. All outcome measures were positively correlated with each other with at least a medium effect size, thereby requiring regression methods to tease apart unique effects.

Academic Outcomes

In terms of academic outcomes, we focused on grades in General Chemistry I and General Chemistry II, and retention from General Chemistry I to General Chemistry II. University grades were reported on a scale from 0 to 4. Another outcome measure that is often important at the department level is the percentage of students in the ends of the grade distribution (meaning the highest grades). On the high end, we computed the percentage of students with A+, A, or A− letter grades (scores from 3.75 to 4.0). On the low end, because most of the students in the class are coming from majors that required a score at least a C to pass a class, we calculated the percentage of students earning a C−, a D, a fail, or a late course withdrawal (called the C-DFW rate). For General Chemistry I, 22% of students scored in the high range of grades, and 17% had a C-DFW. For General Chemistry II, 29% of students had high scores, and 9% had a C-DFW.

Attitudinal Surveys

Students completed an online attitudinal survey (for extra credit in the class) during the first and last few weeks of the semester during General Chemistry II to produce measures for Chemistry Fascination (interest in and mastery goals for chemistry content and skills, 3 items, $\alpha = 0.78$[50,51]), Chemistry Competency Beliefs (beliefs about being having chemistry-related skills and being able to successfully complete chemistry-related tasks, 5 items, $\alpha = 0.84$[9,52]), and Chemistry Identity (self-perception and other perceptions of student as a science type of person, 4 items, $\alpha = 0.93$[53,54]).

The attitude measures were developed by adapting previously validated scales: Colorado Learning Science Survey for Use in Chemistry (CLASS-Chem),[55,56] the Chemistry Self-Concept Inventory (CSCI),[57,58] and a Science Identity survey.[53,54] These surveys have been independently validated in several prior studies in terms of stability, convergent validity, and malleability.[9,50−53,56] After selecting items for our instruments, we conducted pilot studies. The first step was to conduct cognitive interviews.[59] These interviews consisted of think-alouds in which approximately five male and female students from diverse backgrounds read each item, explained in their own words what the item was asking, and explained the reasoning behind their elected response to the item. This process ensured that the items reflected what they were intended to measure in the student population to be studied. Additional psychometric analyses (Exploratory Factor Analyses and Item Response Theory analyses) were then conducted on the basis of pilot data in multiple levels of university coursework to remove poorly behaving survey items and ensure overall acceptable scale properties.

Finally, we also tested for one important measure of reliability for psychometric instruments, that of temporal stability. For an instrument to be useful, it is important for it to have a reasonable level of temporal stability. Given the test−retest interval (3−4 months), one would expect a test−retest coefficient of above $r = 0.60$ to be considered good and above $r = 0.80$ to be considered excellent.[60] Attitudinal constructs are considered to be semi-malleable; that is, under intervention we only expect changes in scale scores to be present for interventions that are at least several weeks or months in duration, not single hour-long experiences. The Chemistry Fascination test−retest coefficient was $r = 0.79$, meaning it is very stable but there was enough change from Time 1 to Time 2 to be considered semimalleable. Test−retest for Chemistry Competency Beliefs was $r = 0.66$, and for Chemistry Identity it was $r = 0.81$. These three attitudinal constructors present sufficient stability to be considered reliable and evidence of semimalleability.

■ ANALYSES

We began with simple statistical tests of differences by condition and then constructed more advanced statistical models that more accurately modeled the effects of instructors and controlled for potentially confounding variables. The specific analysis details varied by the nature of the dependent variable: continuous (like grades and attitudes) or categorical (like retention or particular grade end points).

Differences Across Conditions

Continuous Variables. For continuous variables, we conducted *t*-tests as a way of testing difference across conditions. However, when testing for multiple outcomes at the same time point, it is important to avoid the accumulation of Type-I errors when conducting multiple comparisons.[31] *p*-Values of comparisons were adjusted using Hochberg's method, which involves conducting statistical inference of hypothesis by starting with the largest *p*-value.[32] Cohen's *d* was used as a measure of effect size. With large sample sizes, a difference can be statistically robust but practically meaningless.[33] Effect sizes provide a magnitude of the difference between two groups: by convention, Cohen's *d*-values of 0.2 are called small effect sizes, 0.5 medium, and 0.8 or more large.[33] The effectiveness of a particular intervention can only be interpreted in relation to other interventions that seek to produce the same effect.[34]

Binary and Categorical Variables. For categorical and binary variables, we conducted a chi-square test of independence

https://dx.doi.org/10.1021/acs.jchemed.9b01052
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 3. Means and Standard Deviations by Condition, along with *p*-Values and Effect Sizes for the Condition Contrast for Academic Controls at the Beginning of General Chemistry II

Variable	Type	Mean (SD) N = 941 POGIL	Traditional	p-Values[a]	Effect Size: Cohen's d Values
Standardized Math[b]	Continuous	684 (67)	676 (64)	0.14	0.12
HS GPA[c]	Continuous	4.1 (0.4)	4.0 (0.3)	0.15	0.11
AP Chem Score[d]	Continuous	3.0 (1.0)	2.9 (1.0)	0.77	0.08
AP Chem Group[e]	Categorical				
High Score,[f] %		18	15	0.33	1.27
Low Score,[g] %		8	8	0.77	1.10

