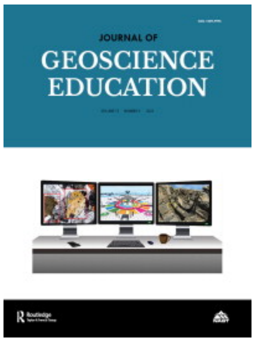


Journal of Geoscience Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ujge20

# Student knowledge gains in polar literacy and statistics after completing guided inquiry modules in an undergraduate statistics course

**Penny M. Rowe, James Bernhard, Jacob Price, Anoushka Adhav, Danielle Dolan, Anna Van Boven, Lea Fortmann, Michael Town & Steven Neshyba**

**To cite this article:** Penny M. Rowe, James Bernhard, Jacob Price, Anoushka Adhav, Danielle Dolan, Anna Van Boven, Lea Fortmann, Michael Town & Steven Neshyba (12 Nov 2024): Student knowledge gains in polar literacy and statistics after completing guided inquiry modules in an undergraduate statistics course, Journal of Geoscience Education, DOI: 10.1080/10899995.2024.2426438

**To link to this article:** https://doi.org/10.1080/10899995.2024.2426438

JOURNAL OF GEOSCIENCE EDUCATION
https://doi.org/10.1080/10899995.2024.2426438



# Student knowledge gains in polar literacy and statistics after completing guided inquiry modules in an undergraduate statistics course

Penny M. Rowe[a], James Bernhard[b], Jacob Price[b], Anoushka Adhav[c], Danielle Dolan[b], Anna Van Boven[b], Lea Fortmann[d], Michael Town[e] and Steven Neshyba[f]

[a]NorthWest Research Associates, Seattle, Washington, USA; [b]Department of Mathematics and Computer Science, University of Puget Sound, Tacoma, Washington, USA; [c]College of the Environment, School of Environmental and Forest Sciences, University of Washington, Seattle, Washington, USA; [d]Economics Department, University of Puget Sound, Tacoma, Washington, USA; [e]Earth and Space Research, Seattle, Washington, USA; [f]Chemistry Department, University of Puget Sound, Tacoma, Washington, USA

**ABSTRACT**

Climate change is a major concern to undergraduate students. Understanding climate change relies on an understanding of polar regions. However, courses on polar regions are rare at undergraduate institutions. Polar ENgagement through GUided INquiry (PENGUIN) modules were designed to give students experience with polar research in a variety of standard courses, including physics, computer science, physical chemistry, and economics, through using course-specific and computational tools to analyze polar data. Here, we present a new PENGUIN module taught in a statistics class, in which students apply statistical tools to ice core data to reconstruct past temperature records. Quantitative student responses on pre- and post-surveys were collected in a quasi-experimental context to assess student knowledge gains for a test group of 91 students and a control group of 73 students (who did not complete the module). Test-group students made statistically significant increases of 25 to 46% on all six statistics questions, with a normalized gain of 56%. By contrast, control group statistics knowledge gains ranged from −4 to 25%, with statistically significant increases for only three questions and a normalized gain of 22%. For polar research questions, the test group demonstrated increases in correct responses to polar research questions (11 to 31%), with statistically significant improvements ($p < .05$) of 22-31% on 3 of 6 polar research questions. These findings support the conclusion that PENGUIN modules can successfully teach course concepts while increasing polar literacy.

**ARTICLE HISTORY**
Received 24 July 2023
Revised 4 November 2024
Accepted 4 November 2024

**KEYWORDS**
Computational guided inquiry; polar research; climate literacy; undergraduate education; R

## Introduction

Undergraduate students express a high level of concern about climate change (Bedford, 2016), and interest in sustainability programs is growing, together with the number of degrees and programs offered (National Academies of Sciences et al., 2020). Central to understanding climate change is an understanding of polar regions (U.S. Global Change Research Program (USGCRP), 2009), where climate change is amplified through a variety of feedback mechanisms that have led to warming about four times the global rate since 1979 (Rantanen et al., 2022). While undergraduate institutions are increasingly offering courses that include climate change, courses that include polar research are nearly absent (Klyce & Ryker, 2023). Furthermore, the cost and logistical challenges of providing research experiences in polar regions to undergraduates make it difficult to reach large numbers of students (Ham & Flood, 2009).

The PENGUIN (Polar ENgagement through GUided INquiry) project has sought to address these deficits by bringing polar research into undergraduate classrooms in established courses like physics, chemistry, and statistics. PENGUIN modules have several key components. They are designed to satisfy course disciplinary learning goals defined by instructors, who often co-create the modules. They are taught within the context of climate change (Rowe et al., 2020). Finally, they give students hands-on experience analyzing, interpreting and visualizing real-world polar data using a computational tool such as Excel, R, or Python. Seven PENGUIN modules were previously developed (available online through the Science Education Resource Center at Carleton College; SERC; https://serc.carleton.edu/penguin), ranging from a spreadsheet module in Economics examining the risk/benefit analysis of a sea wall to address sea level rise linked to polar ice melt (Fortmann et al., 2020), to a Python module in Physics examining heat flux through permafrost (Rowe et al., 2020).

An essential component of PENGUIN modules is data literacy, which lies at the intersection of quantitative, computational, and disciplinary skills (Kjelvik & Schultheis, 2019). Giving students access to authentic data has been shown to improve data literacy (Gould et al., 2014; Kastens et al., 2015; Kjelvik & Schultheis, 2019) and has the potential to engage students in

**CONTACT** Penny M. Rowe penny@nwra.com NorthWest Research Associates, Post Office Box 85535, Seattle, WA 98145-1535, USA.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10899995.2024.2426438.

math and science and improve critical thinking skills such as analyzing and interpreting data, forming arguments based on evidence, and computational and mathematical thinking (Gould et al., 2014; Kastens et al., 2015; Kjelvik & Schultheis, 2019 and references therein; Mosher & Keane, 2021). Although there is a trend toward teachers using larger, 'messy' data sets to teach data literacy and the scientific process in high school classrooms (e.g., Hammett & Dorsey, 2020), recent research suggests that in K-12 education, data analysis typically still involves small datasets collected by students and analyzed *via* calculators and spreadsheets (Rosenberg et al., 2022). Students and instructors only authentically engage with computational tools currently in use by computer and research scientists when the data sets become too large to 'brute force' (Rosenberg et al., 2022). PENGUIN modules provide instructors with access to curated, research-grade data associated with tangible, developmentally appropriate learning goals.

PENGUIN modules share many design similarities with other data focused modules, such as project EDDIE modules (e.g., Carey et al., 2020; Carey & Gougis, 2017; Klug et al., 2017; Soule et al., 2018). However, a key distinction is that PENGUIN modules are intended to improve polar literacy and give students hands-on experience with polar research and data in a wide variety of courses, including courses outside the Earth Sciences.

Prior work (Rowe et al., 2020) demonstrated positive outcomes after working through PENGUIN modules, with students reporting that they enjoyed the module, and students and professors alike feeling that students made significant learning gains and increased their comfort with the computational tool used (spreadsheet or Python). In addition, PENGUIN modules successfully brought polar research to students who reported little to no previous exposure. However, Rowe et al. (2020) and other literature (Madison, 2014) point to a need to demonstrate learning gains quantitatively. Previous work has made important progress in meeting this goal. Studies of student learning after engaging in project EDDIE modules have demonstrated gains in quantitative literacy (Klug et al., 2017), statistical learning and ability to differentiate plausible from unlikely variability in a dataset and understanding of seismological concepts (Soule et al., 2018), ability to provide evidence of systems thinking and proficiency in working with ecosystem models (Carey et al., 2020), and understanding of climate change content (Carey & Gougis, 2017).

The present work builds on these previous studies to quantitatively measure the efficacy of a new PENGUIN module, developed for an introductory statistics course, in teaching students preexisting statistics learning goals in addition to polar literacy goals. The duration of the intervention was increased relative to prior PENGUIN modules, to span approximately two weeks (~8 50-minute class sessions), to increase the ability of students to make significant and sustained knowledge gains. The module was developed for an introductory statistics course in collaboration with statistics professors. Assessment tools were designed to specifically test student knowledge gains in statistics and polar literacy and pre- and post-module assessments were administered to intervention and control groups.

## Purpose and learning goals

Our overarching goal in this paper is to contribute to a larger body of work assessing the effectiveness of PENGUIN modules in teaching students preexisting course learning goals while increasing polar literacy. More specifically, previous work (Rowe et al., 2020) has demonstrated through attitudinal surveys that PENGUIN modules are effective in teaching students preexisting course learning goals while increasing polar literacy. The purpose of the present paper is to provide evidence that supports the conclusion that PENGUIN modules also lead to quantifiable knowledge gains. To do this, a new PENGUIN module was developed, entitled "Statistics: predicting temperature." Statistics and polar literacy learning goals were defined and questions were included in pre and post surveys to test student knowledge gains. The module was developed using the same methodology as prior modules. As a gauge of similarity between the new module and the prior ones, the questions evaluating student perceptions in the prior surveys were also included in the new survey (Rowe et al., 2020).