[a]The p-values are corrected for multiple comparisons using the Hochberg method. [b]The standardized math score represents SAT and converted ACT math scores: M = 680, SD = 66, max = 800, national average = 530. [c]Grade point average may include at least one AP course; thus, it is reported on a 0−5 scale. M = 4.0, SD = 0.4. [d]AP Chemistry exam scores have a scale of 1−5, with 5 a high score. [e]Odds ratio relative to "Didn't Present" AP Chemistry Exam (students who do not have an exam score in the data set). [f]Students who scored 4−5 on the AP Chemistry Exam. [g]Students who scored 1−3 on the AP Chemistry Exam.

(i.e., equal probability distributions by condition). Odds ratios (ORs) were used as a measure of effect size. In this case, it represents the odds that an outcome will occur in the POGIL condition compared with the odds of it occurring in the control condition. An OR = 1 means that treatment does not affect the odds of the outcome; OR > 1 means that treatment is associated with higher odds of outcome, and OR < 1 means treatment is associated with lower odds of outcome. It is also important to report confidence intervals whenever reporting an odds ratio to provide a reader with some information about the precision of the result.[35] As a guideline, 1.5 < OR is considered a small effect size and OR > 5 a large effect size.[36]

Multilevel Models

For the larger and statistically significant differences in outcomes, we then constructed multilevel statistical models. Multilevel models are used to analyze variance on an outcome variables while accounting for influence of various commonalities students might share by having the same teacher and being in the same classroom.[37] Since each instructor had only one section, we used a two-level nested model (students nested within classrooms). All mixed models were fit using the lme4 package[39] in the R statistical computing language.[40]

Continuous Variables. For continuous variables, we ran linear mixed models (multivariate regression accounting for teacher differences). For each outcome, we ran two models: (I) simple contrast, where intervention condition was the only independent variable; and (II) with entry controls (intervention condition, demographic variables, AP scores, and standardized math scores) as independent variables.

Binary and Categorical Variables. For binary and categorical variables, we ran generalized mixed models (multinomial logistic regression accounting for nesting). For each outcome we ran two models: (I) simple contrast, where intervention condition was the only independent variable; and (II) with entry controls (intervention condition, demographic variables, AP scores, and standardized math scores) as independent variables.

Structural Equation Models

Mediation Models. Mediation is the process by which one variable transmits an effect onto another through one or more mediating variables. A mediator variable (Me) is the variable that captures a causal step between the dependent (Y) and the independent (X) variables. For example, students from the POGIL condition might perform better in General Chemistry II because of their greater mastery of General Chemistry I content as measured through higher performance in General Chemistry I. Mediation is tested by measuring the indirect effect, the decrease in the relationship between condition and delayed outcomes after partialing out the association between condition and General Chemistry I grade. Mediation models were tested using Structural Equation Modeling with the lavaan package and the bootstrapping method using the DWLS estimator (diagonally weighted least-squares). Bootstrap estimates are less biased for sample sizes of $N \geq 200$, for both normal and non-normal distributions.

■ RESULTS

Before an examination of the research questions, it was important to first assess whether there were any important differences between populations before the intervention. As shown in Table 3, there were no important differences between students in the POGIL and Traditional sections in terms of incoming academic characteristics. However, to be conservative since there were some small differences and because including important predictors improves statistical power/precision, we control for these variables in later analyses.

What Are the Effects of POGIL on Critical Immediate Outcomes?

The left-most columns of Table 4 show the outcomes for students as a function of having experienced General Chemistry I in POGIL or Traditional Lecture formats on all the key measures (e.g., simple raw means and standard deviations; raw percentages). The next columns to the right show effect sizes and statistical-significance levels of the contrast between conditions using different analytic models. Overall, students were significantly different after the POGIL experience on many outcome variables, both at the end of General Chemistry I and at the end of General Chemistry II.

There were medium-sized differences in General Chemistry I grades, which result from POGIL group students both being more likely to score in the higher-grade range and being less likely to receive a nonsatisfactory grade. POGIL group students were also more likely to enroll in General Chemistry II the next semester.

In terms of attitudes, the effects varied substantially by attitude: (1) the POGIL group was no different on Chemistry Fascination; (2) there was a small difference in Chemistry Competency Beliefs; (3) there was a large difference in Chemistry Identity. In fact, the POGIL group students were 2

https://dx.doi.org/10.1021/acs.jchemed.9b01052
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 4. Means, Standard Deviations, *p*-Values and Effect Sizes for General Chemistry Outcomes

Outcome	Mean (SD) N = 1,352		Effect Size for POGIL vs Traditional[a]	
	POGIL	Traditional	Simple Contrast[b]	With Entry Controls[b]
Gen Chem I Grades[c]				
Mean grade	2.8 (1.0)	2.4 (1.0)	0.32[d]	0.11[e]
High grade, %	26	16	1.6:1[d]	1.5:1[e]
Not Satisfactory grade, %	14	20	0.7:1[e]	0.6:1[e]
Retention to Gen Chem II, %	73	64	1.6:1[d]	1.5:1[e]
Attitudes toward Chemistry at Beginning of Gen Chem II, N = 941[c]				
Fascination	2.9 (0.6)	2.8 (0.5)	0.04[g]	[h]
Competency Beliefs	2.9 (0.5)	2.7 (0.6)	0.24[e]	0.11[e]
Identity	2.8 (0.6)	2.4 (0.7)	0.52[d]	0.23[d]