Statistics learning goals of the module are as follows. After completing the module, students should 1) be able to identify the strength and form of linear association in scatterplots; 2) understand strength of sample correlation as expressed by $r$; 3) understand how the average connection between the value of a response and explanatory variable relates to a linear regression model; 4) be able to articulate the validity of interpolation versus extrapolation in a simple linear regression model, and 5) understand that such a model is insufficient to provide evidence of causation, but rather a physical model is needed.

The primary polar literacy goals of the module are 1) to know that warming is amplified in polar regions; 2) to be able to identify the primary positive feedback associated with this amplification; 3) to recognize that polar amplification means that temperature is increasing faster with increasing $CO_2$; 4) to know that the atmospheric $CO_2$ concentration was much lower before the modern age; 5) to know that the last million years was mostly spent in ice ages with brief warm periods like the present; 6) to understand how ice cores can be used to reconstruct atmospheric temperature and $CO_2$ records, including that isotope abundances are used to determine past temperature; and 7) to know the maximum timespan ice cores have been used to reconstruct timeseries. Goals 1-6 link to climate literacy principles (U.S. Global Change Research Program (USGCRP), 2009) and goals 1, 2, 3 and 6 link to polar literacy principles (McDonnell et al., 2020; https://polar-ice.org/ retrieved 2024/02/14), as detailed in Table S1 in the Supplemental Materials.

## Materials and implementation

### Prior PENGUIN modules

Prior to this work, seven PENGUIN modules were developed (Table 1; see also Rowe et al., 2020). PENGUIN modules are described in detail in Rowe et al. (2018), Fortmann et al. (2020), and Rowe et al. (2020) and are available online at https://serc.carleton.edu/penguin. These modules typically spanned a few class sessions or a long lab period. They were developed by a curriculum development team including polar researchers, education researchers, and instructors who teach the courses. Module development is typically initiated in a backward design process (Wiggins & McTighe, 2018), in that development starts with instructors identifying concepts taught in the course, followed by brainstorming polar research that can be applied and polar literacy concepts.

PENGUIN modules use guided inquiry, in which students conduct inquiry into a scientific topic under the guidance of the instructor (Apedoe et al., 2006; Caspari et al., 2007; Grissom et al., 2015; Lewis & Lewis, 2008; Martin-Hansen, 2002; Weaver et al., 2008). In guided inquiry, students formulate questions, think about what kinds of data and analysis are needed, collect and analyze data, and derive and share conclusions (Jackson et al., 2008; Wells et al., 1995; Windschitl, 2008). In addition, the modules use active learning, in which students learn through working on an activity rather than listening passively to a lecture (Freeman et al., 2014). These approaches have been shown to lead to improvements in student performance, growth, and retention (Apedoe et al., 2006; Freeman et al., 2014; Grissom et al., 2015; Lewis & Lewis, 2008; Weaver et al., 2008). For PENGUIN modules, active learning takes place when students use a computational tool to analyze polar data, which typically occurs while working individually or in pairs at a computer. The instructor typically walks around the room as the students work, guiding inquiry through answering student questions and discussing findings at key points in the activity. Both active learning and guided inquiry occur during small-group and class discussions. Additional guidance comes from the professor through presentations that introduce background and key concepts.

**Table 1.** Previously developed PENGUIN modules. (See Rowe et al., 2020 for more detail).

| Module name | Polar topic |
|---|---|
| Economics: Total economic valuation of the Arctic | Evaluate ecosystem service losses in Arctic attributable to climate change. Read and discuss journal articles related to polar ecosystem services. |
| Economics: Sea level rise | Establish connections between polar ice melt and the effect of sea level rise on a coastal city. |
| Quantum Mechanics: Polar spectra | Develop an understanding of role of temperature and greenhouse gases, particularly water vapor, in the unique polar atmosphere. |
| Thermodynamics: Sea ice melt | Develop awareness of observatories and datasets in the arctic and the effect of climate change on Arctic sea ice. |
| Physics: Permafrost | Learn how permafrost responds to climate change and the consequences of thawing permafrost on the Arctic. |
| Computer Science: Images of Arctic Ice | Learn about and use data from earth observing satellites to examine how arctic sea ice responds to climate change. Learn the role of the ice-albedo effect. |
| Environmental Science: Ice Cores | Know that past temperature and $CO_2$ records over millions of years, and correlations between them, can be determined from polar ice cores. |

Here we focus on a new module to evaluate the effectiveness of PENGUIN modules for student learning. The new module guides students in applying statistics to modern temperature records and records from ice core data and is available online (https://serc.carleton.edu/penguin/modules/statistics_predicting_temperat.html). This module was taught in an introductory statistics class that has high enrollment in order to give a greater number of potential students in the test and control groups. It was developed using the same design process as previously developed modules.

## *Statistics: predicting temperature*

The new module spans several weeks of class time and is divided into four parts. Each part focuses on a set of statistics learning goals and a polar research question. Students are guided through the inquiry steps outlined above by working through the following questions: What is the research question and why should we care about it? What data are needed to answer the question? How and where can the data be collected? How should errors and outliers be handled? To what extent is the analysis valid? Students work through these questions, learn background polar research material and statistics tools and concepts, and perform analysis through a combination of lecture (*via* PowerPoint presentations given by the instructors), hands-on activities in R, in-class discussions, and reading and analyzing a journal article. Students spend an estimated 1/3 of their in-class time working actively. They are encouraged to work in pairs but may choose to work individually. Because the class is introductory, all data are provided for the students in an Rdata file (instructions for downloading Arctic data are shared with the instructor).

Through a PowerPoint presentation, students begin by learning about the consequences of climate change, such as how melting of polar land ice leads to sea level rise, and how thawing Arctic permafrost causes infrastructure damage. They also learn about polar amplification *via* feedbacks that amplify Arctic climate change, including the ice-albedo feedback and the feedback mechanism by which permafrost thaw releases methane, a potent greenhouse gas. This leads to the first question: how fast is climate changing in the Arctic compared to the globe as a whole? The statistics content focuses on learning about scatterplots, including form, direction, and strength of association. In an in-class exercise, students use R in RStudio to create and analyze scatterplots of Arctic temperature with year (from Utqiagvik, Alaska) as well as the global average temperature anomaly with year. The first part ends with a discussion of the validity of the analysis.

The second part of the module addresses the question of how climate has changed in the distant past. Following a presentation of background material, students discuss what data are needed and how it can be collected, leading to a presentation about how isotopes in ice cores can be used to reconstruct the temperature record. They learn about interpreting the coefficients of simple linear regression and about identifying and dealing with outliers. They use R to determine the coefficients of simple linear regression and discuss in small groups questions asked when discovering an association (measurements needed, where/how to collect them, how to test the association). As a homework assignment, they read a journal article (Dahe et al., 1994) and use R to examine the association between the isotopic abundance difference and temperature measured by Dahe et al. (1994); perform the linear regression; and interpret the coefficients. A journal article from the 1990s, before online journals and data became ubiquitous, was purposely chosen because the data are included within the print article itself, giving students easy access to explore real data, including identification of an outlier due to a printing typo.

The third part of the module continues with the question of how climate has changed in the distant past, but with the statistics focus on prediction. The homework is reviewed, including the validity of their results: because the isotopic concentration and temperature are strongly correlated, prediction within the range is valid. As part of in-class activities in R, they apply the linear regression model they developed to predict temperature from isotope abundance difference in the ice core record and compare their predicted temperatures to literature values, discussing possible reasons for differences.

In the final part of the module, students learn about the variations in temperature over the last 800,000 years from the ice core record and how these variations are enhanced by $CO_2$. Students learn about how $CO_2$ causes greenhouse warming and how $CO_2$ and temperature are correlated in the ice core record, in the modern global average, and in polar regions. This is followed by a discussion of correlation and causation, and the need for a predictive physical model to infer causation. Students also learn about sample correlation, how correlation does not prove causation, and that a model built on regression cannot be used to extrapolate outside the range of the data, but rather a physically-based predictive model is needed. Students plot $CO_2$ versus time and temperature versus time in the ice core record and note that they appear correlated. They then plot $CO_2$ versus temperature and compute the correlation coefficient. Finally, they repeat this for the modern polar site (Utqiagvik, Alaska) and modern globe and compare all three.

### Assessment mechanisms for instructors

Assessment mechanisms are composed of the homework assignment, student in-class activities, and a final quiz. Keys are provided for all assessments. The homework assignment and activity keys include short-answer responses and R coding inputs and outputs, as well as figures that are produced.