[a]Effect sizes of the condition contrast without and with academic and demographic controls. [b]The *p*-values are corrected for multiple comparisons using the Hochberg method. [c]Std Beta is reported for continuous variables and Odd Ratios for binomial and multinomial outcomes. N reported for students enrolled in General Chemistry II. [d]$p < 0.001$. [e]$p < 0.01$. [f]$p < 0.05$. [g]Not statistically significant. [h]Follow-up analyses omitted for nonstatistically significant simple contrast effects.

times more likely to strongly consider themselves a chemistry type of person (23% in POGIL vs 10% in Traditional).

What Are the Effects of POGIL on Critical Distal Outcomes?

Moving to distal outcomes, Figure 3 shows the percentage of students with high chemistry-related attitudes and grades in each condition at both the beginning and end of General Chemistry II. In this section we only describe the differences, while on the next Research Question we explore possible mechanisms. At the end of General Chemistry II, there continued to be differences in the overall grade distribution between both conditions, including a marginally significant effect of POGIL students being less likely to fail General Chemistry II (see Table 5). Finally, students in the POGIL condition still report higher

Table 5. Means, Standard Deviations, *p*-Values, and Effect Sizes for Condition Differences in General Chemistry II Outcomes

Outcome	Mean (SD) N = 941		Effect Size for POGIL vs Traditional[a]	
	POGIL	Traditional	Simple Contrast[b]	Simple Contrast[b]
Gen Chem II Grades[c]				
Mean grade	2.9 (0.8)	2.7 (0.9)	0.18[d]	0.12[t]
High grade, %	30	26	1.2:1	
Not Satisfactory grade, %	7	11	0.8:1[f]	0.6:1[g]
Attitudes toward Chemistry at End of Gen Chem II[c]				
Fascination	2.8 (0.6)	2.9 (0.6)	0.05[h]	[i]
Competency Beliefs	2.8 (0.5)	2.7 (0.5)	0.19[e]	0.18[e]
Identity	2.8 (0.7)	2.6 (0.7)	0.27[e]	0.21[f]

[a]Effect sizes of the condition contrast without and with academic and demographic controls. [b]The *p*-values are corrected for multiple comparisons using the Hochberg method. [c]Std Beta is reported for continuous variables and Odd Ratios for binomial and multinomial outcomes. [d]$p < 0.001$. [e]$p < 0.01$. [f]$p < 0.05$. [g]$p < 0.1$. [h]Not statistically significant. [i]Follow-up analyses omitted for nonstatistically significant simple contrast effects.

levels of Chemistry Competency Beliefs and Chemistry Identity, though the size effects were both small at this time point.

For all outcomes showing significant simple differences, additional models were run that control for incoming academic factors and demographic factors to ensure the observed differences were not a product of incoming population differences (see Tables 4 and 5, "With Entry Controls"). All the differences remained both statistically significant and similarly sized, except for Chemistry Identity at the end of General Chemistry II. Given that more students were retained to