### Study population and setting

The module was taught in-person in a liberal arts college in the U.S. during 2021-2022. All courses in which the module was taught were sections of Math 160, Introduction to Statistics, taught at the University of Puget Sound, with typically about 24 students per class, using the following course description: "This course provides an introduction to statistics, concentrating on statistical concepts and the 'why and when' of statistical methodology. The course focuses on learning to ask appropriate questions, collect data effectively, summarize and interpret information, and understand the limitations of statistical inference." This description is in line with our inquiry-based framework.

Two instructors taught the module to the test group in 7 statistics classes (91 survey respondents; response rate of 55%). In the first semester of the study, the polar- and climate-relevant parts of the module were taught to the students by a guest lecturer who is a woman polar researcher, while in the second two semesters it was taught by the test-group instructors (both men). Three different instructors (one man and two women) taught the control group, which consisted of students in 5 statistics classes who did not work through the module (73 survey respondents; response rate of 63%) in the first year of the study.

Demographics for students in the test group are shown in Figure 1. Only students 18 and over were surveyed. The majority of students identified as white (79%), Asian (9%), and Hispanic/Latino/a (9%), and gender identity was mainly split between women (51%) and men (43%). Of note for this study, most students reported having no (36%) to little (36%) prior exposure to polar research, and the majority of students were not STEM majors (73%). Demographic information was not collected in the control group but is expected to be similar since students in all classes were from the same university.

### Comparison between test and control groups

The two instructors who taught the test group helped develop the module, while the other instructors did not. All instructors were provided with a copy of the pre-survey when it was administered. Control group instructors indicated that they covered all topics in the pre-survey, with some differences as follows:

In control group 1 (instructor 3; 22% of control group respondents) the instructor indicated that the class was mainly lecture-driven, that they did not discuss interpolation in class (interpolation was used in a survey question associated with goal 4) and that they generally used the term "least square regression" rather than "linear regression" (associated with goals 3 and 4). Students worked with real-world datasets from the textbook and from a survey of math students, which was collected, analyzed, and presented in a project. Students learned basic R and spent 1 to 4 h per week using R.

In control group 2 (instructor 4; 37% of control group respondents) the instructor indicated they taught all information on the pre-survey questions with no major differences in vocabulary and that the class was mainly lecture-driven. Students used real-world datasets from the textbook and did not use a computational tool apart from a calculator.

In control group 3 (instructor 5; 41% of control group respondents), the instructor indicated that they touched on all

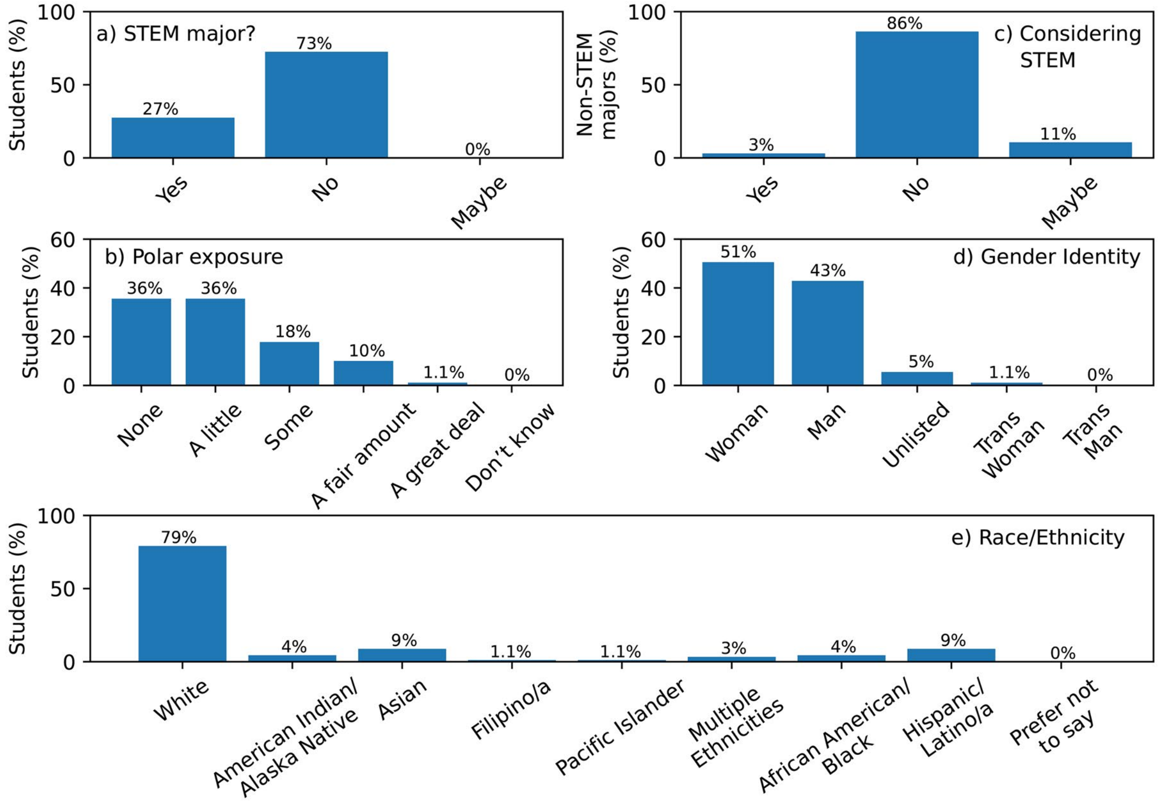


**Figure 1.** Demographics of the 91 survey respondents who completed the PENGUIN module. Full questions are given in the surveys. For panel e, students could select more than one choice, so percentages total to more than 100%.

ideas in the survey with no significant differences, except that the textbook refers to "predictor and response" rather than "explanatory and response" variables (used in statistics questions related to learning goals 3 and 4). The class was a mix of lecture and group work on worksheets. Students used real-world datasets from the textbook and from a survey of math students and made significant use of R both in class and in projects.

## Evaluation

### Overall design and strategy

The overall design of our study is a quantitative, quasi-experimental design, in which a group of students are taught a set of statistics topics *via* a PENGUIN module that also includes polar research and data, while another group is taught without using the module. A quasi-experimental study is one in which existing groups of participants are used, rather than randomly assigning participants to test and control groups. Here the two groups of students are students in the same statistics course at the same college, but with two different sets of instructors: one group of two instructors taught the course with the module, while the other group of three instructors taught the course without the module. Additional differences exist between how the instructors taught the courses. This design was chosen to maximize the number of students in the two groups, since the three control-group instructors preferred not to teach the module but were willing to administer surveys.

Pre- and post-survey knowledge test results were collected and analyzed and quantitative improvements for the test group and control group were compared. To test knowledge gains, students were given eight multiple-choice polar questions and six multiple-choice statistics questions on pre- and post-surveys. The post-survey given to students in the experiment group also asked questions that assessed student perceptions of knowledge gains and what they liked and did not like about the module. These student-perception questions were the same as in previously created modules, allowing for comparison to determine if students overall viewed the modules the same (Rowe et al., 2020).

### Data sources and collection

In the first semester of the study (Fall 2021), the pre-survey was administered to all classes early in the semester (in the third week of class) and the post-survey was administered near the end (~3 to 4 wk before classes ended). This ensured that the treatment (or relevant statistics material for the control classes) occurred between the pre- and post-surveys. The PENGUIN module was taught six weeks after the pre-survey, whereas the relevant statistics material was taught in control group classes at various points throughout the semester. This led to a time span between pre- and post-surveys of about 10 wk. In the second two semesters of the study (Spring and Fall 2022), only test groups were taught. For these test groups, there were 4 to 6 wk between the pre- and post-surveys.

The method for administering the survey was as follows. First, a script about the survey (provided in Supplemental Materials) was read to students in-person by one of the authors. In the first year, a cash incentive was provided to students, although in the second year the cash incentive was discontinued. Students were then provided with an online link to the survey, given in-class time to complete it with the option of finishing as needed after class (with the exception of one experiment group in Spring 2022 in which it was assigned as homework).

### Data analysis, validity and reliability

Pre- and post-surveys were matched based on anonymized identifying information requested in the survey. Only surveys that could be matched were retained for determining knowledge gains. For the test group, there were 129 students in the pre-survey, 105 in the post-survey, and 91 matched students. For the control group, these numbers were 94, 87, and 73, respectively. Student knowledge gains from pre- to post-survey were then assessed using Fisher's exact test to compute the p-value and the odds ratio. The null hypothesis is that the odds of answering correctly are the same before and after the intervention. The odds ratio is a measure of effect size that is appropriate for binary data (Ialongo, 2016). It is defined as OR = $(A_c/A_i)/(B_c/B_i)$, where each term gives the number of correct (subscript c) or incorrect (subscript i) answers either before (B) or after (A) the intervention. OR = 1 indicates no correlation, OR < 1 indicates negative correlation, and OR > 1 indicates positive correlation. OR > 1 is consistent with students making knowledge gains, with a higher odds ratio indicating greater likelihood of an effect.

An additional effect size was also computed: the normalized gain (Hake 1998). The normalized gain is defined as (<post>—<pre>)/(100—<pre>), where brackets refer to the class average (as a percent). It can thus be thought of as "the fraction of concepts learned by a class that were not known at the beginning of the course," and has the advantage that classes with different averages can have the same value (Coletta & Steinert, 2020).

Regarding reliability, there were differences between the experiment and control groups that could cause different outcomes (described in the comparison between control and experiment groups section above). To assess potential biases, the analysis described above was repeated for different subsets of the survey data: by group (test or control), by instructor, and by semester. The experiment group was also subsetted by STEM versus non-STEM and men versus women. Given the small proportion and racial diversity of students who identified as non-white, subsetting by race/ethnicity was not done.

## Results

### Statistics knowledge gains

Figure 2 shows student responses to questions on statistics topics for the pre- and post-surveys and for the test and control group. Summary statistics and full survey questions are given in Table 2 for the test group, while Table 3 gives summary statistics for the control group. Table 2 gives survey questions in full, as well as abbreviations, whereas in Table 3 and Figure 2 only abbreviations are used, for brevity. Statistics survey questions are numbered in the table according to their associated statistics learning goals, with two questions (4a and 4b) associated with goal 4.

Students in the test group made statistically significant ($p < .05$) improvements on all statistics test questions. Moreover, students in the test group had greater gains on the statistics questions than students in the control group. For the control group (Table 3), improvements were statistically significant ($p < .05$) for three questions and odds ratios varied from 0.8 to 3.4, compared to statistically significant improvements for all questions and odds ratios of 2.6 to 8.0 for the test group. Differing outcomes for the two treatments were also evident in the normalized gains, which were 56% for the test group and 22% for the control group.

For subsets of the test group, no major differences were found between STEM and non-STEM majors (normalized gains of 50% and 59%; see also Figure S3 of the Supplemental Materials), or for students who identify as men or women (normalized gains of 57% for each; see also Figure S4 of the Supplemental Materials). For the test group, there was a small drop in normalized gain over the three semesters of the survey (normalized gains of 65%, 53%, and 51%, chronologically). However, when the results are subsetted by semester and instructor, a more complicated picture emerges: for the instructor who taught the module in all three semesters, the normalized gains were 65%, 47%, and 63%, chronologically, while for the other instructor, who taught the module in the second two semesters, the normalized gains were 58% and 43%. (Compositing by student demographics or semester for the control group was not possible because demographic information was not collected, and all surveys were conducted in the first semester.)

Compositing by instructor, for the test group the normalized gains were about the same for the two instructors (59% and 52%). For the control group, by contrast, normalized gains varied widely (−27%, 20%, and 46%).

Student self-reported assessments of knowledge gains are given in Figure S1 of the Supplemental Materials. Students overall reported increases in knowledge of the climate and ice cores after completing the module. They also placed a higher value overall on the importance of polar data in the context of climate change and reported increased comfort with the computational tool after completing the module. Figure S2 of the Supplemental shows that students had an overall favorable ranking of the module, with 64% ranking it "good" and 12% "excellent," and that many (42%) would be interested in learning more about polar research.

### Polar literacy gains

Figures 3 and 4 show student responses to the survey questions testing polar literacy for the pre- and post-survey and for the test and control groups. Tables 4 and 5 give summary statistics. Survey questions in Table 4 are given in full, with abbreviations that are used for brevity in Table 5 and Figures

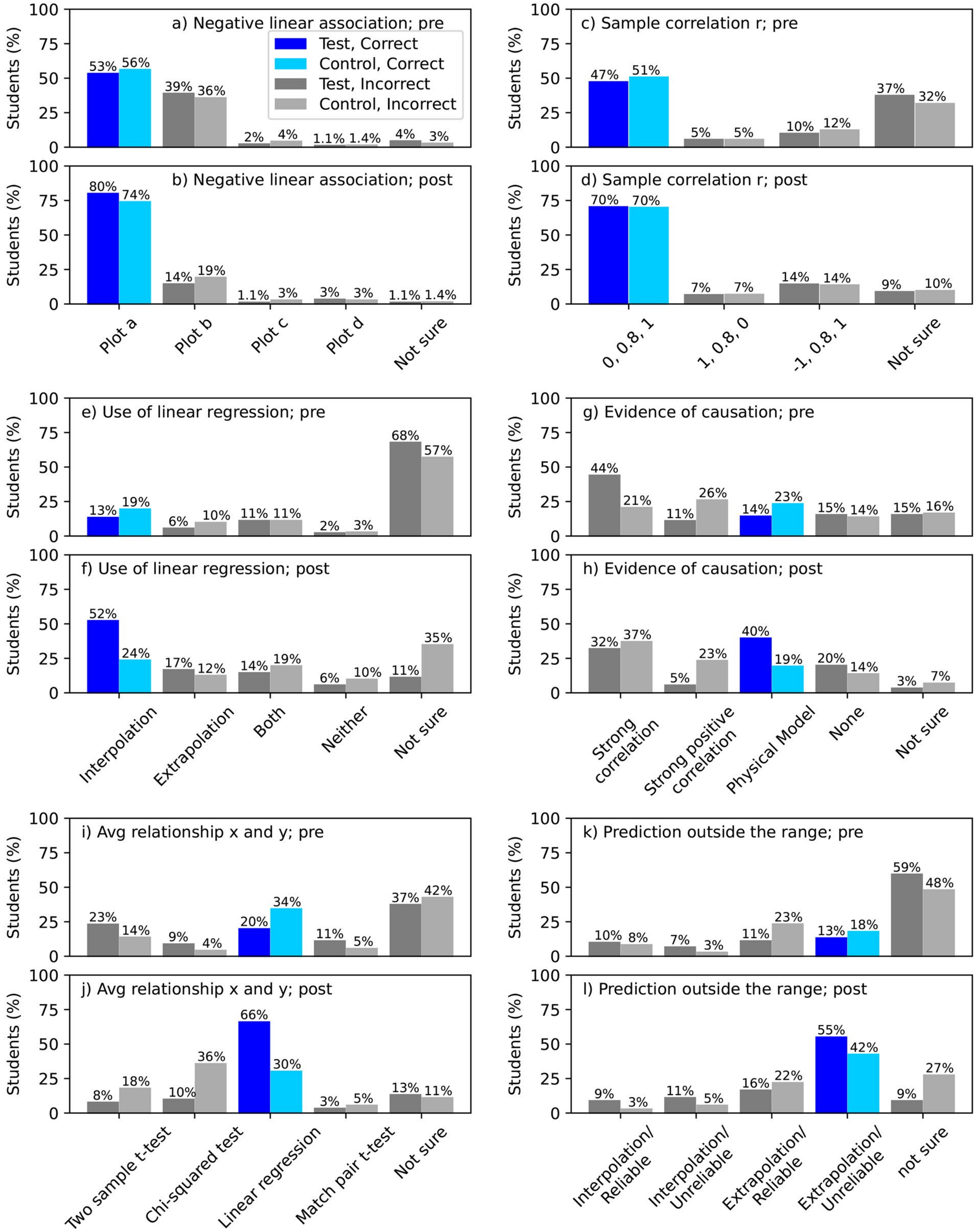


**Figure 2.** Student survey responses for statistics questions for the test (91 students) and control group (73 students). Panel labels give abbreviated questions and survey type (pre or post). The legend given in panel a applies to all panels and indicates that the correct answers are shown in blues while incorrect answers are shown in grays.

3 and 4. As for the statistics questions, polar literacy questions are numbered according to their associated polar literacy goals, with questions 3a and 3b both corresponding to goal 3.

The percentages of students who answered polar literacy questions correctly increased in the post-survey for all questions. Half of these improvements were statistically significant with p-values < .05. Not surprisingly, students in the control group, who were not taught the PENGUIN module, did not demonstrate statistically significant gains in polar literacy, with one exception. There was a statistically significant increase of 17% in the answer to the question of which region is warming the fastest (polar regions).