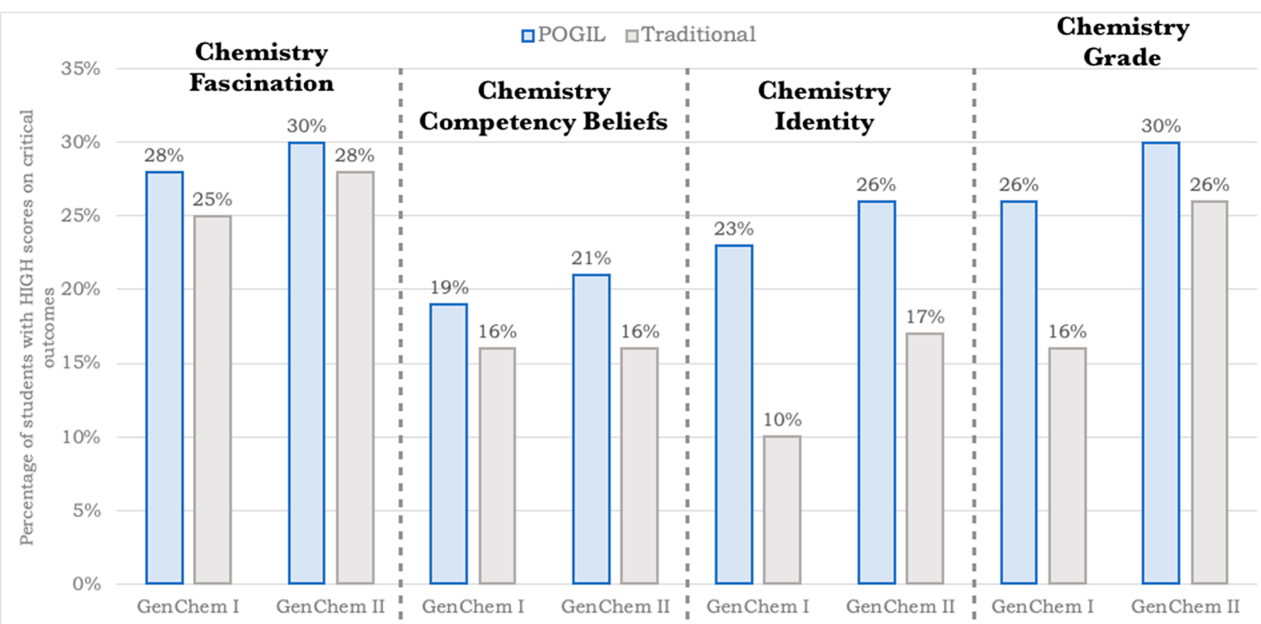

Figure 3. Percentage of students in each condition with high chemistry-related attitudes and grades after General Chemistry I and after General Chemistry II.

https://dx.doi.org/10.1021/acs.jchemed.9b01052
J. Chem. Educ. XXXX, XXX, XXX−XXX

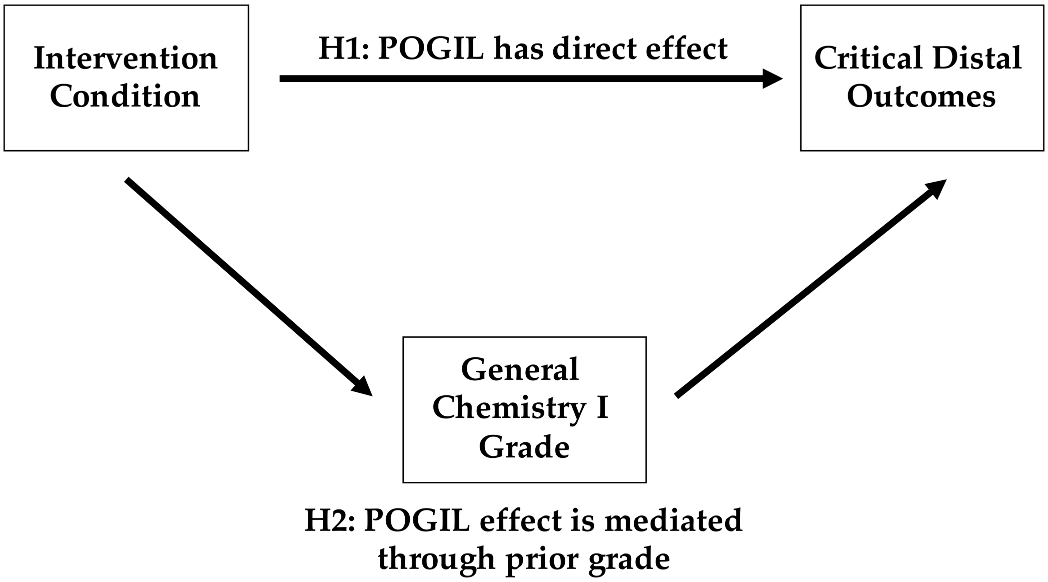

Figure 4. Proposed mechanisms behind the differences in the distal outcomes.

General Chemistry II in the POGIL section, some adjustments of effects sizes were possible with entry controls. Importantly, the initial attitude effects, the retention effect, and the grade effects were found to be robust across the simple contrast and with the addition of entry controls.

The reduction in size of the Chemistry Identity effect by the end of General Chemistry II was surprising, especially given its large initial size and the robustness of the other differences. In exploring the variability in outcomes by section (by General Chemistry I section and by General Chemistry II section), the sectional variation by General Chemistry II stood out. In particular, there was substantial variation in whether Chemistry Identity significantly increased (one section), stayed approximately the same (two sections), or significantly decreased (two sections). In other words, the Chemistry Identity effects seemed to be partially overwhelmed by the new instructional context.

Do Immediate Effects of POGIL Explain the Distal Outcome

Finally, we explored the possible mechanisms behind the differences in the distal outcomes. It is possible the differences observed in General Chemistry II (retention, attitudes, and grades) are a result of the improved knowledge and skills in General Chemistry I (as might be indicated by higher General Chemistry I grades). So, we compared models and tested two alternative hypotheses: H1-POGIL had a direct effect on the differences observed in General Chemistry II, or H2-POGIL reflected that the General Chemistry II differences are mediated through General Chemistry I grades (Figure 4).

The hypothesis was tested through Structural Equation Modeling (SEM) with the bootstrap approach. Table 6 presents the results from the formal mediation test. It is important to note that the SEM test included all critical distal outcomes and controls in the same model because the hypothesized mediator is the same for all outcomes and it is necessary to account for the correlations between the outcomes. In Table 6, the columns "Indirect Effect → Outcome" and "Condition → Outcome" show the key mediation results. If the effect of POGIL is at least partially direct (meaning it is caused by some other underlying mechanism), we would see significant coefficients in the "Condition → Outcome" column. If the effect is significantly

Table 6. Standardized Coefficients for Mediation Test for the Relationship of Condition to General Chemistry I Grades, General Chemistry I Grades to Each Outcome, the Remaining Direct Effect of Condition to Each Outcome, and the Estimated Indirect Contribution of Condition to Each Outcome

	Structural Equation Modeling Results[a]			
Distal Outcome	Condition → GenChem1	GenChem1 → Outcome	Indirect Effect → Outcome	Condition → Outcome
Gen Chem II grade	0.28[b]	0.81[b]	0.23[b]	0.03
Retention to Gen Chem II	0.28[b]	0.70[b]	0.11[b]	0.01
Fascination T2	0.28[b]	0.18[c]	0.05[c]	0.01
Competency Beliefs T2	0.28[b]	0.47[b]	0.13[b]	0.05
Identity T2	0.28[b]	0.50[b]	0.14[b]	0.03

[a]The SEM results include demographic and academic controls. [b]$p < 0.001$. [c]$p < 0.01$.

mediated by the General Chemistry I grade, we would see significant coefficients in the "Indirect Effect → Outcome".

Mediation results showed that performance in General Chemistry I appeared to be a core mediator of all the observed differences in General Chemistry II. In other words, students appeared to be more likely to enroll in the second course, feel more confident, have a stronger identity in chemistry, and do well in the second course because of their improved performance in the first course.

DISCUSSION

Despite the call for greater use for student-centered approaches in chemistry, there are very few studies focusing on understanding what chemistry-related attitudes are developed through these approaches. Furthermore, even though there is a lot of research focusing on the immediate effects of these interventions, there is less evidence on the delayed effects on retention and performance on follow-up courses when compared to other active learning approaches.[7]

https://dx.doi.org/10.1021/acs.jchemed.9b01052
J. Chem. Educ. XXXX, XXX, XXX−XXX

In this study, we focused on POGIL, a type of active learning that provides structured guidance to engage in inquiry learning. Our results support prior work showing that POGIL is more effective at increasing students' performance in General Chemistry I when compared with Traditional instruction.[10,11] However, we also show that POGIL is more effective at increasing students' chemistry-related attitudes, an often claimed but underexplored topic. In particular, there was a medium-sized effect on students' chemistry identity and a small effect on competency beliefs. These two results supported our proposed theoretical framework regarding the expected effectiveness given the specific structures within POGIL. However, to our surprise there were no important differences in chemistry fascination; since most students in this course are intending prehealth or biology majors, POGIL exercises with a health focus or application may have been needed to increase fascination.[43] In addition, the General Chemistry I grade proved to be a strong mediator of the differences in competency beliefs, which is to be expected given the strong relationship between competency beliefs and positive performance feedback. Differences in chemistry identity were also explained by grades, pointing toward identity as a dependent skill as well as knowledge development.[8,44,53]

Another important aspect of this study was related to showing distal effects. Once students left the active learning instruction classroom, they were still more likely to perform better rather than revert to performing at the same level. Furthermore, the mediation analysis showed that this increased performance was connected to their General Chemistry I grade. This finding is important because it provides evidence that the higher grades exhibited in General Chemistry I were not due to arbitrarily inflated grades from the instructors who were implementing active learning. Further, being a stronger mediator than other attitudinal factors points to the mechanism by which later performance rests: stronger developed knowledge or skill within the prior course. Finally, the differences in chemistry-related attitudes persisted to the end of General Chemistry II, meaning that returning to a Traditional environment did not erase these attitudinal gains. Such attitudinal effects are important because they influence student retention within STEM; indeed, differences in retention into General Chemistry II were observed.

■ LIMITATIONS

The aim of this study was to deepen understanding of the broad and potentially long-lasting effects of POGIL as a form of active learning instruction. This study used survey instruments to measure the effectiveness of the intervention. Survey instruments are an effective way to assess impact in this context: Given the length of the semester, it would be unlikely to see radical changes on attitude endorsement that could be captured in a small number of interviews. However, when it comes to measuring attitudes, the particular surveys did not examine how well students integrated their chemistry identity with other identities (e.g., as a woman, as an athletic person, as a biology kind of person), nor the ways in which identity was negotiated during the class or why some students report lower competency beliefs than others after the intervention. Further, these surveys are not sufficient to assess the day-to-day activities or systemic barriers that some students may have experienced. However, given the validation work done on our instruments, we are confident our surveys were useful in revealing the macrolevel effects of the intervention on students' motivation and grades.

Second, all results (both between section contrasts and mechanism analyses) are fundamentally correlational in nature. We did examine many plausible confounding variables and use strategic research design (e.g., temporal measurement of variables to determine directionality), and these decisions allow us to get closer to causality.

Finally, the results also are more likely to generalize to other similar contexts, such as large research-intensive universities (rather than teaching focused institutions that may already have more highly optimized non-POGIL instruction) serving predominantly white students (rather than more diverse institutions in which traditional underserved students are not a numerical minority). Mechanistically, similar benefits should also be observed in other contexts but may require other adaptations to the approach to match the needs of the context.