**Table 2.** Answers to statistics questions before and after completing the polar module. The first column gives the abbreviation and survey question along with an index to the statistics knowledge goal (in parentheses); the next three columns are the percent of students who correctly answered the question before and after completing the module and the difference; n is the total number of responses; and the final two columns give the odds ratio and p-value (statistically significant results to $p < .05$ are in bold).

| Question | Correct (%) | | Change | n | Odds | P |
|---|---|---|---|---|---|---|
| | Before | After | (pts) | | Ratio | |
| Negative linear association: Choose the scatterplot that is best described as having a moderately strong, negative, linear association between the two variables. (1) | 53 | 80 | 27 | 90 | 3.5 | **< .001** |
| Sample correlation r: The sample correlation, r, is a measure of the strength of linear association. Which values of r would represent (in order); no linear association, moderate linear association, and perfect linear association? (2) | 47 | 70 | 23 | 91 | 2.6 | **.003** |
| Avg relationship x and y: Which statistical method would be appropriate for assessing the average relationship between the value of a response variable and the value of a numerical explanatory variable? (3) | 20 | 66 | 46 | 91 | 7.8 | **< .001** |
| Prediction outside range: Suppose we … use the regression equation to predict the value of the response variable when the explanatory variable is well outside the range of the values we observed for it. This is an example of (4a) | 13 | 55 | 42 | 91 | 8.0 | **< .001** |
| Use of linear regression: For an explanatory and response variable showing a strong linear association … which is an appropriate use of a simple linear regression model? (4b) | 13 | 52 | 39 | 90 | 7.1 | **< .001** |
| Evidence of causation: Which of the following give strong evidence of causation? (5) | 14 | 40 | 25 | 91 | 3.9 | **< .001** |

**Table 3.** Answers to statistics questions in control-group statistics classes, where the polar module was not taught. The first column is the abbreviated question and the index to the statistics knowledge goal (in parentheses; see text); the next three columns are the percent of students who correctly answered the question before and after completing the module and the difference; and n is the total number of responses. The odds ratio and p-value are from Fisher's exact test. Full survey questions are given in Table 3.

| Question | Correct (%) | | Change | n | Odds | P |
|---|---|---|---|---|---|---|
| | Before | After | (pts) | | Ratio | |
| Negative linear association (1) | 56 | 74 | 18 | 73 | 2.2 | **.037** |
| Sample Correlation r (2) | 51 | 70 | 19 | 73 | 2.3 | **.027** |
| Avg relationship x and y (3) | 34 | 30 | −4 | 73 | 0.8 | .723 |
| Prediction outside range (4a) | 18 | 42 | 25 | 73 | 3.4 | **.002** |
| Use of linear regression (4b) | 19 | 24 | 4 | 72 | 1.3 | .686 |
| Evidence of causation (5) | 23 | 19 | −4 | 73 | 0.8 | .686 |

Normalized gains were 36% for the test group and 4% for the control group.

For the test group, no overall differences in normalized gains on polar questions were found between instructors, nor were there large differences between STEM and non-STEM majors. However, test-group students who identified as men made higher normalized gains than those who identified as women (41% versus 28%) on polar questions, and there was a decrease in normalized gain by semester (47%, 32%, and 26%).

## Discussion

### *Statistics knowledge gains*

Our results demonstrate significant gains ($p<.05$) for all statistics learning goals for students who worked through the module. Notably, test-group students outperformed students in the control classes by a wide margin. Despite these differences, some trends were the same for the two groups. Both scored better on questions related to goals 1-2 than 4-5 on the pre-survey, and again on the post-survey (with intermediate/mixed results for goal 3). The pre-survey score differences are consistent with expected differences in prior exposure, given that the common core standards for High School (Common Core State Standards Initiative, 2010), explicitly mention form of association and sample correlation but not interpolation and extrapolation or physical models. Recognizing the form and direction of a scatterplot (goal 1) and strength of linear association (goal 2) is also more straightforward than the more subtle understandings of goals 3-5, which require students to interpret linear models and understand their limitations. The stronger gains on these questions for the test group are important indicators of development of more nuanced understanding.

While important gains were made in the test group for goal 4, concerning validity of interpolation and extrapolation, and goal 5, related to knowing that a physical model is needed for evidence of causation, progress is nevertheless needed to raise student post-survey scores, particularly for goal 5, for which only 40% of the test group answered the post-survey question correctly.

It is important to consider to what extent confounding factors could explain test/control group differences. One possibility could be differences in time between teaching the material and giving the post-survey, given that in the latter two semesters, when only test groups were taught, the post-surveys were given soon after the module ended. However, no consistent differences were found between the first semester and the latter two semesters for the instructor who taught the test group in all three semesters (normalized gains of 65%, 47%, and 63%, by semester).

Since different instructors taught the test and control groups, differences in topic emphasis and vocabulary likely resulted in differences across classes. Figure S5 of the Supplemental Materials shows the results after sub-setting by instructor and Tables S2-S6 give results by instructor, for both test and control groups. Standing out in Figure S5 panels i and j is the decrease in correct response rate for the question

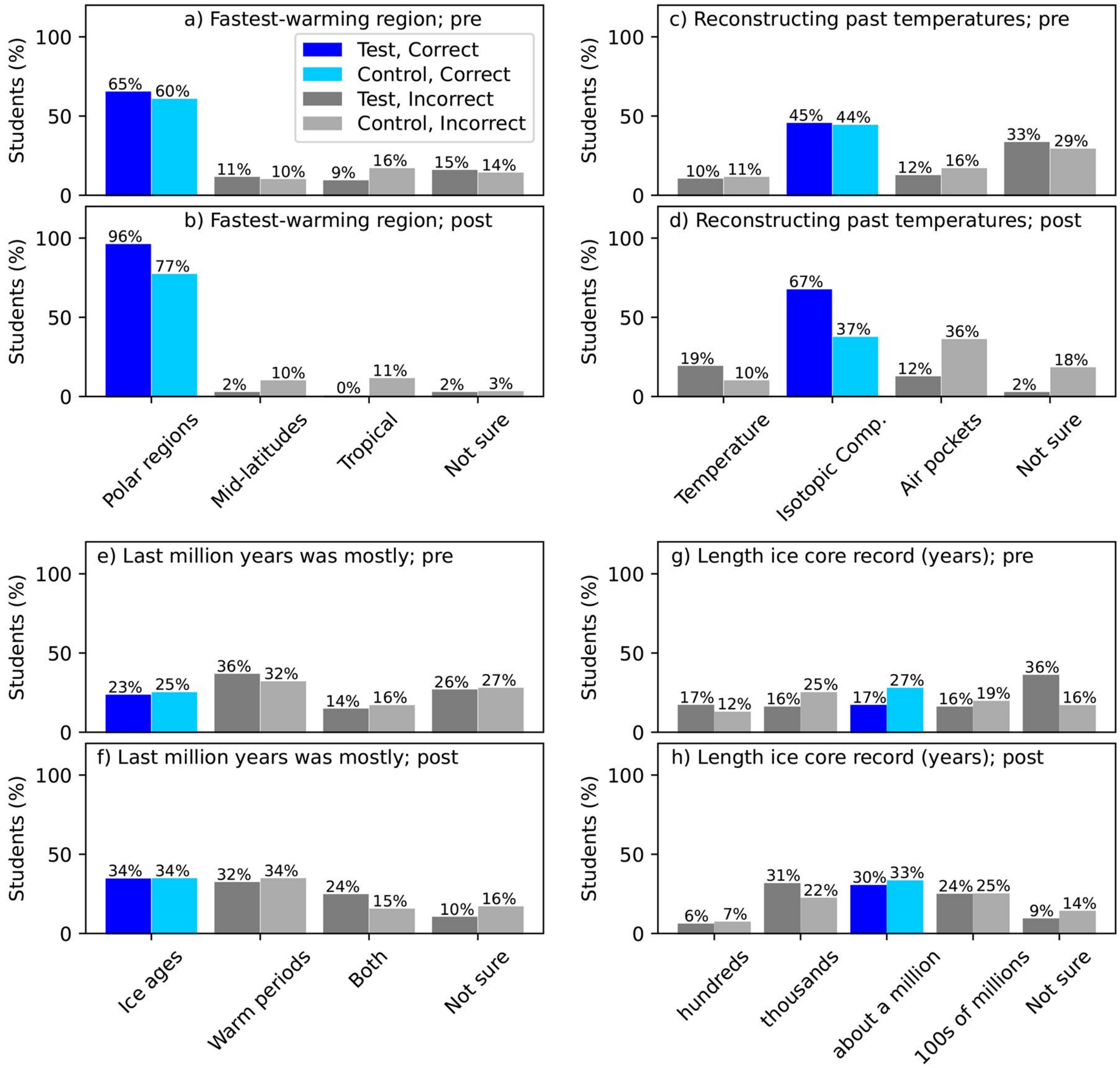


**Figure 3.** Student responses to polar literacy questions on the pre- and post-survey for the test group (91 students) and the control group (73 students). Panel labels give abbreviated survey questions, followed by the survey type (pre or post). The legend given in panel a applies to all panels and indicates that the correct answers are shown in blues while incorrect answers are shown in grays.

regarding the average relationship between x and y (goal 3): in the pre-survey more than two-thirds of students (69%) selected the correct response, dropping to only one student (6%) in the post-survey. This seems likely to be due to the instructor using the alternate vocabulary term, "least-squares regression," which was not a multiple-choice option, instead of the assumed term, "simple linear regression." Overall, the largest gains in the control group were found for instructor 4, who reported teaching all topics on the survey with no significant differences in vocabulary; for this instructor, the normalized gain for statistics questions (46%) was much closer to the gains found for the test group (56%). This was the case despite that this instructor used a mostly lecture-driven approach, with no computational tool apart from a calculator. These findings point to a need for more work to understand the benefits of an inquiry-based approach, and to the need to standardize vocabulary, learning goals, and ideally instructors, across test and control groups.