■ CONCLUSIONS

The aim of this study was to deepen understanding of the broad and potentially long-lasting effects of POGIL as a form of active learning instruction. Our results not only are consistent with the literature but also further show that active learning can have lasting effects on students into the follow-on course both in terms of performance and chemistry attitudes. Furthermore, we presented a new theoretical framework that ties specific aspects of POGIL structure to chemistry-related attitudes. This framework and supporting findings can guide other researchers to better design POGIL activities that support the development of chemistry competency beliefs and chemistry identity. Most importantly, we hope that this study assuages the worries of many practitioners about whether it is possible to implement POGIL in large-enrollment chemistry classes. As shown here, not only did instructors implement POGIL in a consistent manner across sections but also they better prepared their students for the follow-up course, even in their first implementation of this approach.

■ AUTHOR INFORMATION

Corresponding Author

Paulette Vincent-Ruz − *Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, United States;* orcid.org/0000-0002-8244-8700; Email: pvinruz@umich.edu

Authors

Tara Meyer − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0002-9810-454X

Sean G. Roe − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0001-6199-8773

Christian D. Schunn − *Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States;* orcid.org/0000-0003-3589-297X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jchemed.9b01052

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Kang, J.; Keinonen, T. The Effect of Student-Centered Approaches on Students' Interest and Achievement in Science:

Relevant Topic-Based, Open and Guided Inquiry-Based, and Discussion-Based Approaches. *Research in Science Education* **2018**, *48* (4), 865−885.
(2) National Academies of Science, Engineering, and Medicine. *How People Learn II*; National Academies Press, 2018. DOI: 10.17226/24783.
(3) Baepler, P.; Walker, J.; Driessen, M. It's Not about Seat Time: Blending, Flipping, and Efficiency in Active Learning Classrooms. *Computers & Education* **2014**, *78*, 227−236.
(4) Prince, M. Does Active Learning Work? A Review of the Research. *Journal of engineering education* **2004**, *93* (3), 223−231.
(5) Deslauriers, L.; McCarty, L. S.; Miller, K.; Callaghan, K.; Kestin, G. Measuring Actual Learning versus Feeling of Learning in Response to Being Actively Engaged in the Classroom. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (39), 19251−19257.
(6) Yezierski, E. J.; Bauer, C. F.; Hunnicutt, S. S.; Hanson, D. M.; Amaral, K. E.; Schneider, J. P. POGIL Implementation in Large Classes: Strategies for Planning, Teaching, and Management. In *Process Oriented Guided Inquiry Learning (POGIL)*; Moog, R. S., Spencer, J. P., Eds.; ACS Symposium Series; American Chemical Society, 2008; pp 60−71.
(7) Lewis, S. E. Investigating the Longitudinal Impact of a Successful Reform in General Chemistry on Student Enrollment and Academic Performance. *J. Chem. Educ.* **2014**, *91* (12), 2037−2044.
(8) Vacek, J. J. Process Oriented Guided Inquiry Learning (POGIL), a Teaching Method from Physical Sciences, Promotes Deep Student Learning in Aviation. *Collegiate Aviation Review International* **2011**, *29* (2). DOI: 10.22488/okstate.18.100424
(9) Vincent-Ruz, P.; Binning, K.; Schunn, C. D.; Grabowski, J. The Effect of Math SAT on Women's Chemistry Competency Beliefs. *Chem. Educ. Res. Pract.* **2018**, *19* (1), 342−351.
(10) Lo, S. M.; Mendez, J. I. "L" Learning—The Evidence. In *POGIL: An Introduction to Process Oriented Guided Inquiry Learning for Those Who Wish to Empower Learners*; Simonson, S. R., Ed.; Stylus Publishing, LLC, 2019.
(11) Brown, P. J. Process-Oriented Guided-Inquiry Learning in an Introductory Anatomy and Physiology Course with a Diverse Student Population. *Advances in physiology education* **2010**, *34* (3), 150−155.
(12) Farrell, J. J.; Moog, R. S.; Spencer, J. N. A Guided-Inquiry General Chemistry Course. *J. Chem. Educ.* **1999**, *76* (4), 570.
(13) Schroeder, J. D.; Greenbowe, T. J. Implementing POGIL in the Lecture and the Science Writing Heuristic in the Laboratory—Student Perceptions and Performance in Undergraduate Organic Chemistry. *Chem. Educ. Res. Pract.* **2008**, *9* (2), 149−156.
(14) Cannady, M. A.; Vincent-Ruz, P.; Chung, J. M.; Schunn, C. D. Scientific Sensemaking Supports Science Content Learning across Disciplines and Instructional Contexts. *Contemporary Educational Psychology* **2019**, *59*, 101802.
(15) Bathgate, M.; Crowell, A.; Schunn, C.; Cannady, M.; Dorph, R. The Learning Benefits of Being Willing and Able to Engage in Scientific Argumentation. *International Journal of Science Education* **2015**, *37* (10), 1590−1612.
(16) Lesh, R.; Doerr, H. M. Foundations of a Models and Modeling Perspective on Mathematics Teaching, Learning, and Problem Solving. *Beyond Constructivism: Models and Modeling Perspectives on Mathematics Problem Solving, Learning, and Teaching*; Routledge, 2003; pp 3−33.
(17) Haak, D. C.; HilleRisLambers, J.; Pitre, E.; Freeman, S. Increased Structure and Active Learning Reduce the Achievement Gap in Introductory Biology. *Science* **2011**, *332* (6034), 1213−1216.
(18) Walker, L.; Warfa, A.-R. M. Process Oriented Guided Inquiry Learning (POGIL®) Marginally Effects Student Achievement Measures but Substantially Increases the Odds of Passing a Course. *PLoS One* **2017**, *12* (10), No. e0186203.
(19) Belland, B. R.; Walker, A. E.; Kim, N. J.; Lefler, M. Synthesizing Results from Empirical Research on Computer-Based Scaffolding in STEM Education: A Meta-Analysis. *Review of Educational Research* **2017**, *87* (2), 309−344.
(20) Chase, A.; Pakhira, D.; Stains, M. Implementing Process-Oriented, Guided-Inquiry Learning for the First Time: Adaptations and Short-Term Impacts on Students' Attitude and Performance. *J. Chem. Educ.* **2013**, *90* (4), 409−416.
(21) Stanford, C.; Moon, A.; Towns, M.; Cole, R. Analysis of Instructor Facilitation Strategies and Their Influences on Student Argumentation: A Case Study of a Process Oriented Guided Inquiry Learning Physical Chemistry Classroom. *J. Chem. Educ.* **2016**, *93* (9), 1501−1513.
(22) Lamnina, M.; Chase, C. C. Developing a Thirst for Knowledge: How Uncertainty in the Classroom Influences Curiosity, Affect, Learning, and Transfer. *Contemporary Educational Psychology* **2019**, *59*, 101785.
(23) Allchin, D. Teaching the Nature of Science through Scientific Errors. *Sci. Educ.* **2012**, *96* (5), 904−926.
(24) Giunta, C. J. Using History to Teach Scientific Method: The Role of Errors. *J. Chem. Educ.* **2001**, *78* (5), 623.
(25) Cohen, G. L.; Garcia, J. Identity, Belonging, and Achievement: A Model, Interventions, Implications. *Current Directions in Psychological Science* **2008**, *17* (6), 365−369.
(26) Fixsen, D. L.; Naoom, S. F.; Blase, K. A.; Friedman, R. M. *Implementation Research: A Synthesis of the Literature*; National Implementation Research Network, University of South Florida: Tampa, FL, 2005.
(27) Stains, M.; Vickrey, T. Fidelity of Implementation: An Overlooked yet Critical Construct to Establish Effectiveness of Evidence-Based Instructional Practices. *CBE—Life Sciences Education* **2017**, *16* (1), No. rm1.
(28) Moir, T. Why Is Implementation Science Important to Intervention Design and Evaluation, within Educational Settings? *Frontiers in Education* **2018**, *3*, 61.
(29) Perry, M. D.; Wight, R. D. Using an ACS General Chemistry Exam to Compare Traditional and POGIL Instruction. *Process Oriented Guided Inquiry Learning (POGIL)*; American Chemical Society; 2008; Vol. *994*, pp 240−247. .
(30) Vishnumolakala, V. R.; Southam, D. C.; Treagust, D. F.; Mocerino, M.; Qureshi, S. Students' Attitudes, Self-Efficacy and Experiences in a Modified Process-Oriented Guided Inquiry Learning Undergraduate Chemistry Classroom. *Chem. Educ. Res. Pract.* **2017**, *18* (2), 340−352.
(31) Warburton, E. C.; Bugarin, R.; Nunez, A.-M. *Bridging the Gap: Academic Preparation and Postsecondary Success of First-Generation Students*; Statistical Analysis Report. Postsecondary Education Descriptive Analysis Reports; National Center for Education Statistics, 2001.
(32) Sadler, P. M.; Tai, R. H. Advanced Placement Exam Scores as a Predictor of Performance in Introductory College Biology, Chemistry and Physics Courses. *Science Educator* **2007**, *16* (2), 1−19.
(33) Tai, R. H.; Sadler, P. M.; Mintzes, J. J. Factors Influencing College Science Success. *Journal of College Science Teaching* **2006**, *36* (1), 52.
(34) Ferrini-Mundy, J.; Gaudard, M. Secondary School Calculus: Preparation or Pitfall in the Study of College Calculus? *J. Res. Math Educ* **1992**, *23* (1), 56−71.
(35) Chajewski, M.; Mattern, K. D.; Shaw, E. J. Examining the Role of Advanced Placement® Exam Participation in 4-Year College Enrollment. *Educ Meas Issues Pract* **2011**, *30* (4), 16−27.
(36) Contreras, F. E. Access, Achievement, and Social Capital: Standardized Exams and the Latino College-Bound Population. *J. Hispanic High Educ* **2005**, *4* (3), 197−214.
(37) Park, J. J. It Takes a Village (or an Ethnic Economy): The Varying Roles of Socioeconomic Status, Religion, and Social Capital in SAT Preparation for Chinese and Korean American Students. *Am. Educ Res. J.* **2012**, *49* (4), 624−650.
(38) Venezia, A.; Kirst, M. W. Inequitable Opportunities: How Current Education Systems and Policies Undermine the Chances for Student Persistence and Success in College. *Educational Policy* **2005**, *19* (2), 283−307.
(39) Angrist, J. D.; Cohodes, S. R.; Dynarski, S. M.; Pathak, P. A.; Walters, C. R. Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice. *J. Labor Econ* **2016**, *34* (2), 275−318.