Introductory statistics is a popular math class for non-STEM majors, who made up 73% of the test group. Comparing STEM majors to non-STEM majors for the test group (see Figure S3 in Supplemental Materials), no significant differences were found, suggesting that the module is equally successful in reaching STEM and non-STEM students.

Regarding gender, only groups of students who identified as men or women were large enough for comparison. For the test group statistics questions, men and women had the same normalized gains (57%), a finding that is contrary to other studies of inquiry-based math activities (Johnson et al., 2018). Furthermore, women made statistically significant gains for all questions, while men did not make statistically significant gains for the question regarding evidence of causation. This suggests that, for statistics, the module was robust against stereotype threat (Steele, 2010).

### *Polar literacy gains*

Table 4 indicates that test-group students made statistically significant gains ($p<.05$) on questions related to polar literacy goals 1, 4, and 6, with mixed results for goal 3. Unsurprisingly,

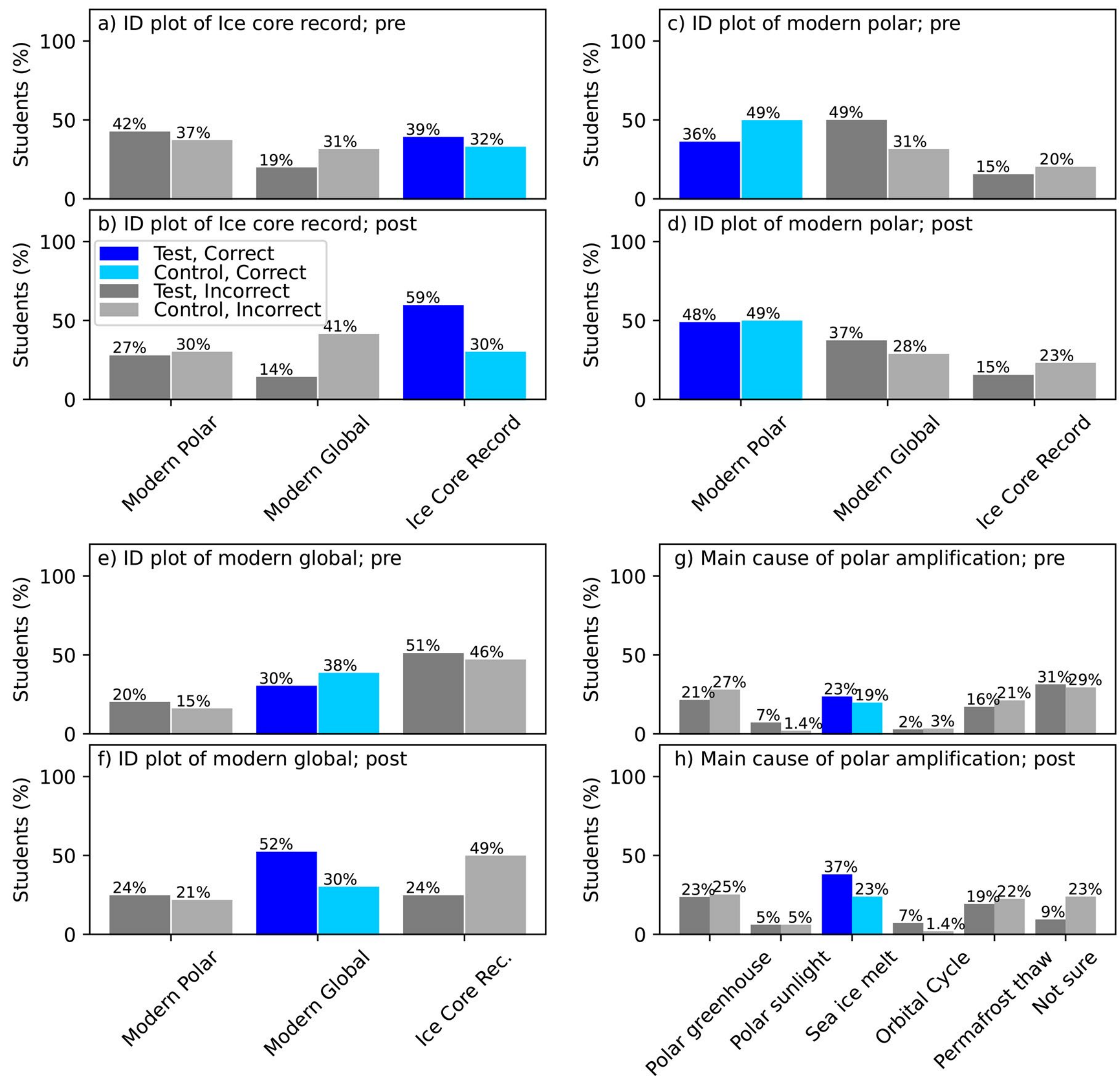


**Figure 4.** Student responses to polar literacy questions on the pre- and post-survey for the test group (91 students) and the control group (73 students). Panel labels give abbreviated survey questions, followed by the survey type (pre or post). The legend given in panel b applies to all panels and indicates that the correct answers are shown in blues while incorrect answers are shown in grays.

**Table 4.** Percentage of correct answers to polar research questions before and after completing an ice core module in a statistics class. The first column gives the abbreviation and survey question along with an index to the polar literacy knowledge goal; the next three columns are the percent of students who correctly answered the question before and after completing the module and the difference; n is the total number of responses; and the odds ratio and p-value are from Fisher's exact test. and the final two columns give the odds ratio and p-value (statistically significant results to $p < .05$ are in bold).

| | Correct (%) | | Change | | Odds | |
|---|---|---|---|---|---|---|
| Question | Before | After | (pts) | n | Ratio | P |
| Fastest-warming region: Over the last 100 years, which region has warmed the fastest due to climate change? (1) | 65 | 96 | 31 | 91 | 12 | **< .001** |
| Polar Amp. Primary Cause: Which is currently believed to be the strongest contributor to polar amplification? (2) | 23 | 37 | 14 | 91 | 2.0 | .052 |
| ID plot modern polar: Temperature and $CO_2$ are positively correlated in the modern era and in the ice core record. However, the temperature change for a change in $CO_2$ is not always the same. Choose the best description for scatterplot b. (3a) | 36 | 48 | 13 | 87 | 1.7 | .124 |
| ID plot modern global: As above, but for scatterplot c. (3b) | 30 | 52 | 22 | 87 | 2.5 | **.010** |
| ID plot ice core record: As above, but for scatterplot a. (4) | 39 | 59 | 20 | 88 | 2.3 | **.010** |
| Last million years mostly: Over the last few million years, Earth's transitions to/from "ice age" conditions were such that most time was spent in… (5) | 23 | 34 | 11 | 91 | 1.7 | .139 |
| Reconstructing past temps: In reconstructing the temperature record of past climates, scientists mainly use the following feature of the ice core. (6) | 45 | 67 | 22 | 91 | 2.5 | **.004** |
| Length ice core record: The longest time range for which climate scientists have obtained useful information from ice cores is on the order of… (7) | 17 | 30 | 13 | 90 | 2.1 | .052 |

**Table 5.** Answers to questions related to polar research in control-group statistics classes, where the polar module was not taught. The first column is the abbreviated question, where full questions are given in Table 1; the next three columns are the percent of students who correctly answered the question before and after completing the module and the difference; n is the total number of responses; and the final two columns give the odds ratio and p-value (statistically significant results to $p < .05$ are in bold).

| | Correct (%) | | Change | | Odds | |
|---|---|---|---|---|---|---|
| Question | Before | After | (pts) | n | Ratio | P |
| Fastest-warming region (1) | 60 | 76 | 17 | 72 | 2.2 | **.049** |
| Polar Amp. primary cause (2) | 19 | 24 | 4 | 72 | 1.3 | .686 |
| ID plot modern polar (3a) | 49 | 49 | 0 | 70 | 1.0 | 1.000 |
| ID plot modern global (3b) | 39 | 29 | −10 | 70 | 0.6 | .283 |
| ID plot ice core record (4) | 33 | 29 | −4 | 70 | 0.8 | .714 |
| Last million years mostly (5) | 25 | 35 | 10 | 72 | 1.6 | .275 |
| Reconstructing past temps (6) | 44 | 38 | −7 | 72 | 0.8 | .498 |
| Length ice core record (7) | 28 | 33 | 6 | 72 | 1.3 | .588 |

since they were not taught the module, students in the control group did not make significant gains on any polar literacy questions (Table 5), with the exception of question 1 regarding polar regions warming fastest; it seems likely they guessed this from hearing about the "polar module". The remainder of this section therefore discusses the test group.