https://dx.doi.org/10.1021/acs.jchemed.9b01052
J. Chem. Educ. XXXX, XXX, XXX−XXX

(40) Dorans, N. J. Correspondences between ACT™ and SAT® I Scores. *ETS Research Report Series* **1999**, *1999* (1), i−18.
(41) Campbell, C. M.; Cabrera, A. F. Making the Mark: Are Grades and Deep Learning Related? *Research in Higher Education* **2014**, *55* (5), 494−507.
(42) Severiens, S.; Wolff, R. A Comparison of Ethnic Minority and Majority Students: Social and Academic Integration, and Quality of Learning. *Stud High Educ* **2008**, *33* (3), 253−266.
(43) Ramsey, P.; Franklin, T.; Ramsey, D.; Wells, R. Rethinking Grades for Sustainable Learning. *Innov Educ Teach Int.* **2002**, *39* (2), 117−123.
(44) Allen, J. D. Grades as Valid Measures of Academic Achievement of Classroom Learning. *Clearing House: A Journal of Educational Strategies, Issues and Ideas* **2005**, *78* (5), 218−223.
(45) Sabot, R.; Wakeman-Linn, J. Grade Inflation and Course Choice. *Journal of Economic Perspectives* **1991**, *5* (1), 159−170.
(46) Berry, C. M.; Sackett, P. R. Individual Differences in Course Choice Result in Underestimation of the Validity of College Admissions Systems. *Psychol Sci.* **2009**, *20* (7), 822−830.
(47) Witherspoon, E. B.; Vincent-Ruz, P.; Schunn, C. D. When Making the Grade Isn't Enough: The Gendered Nature of Premed Science Course Attrition. *Educ Researcher* **2019**, *48* (4), 193−204.
(48) Witherspoon, E. B.; Schunn, C. D. Locating and Understanding the Largest Gender Differences in Pathways to Science Degrees. *Sci. Educ.* **2020**, *104*, 144−163.
(49) Kalender, Z. Y.; Marshman, E.; Schunn, C. D.; Nokes-Malach, T. J.; Singh, C. Why Female Science, Technology, Engineering, and Mathematics Majors Do Not Identify with Physics: They Do Not Think Others See Them That Way. *Phys. Rev. Phys. Educ Res.* **2019**, *15* (2), 020148.
(50) Liu, Y.; Ferrell, B.; Barbera, J.; Lewis, J. E. Development and Evaluation of a Chemistry-Specific Version of the Academic Motivation Scale (AMS-Chemistry). *Chem. Educ. Res. Pract.* **2017**, *18* (1), 191−213.
(51) Ferrell, B.; Phillips, M. M.; Barbera, J. Connecting Achievement Motivation to Performance in General Chemistry. *Chem. Educ. Res. Pract.* **2016**, *17* (4), 1054−1066.
(52) Ferrell, B.; Barbera, J. Analysis of Students' Self-Efficacy, Interest, and Effort Beliefs in General Chemistry. *Chem. Educ. Res. Pract.* **2015**, *16* (2), 318−337.
(53) Hosbein, K. N.; Barbera, J. Alignment of Theoretically Grounded Constructs for the Measurement of Science and Chemistry Identity. *Chem. Educ. Res. Pract.* **2020**, *21* (1), 371−386.
(54) Vincent-Ruz, P.; Schunn, C. D. The Nature of Science Identity and Its Role as the Driver of Student Choices. *International journal of STEM education* **2018**, *5* (1), 48.
(55) Heredia, K.; Lewis, J. E. A Psychometric Evaluation of the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *J. Chem. Educ.* **2012**, *89* (4), 436−441.
(56) Adams, W. K.; Wieman, C. E.; Perkins, K. K.; Barbera, J. Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *J. Chem. Educ.* **2008**, *85* (10), 1435.
(57) Lewis, S. E.; Shaw, J. L.; Heitz, J. O.; Webster, G. H. Attitude Counts: Self-Concept and Success in General Chemistry. *J. Chem. Educ.* **2009**, *86* (6), 744.
(58) Bauer, C. F. Beyond" Student Attitudes“: Chemistry Self-Concept Inventory for Assessment of the Affective Component of Student Learning. *J. Chem. Educ.* **2005**, *82* (12), 1864.
(59) Desimone, L. M.; Le Floch, K. C. Are We Asking the Right Questions? Using Cognitive Interviews to Improve Surveys in Education Research. *Educational evaluation and policy analysis* **2004**, *26* (1), 1−22.
(60) Cirino, P. T.; Rashid, F. L.; Sevcik, R. A.; Lovett, M. W.; Frijters, J. C.; Wolf, M.; Morris, R. D. Psychometric Stability of Nationally Normed and Experimental Decoding and Related Measures in Children with Reading Disability. *Journal of learning disabilities* **2002**, *35* (6), 526−539.