Test-group students overwhelmingly learned that polar regions are among the fastest-warming on Earth (question 1). However, it proved difficult for students to connect this understanding of polar amplification to the idea that a plot of temperature with atmospheric $CO_2$ concentration will have a steeper slope for polar regions than for the modern globe. While they made significant improvements identifying the plot for the modern globe (question 3b), their improvement in identifying the plot for polar regions were not significant to $p < .05$, with only 48% identifying the plot for the modern poles correctly on the post-survey and most of the remainder (37%) incorrectly choosing the modern globe.

The gains found for goal 4 required identifying the ice core plot as the one with lower $CO_2$ values relative to the modern poles and modern globe, corresponding to knowing that $CO_2$ was lower in the pre-modern era.

Students failed to make statistically significant improvements on questions related to polar literacy goals 2, 5, or 7, suggesting revisions are needed to the learning goals or modules, as discussed in the implications below.

Unlike for the statistics learning goals, for polar literacy goals significant differences were found between students who identify as women and men, with normalized gains of 28% for women compared to 41% for men, and with women only making significant gains ($p < .05$) on the first question, compared to gains on three questions for men. (Gains were positive but not significant for women for questions 2-6). Broken down by semester and instructor, women achieved lower normalized gains for both instructors in every semester, with one exception: women's normalized gains were higher than men's in the first semester (51% versus 36%). While many factors could explain this difference, one stands out. In the first semester, a woman polar researcher presented the polar-related parts of the module, while in semesters 2 and 3 the entire module was presented by the instructors, both men without prior polar research expertise. Additional studies are needed to determine to what extent gender differences could be due to 1) having a woman teach the polar material, and 2) having a polar researcher present the polar material. Findings from such studies would hopefully also suggest avenues for improving gender parity. In addition, improvements for students under stereotype threat could be made by coupling PENGUIN modules with meta-cognitive treatments about gender or race in STEM contexts (e.g., Steele, 2010; Liu et al., 2021).

### *Implications for previous PENGUIN modules*

While topics vary among PENGUIN modules, all share a common design and pedagogical strategy, and all use polar data to teach preexisting learning goals in a widely taught (i.e., conventional) disciplinary course using a hands-on computational tool. Therefore, we hypothesize that this work speaks to the efficacy of PENGUIN modules more broadly. Similarities in student self-assessments between the new module and previous modules are consistent with this hypothesis (compare Figures S1 and S2 of the Supplemental Material to Figures 1 and A2 of Rowe et al., 2020). As examples, positive student self-assessments of knowledge gains, comfort with the computational tool, and overall module ranking, as well as increases in the importance placed on polar regions in the context of climate change reported here are echoed in self-assessments of the other PENGUIN modules. These similarities are consistent with documented learning gains in the statistics PENGUIN module applying to PENGUIN modules more broadly.

### *Context for geoscience education*

The learning gains in test versus control assessments we report here add to the body of evidence in support of inquiry-based approaches in general, and more particularly to the efficacy of meeting learning goals using geoscientific questions and data. As mentioned in the introduction, PENGUIN modules are embedded in "conventional" disciplines that exist in most undergraduate institutions, even small ones. The motivation for this lies in the fact that instructors of such disciplines who would like to address pressing geoscientific questions (such as climate change), may nevertheless resist doing so if it means that disciplinary learning goals must be forfeited. The PENGUIN ambition is to provide resources that will allow such instructors to reach their disciplinary learning goals when contextualized through a polar scientific question with comparable (or better) effect and efficiency. That this dual purpose is achievable, we assert, lies in the fact that polar science is intrinsically transdisciplinary. For example, in the PENGUIN module a major emphasis is placed on understanding the limits of interpreting a linear model, because helping students make judgements about when interpolation/extrapolation may be statistically justified is a key learning goal of the course in which the module was taught.

## Limitations

A limitation of this work is that professors who taught the test classes co-developed the module, while the professors who taught the control classes did not. The involvement of control-class professors was limited to administering pre- and post-surveys and describing differences in how they taught their courses. Different professors likely emphasized different topics, and some used different vocabulary than used in the survey questions. This points to a need to align statistics learning goals and topic emphases as well as standardize vocabulary across test and control classes. This could be done by having the same professors teach the course alternately with and without the module.

For statistics questions in the test group, we did not find differences in normalized gains between STEM and non-STEM majors or between students who identify as men and women. Because demographics were not collected for the control group, it could not be determined if the same lack of differences held in statistics classes overall, or if the module mitigated differences.

An additional limitation regards the small sample size of the test and control groups. Finally, there is a possibility of correlation between knowledge gains and likelihood of completing the surveys. Response rates were higher in the first semester of the study (63% versus 55% for the second two semesters) when all surveys were done in class and a financial incentive was given. A large fraction of students in the first semester of the study (32%) completed the pre-survey but either failed to complete the post-survey or had pre- and post-surveys that could not be matched. Putting the identifying information needed to match pre- and post-surveys early in the survey would allow for partial use of incomplete surveys.

## Implications

The new modules are similar in design and implementation to existing PENGUIN modules, and student responses to how well they liked the module and how much they believe they learned are in line with results for those modules (reported by Rowe et al., 2020). Taken together, these results suggest that PENGUIN modules can successfully teach course concepts—potentially even better than for the same course without the module—in addition to giving students exposure to polar research and computational tools. The knowledge gains found suggest that there is no risk to student performance for instructors who try these modules, and significant gains in student performance are likely.

Regarding improving the polar part of the module, men were found to score better on polar literacy questions than women, except for the one group that was taught by a woman polar researcher. More work is needed to investigate this correlation. In addition, the survey results suggest that there were concepts students learned well and others that students did not learn. Given that the course is not a polar science class, a reasonable path forward is to reduce the number of polar concepts for students to learn, and strengthen the module around the remaining ones, while providing clearer linkages to statistics concepts to motivate the learning. We therefore suggest modifications to goals and questions, as follows. Goal 2 should be generalized to, "to be familiar with positive feedback mechanisms that cause amplification" (e.g., the students would select a mechanism but would not need to select the primary mechanism). Correctly answering the questions related to goals 3 and 4 requires both polar literacy and statistical analysis skill, thus, additional effort should be placed on helping students make the connection, for goal 3, that faster increases in temperature mean a steeper slope in a scatterplot of $CO_2$ versus temperature, and for goal 4, that human fossil fuel emissions in the modern era imply that scatterplots will show higher $CO_2$ values for modern times than for the distant past. To avoid confusion regarding time spans, the related survey questions should refer to the "ice core record for the past ~1 million years," rather than just the "ice core record." Finally, we suggest omitting goals 5 and 7 and related questions, regarding warm periods and ice ages over the last million years and the maximum length of the ice core record. Students, particularly non-STEM majors, did not make significant knowledge gains for these goals, and they do not have associated polar literacy principles (McDonnell et al., 2020; https://polar-ice.org/ retrieved 2024/02/14); goal 7 also lacks an associated climate literacy principle; U.S. Global Change Research Program (USGCRP), 2009).

## Conclusions

Survey results for a PENGUIN module (*Statistics: Predicting Temperature*) taught in statistics classes indicate that students made significant knowledge gains in both statistics and in polar literacy after completing the module.

Unsurprisingly, students who completed the polar module did better on survey questions testing polar literacy than a control group in a statistics class that did not complete the module. For polar research questions, correct responses on a student survey increased by 11 to 31% (average 18%), with statistically significant improvements ($p < .05$) of 22-31% on 3 of 6 polar research questions. STEM and non-STEM students in the test group made similar gains in polar literacy, while students who identified as men were found to make larger gains than students who identified as women, except when a woman polar researcher delivered the polar material, when women made the larger normalized gains.

Of note, we also found that students who completed the module scored better than control-group students on questions related to statistics, with statistically significant increases of 25 to 46% (average 34%) for all questions. By contrast, for the control group, changes ranged from −4 to 25% (average 10%), with a statistically significant increase for only three questions. Normalized gains for the statistics questions with the PENGUIN intervention were 56%, whereas the control group experienced a normalized gain of 22%. For statistics questions, differences in normalized gains were not found between STEM and non-STEM majors or men and women.

Our results suggest that through PENGUIN modules, polar research can successfully be taught in a range of courses without sacrificing course learning goals, but rather with the potential to enhance them.

## Acknowledgements

We are grateful to Dr. Spruce Schoenemann for sharing his module "From Isotopes to Temperature: Using Ice Core Data!", available at https://uwpcc.ocean.washington.edu/file/Oxygen_Isotopes_in_Ice_Cores, which inspired this module; for help understanding how ice cores are used to infer the temperature record; and for sharing images used in the module presentations. We are also grateful to the Social and Economic Sciences Research Center of Washington State for helping develop surveys, deploying them, and for conducting external evaluations, with particular thanks to Candiya Mann and Adam McKee. John Woodward provided valuable help with planning the design of the module. We thank the Science Education and Resource Center (SERC) for hosting our PENGUIN modules, and the instructors for disseminating the control surveys to their statistics students.

## Disclosure statement

No potential conflicts of interest were reported by the authors.

## Funding

This work was supported by the National Science Foundation under Grant 2021429 and 2021213.

## ORCID

Penny M. Rowe http://orcid.org/0000-0002-7594-0553

## References

Apedoe, X. S., Walker, S. E., & Reeves, T. C. (2006). Integrating inquiry-based learning into undergraduate geology. *Journal of Geoscience Education*, *54*(3), 414–421. https://doi.org/10.5408/1089-9995-54.3.414

Bedford, D. (2016). Does climate literacy matter? A case study of US students' level of concern about anthropogenic global warming. *Journal of Geography*, *115*(5), 187–197. https://doi.org/10.1080/00221341.2015.1105851

Carey, C. C., & Gougis, R. D. (2017). Simulation modeling of lakes in undergraduate and graduate classrooms increases comprehension of climate change concepts and experience with computational tools. *Journal of Science Education and Technology*, *26*(1), 1–11. https://doi.org/10.1007/s10956-016-9644-2

Carey, C. C., Farrell, K. J., Hounshell, A. G., & O'Connell, K. (2020). Macrosystems EDDIE teaching modules significantly increase ecology students' proficiency and confidence working with ecosystem models and use of systems thinking. *Ecology and Evolution*, *10*(22), 12515–12527. https://doi.org/10.1002/ece3.6757

Caspari, A., Kuhlthau, C. C., & Maniotes, L. (2007). *Guided inquiry: Learning in the 21st century*. Libraries Unlimited.

Coletta, V. P., & Steinert, J. J. (2020). Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories. *Physical Review Physics Education Research*, *16*(1), 010108. https://doi.org/10.1103/PhysRevPhysEducRes.16.010108

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. National Governors Association Center for Best Practices and the Council of Chief State School Officers. Washington, DC. Retrieved from https://learning.ccsso.org/wp-content/uploads/2022/11/Math_Standards1.pdf

Dahe, Q., Petit, J. R., Jouzel, J., & Stievenard, M. (1994). Distribution of stable isotopes in surface snow along the route of the 1990 International Trans-Antarctica Expedition. *Journal of Glaciology*, *40*(134), 107–118. https://doi.org/10.3189/S0022143000003865

Fortmann, L., Beaudoin, J., Rajbhandari, I., Wright, A., Neshyba, S., & Rowe, P. (2020). Teaching modules for estimating climate change impacts in economics courses using computational guided inquiry. *The Journal of Economic Education*, *51*(2), 143–158. https://doi.org/10.1080/00220485.2020.1731383

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gould, R., Sunbury, S., & Dussault, M. (2014). In praise of messy data. *The Science Teacher*, *081*(08), 31. https://doi.org/10.2505/4/tst14_081_08_31

Grissom, A. N., Czajka, C. D., & McConnell, D. A. (2015). Revisions of physical geology laboratory courses to increase the level of inquiry: Implications for teaching and learning. *Journal of Geoscience Education*, *63*(4), 285–296. https://doi.org/10.5408/14-050.1

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74.

Ham, N. R., & Flood, T. P. (2009). International field trips in undergraduate geology curriculum: Philosophy and perspectives. *Geological Society of America Special Papers*, *461*, 99–104.

Hammett, A., & Dorsey, C. (2020). Messy data, real science. *The Science Teacher*, *87*(8), 40–48. https://doi.org/10.1080/00368555.2020.12293530

Ialongo, C. (2016). Understanding the effect size and its measures. *Biochemia Medica*, *26*(2), 150–163. https://doi.org/10.11613/BM.2016.015

Jackson, J., Dukerich, L., & Hestenes, D. (2008). Modeling instruction: An effective model for science education. *Science Educator*, *17*(1), 10–17.

Johnson, E., Andrews-Larson, C., Keene, K., Keller, R., Fortune, N., & Melhuish, K. (2018). Inquiry and inequity in the undergraduate mathematics classroom. In *Proceedings of the 40th Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education*.

Kastens, K., Krumhansl, R., & Baker, I. (2015). Thinking big. *The Science Teacher*, *082*(05), 25. https://doi.org/10.2505/4/tst15_082_05_25

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE Life Sciences Education*, *18*(2), es2. https://doi.org/10.1187/cbe.18-02-0023

Klug, J. L., Carey, C. C., Richardson, D. C., & Darner Gougis, R. (2017). Analysis of high-frequency and long-term data in undergraduate ecology classes improves quantitative literacy. *Ecosphere*, *8*(3), e01733. https://doi.org/10.1002/ecs2.1733

Klyce, A., & Ryker, K. (2023). What does a degree in geology actually mean? A systematic evaluation of courses required to earn a bachelor of science in geology in the United States. *Journal of Geoscience Education*, *71*(1), 3–19. https://doi.org/10.1080/10899995.2022.2076201

Lewis, S. E., & Lewis, J. E. (2008). Seeking effectiveness and equity in a large college chemistry course: An HLM investigation of peer-led guided inquiry. *Journal of Research in Science Teaching*, *45*(7), 794–811. https://doi.org/10.1002/tea.20254

Liu, S., Liu, P., Wang, M., & Zhang, B. (2021). Effectiveness of stereotype threat interventions: A meta-analytic review. *The Journal of Applied Psychology*, *106*(6), 921–949. https://doi.org/10.1037/apl0000770

Madison, B. L. (2014). How does one design or evaluate a course in quantitative reasoning? *Numeracy*, *7*(2), 3. https://doi.org/10.5038/1936-4660.7.2.3

Martin-Hansen, L. (2002). Defining inquiry: Exploring the many types of inquiry in the science classroom. *The Science Teacher*, *69*(2), 34–37.

McDonnell, J., Hotaling, L., Schofield, O., & Kohut, J. (2020). Key concepts in polar science: Coming to consensus on the essential polar literacy principles. *Current: The Journal of Marine Education*, *34*(1), 2–8. https://doi.org/10.5334/cjme.42

Mosher, S. and Keane, C. (Eds.). (2021). *Vision and change in the geosciences: The future of undergraduate geoscience education*. American Geosciences Institute.

National Academies of Sciences, Engineering, and Medicine (2020). *Strengthening sustainability programs and curricula at the undergraduate and graduate levels*. National Academies Press.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., & Laaksonen, A. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, *3*(1), 168. https://doi.org/10.1038/s43247-022-00498-3

Rosenberg, J. M., Schultheis, E. H., Kjelvik, M. K., Reedy, A., & Sultana, O. (2022). Big data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Educational Technology*, *53*(5), 1179–1201. https://doi.org/10.1111/bjet.13245

Rowe, P. M., Cheng, H., Fortmann, L., Wright, A., & Neshyba, S. (2018). Teaching image processing in an upper level CS undergraduate class using computational guided inquiry and polar data. *Journal of Computing Sciences in Colleges*, *34*(1), 171–179.

Rowe, P. M., Fortmann, L., Guasco, T. L., Wright, A., Ryken, A., Sevier, E., Stokes, G., Mifflin, A., Wade, R., Cheng, H., Pfalzgraff, W., Beaudoin, J., Rajbhandari, I., Fox-Dobbs, K., & Neshyba, S. (2020). Integrating polar research into undergraduate curricula using computational guided inquiry. *Journal of Geoscience Education*, *69*(2), 178–191. https://doi.org/10.1080/10899995.2020.1768004

Steele, C. M. (2010). *Whistling Vivaldi: How stereotypes affect us and what we can do*. WW Norton & Company.

Soule, D., Darner, R., O'Reilly, C. M., Bader, N. E., Meixner, T., Gibson, C. A., & McDuff, R. E. (2018). EDDIE modules are effective learning tools for developing quantitative literacy and seismological understanding. *Journal of Geoscience Education*, *66*(2), 97–108. https://doi.org/10.1080/10899995.2018.1411708

U.S. Global Change Research Program (USGCRP). (2009). Climate literacy: The essential principles of climate science.

Weaver, G. C., Russell, C. B., & Wink, D. J. (2008). Inquiry-based and research-based laboratory pedagogies in undergraduate science. *Nature Chemical Biology*, *4*(10), 577–580. https://doi.org/10.1038/nchembio1008-577

Wells, M., Hestenes, D., & Swackhamer, G. (1995). A modeling method for high school physics instruction. *American Journal of Physics*, *63*(7), 606–619. https://doi.org/10.1119/1.17849

Wiggins, G., & McTighe, J. (2018). *Understanding by design* (pp. 15–18). Association for Supervision and Curriculum Development (ASCD).

Windschitl, M. (2008). What is inquiry? A framework for thinking about authentic scientific practice in the classroom. In J. Luft, R. L. Bell & J. Gess-Newsome (Eds.), *Science as inquiry in the secondary setting*, (pp. 1–20). National Science Teachers Association Press.