RESEARCH   Open Access

Web-based authentic inquiry experiences in large introductory classes consistently associated with significant learning gains for all students

X. Ben Wu[1*] , Carolyn Sandoval[2], Stephanie Knight[3], Xavier Jaime[1], Maria Macik[1] and Jane F. Schielack[4]

Abstract

Background: Continuous calls for reform in science education emphasize the need to provide science experiences in lower-division courses to improve the retention of STEM majors and to develop science literacy and STEM skills for all students. Open or authentic inquiry and undergraduate research are effective science experiences leading to multiple gains in student learning and development. Most inquiry-based learning activities, however, are implemented in laboratory classes and the majority of them are guided inquiries. Although course-based undergraduate research experiences have significantly expanded the reach of the traditional apprentice approach, it is still challenging to provide research experiences to nonmajors and in large introductory courses. We examined student learning through a web-based authentic inquiry project implemented in a high-enrollment introductory ecology course for over a decade.

Results: Results from 10 years of student self-assessment of learning showed that the authentic inquiry experiences were consistently associated with significant gains in self-perception of interest and understanding and skills of the scientific process for all students—both majors and nonmajors, both lower- and upper-division students, both women and men, and both URM and non-URM students. Student performance in evaluating the quality of an inquiry report, before and after the inquiry project, also showed significant learning gains for all students. The authentic inquiry experiences proved highly effective for lower-division students, nonmajors, and women and URM students, whose learning gains were similar to or greater than those of their counterparts. The authentic inquiry experiences were particularly helpful to students who were less prepared with regard to the ability to evaluate a scientific report and narrowed the performance gap.

Conclusions: These findings suggest that authentic inquiry experiences can serve as an effective approach for engaging students in high-enrollment, introductory science courses. They can facilitate development of science literacy and STEM skills of all students, skills that are critical to students' personal and professional success and to informed engagement in civic life.

Keywords: Inquiry experiences, Ecology, Large introductory classes, Majors and nonmajors, Science literacy, STEM skills, Underrepresented minority, Women

* Correspondence: xbw@tamu.edu
[1]Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX 77843, USA
Full list of author information is available at the end of the article

Introduction

Continuous calls for reform in science education emphasize the need to introduce undergraduate students to the scientific process early in their careers and improve students' ability to apply the process of science (AAAS 2011). The "Engage to Excel" report by the President's Council of Advisors on Science and Technology also stresses the importance of improving the first two years of undergraduate STEM (Science, Technology, Engineering, and Math) education by providing STEM experiences (PCAST 2012). In the report *Science Literacy: Concepts, Contexts, and Consequences*, the National Academies of Science, Engineering, and Medicine (NASEM) articulated "understanding of scientific practices" and "understanding of science as a social process" as two of the three key aspects of science literacy, along with content knowledge (NASEM 2016). The report also argued the necessity for developing science literacy for all citizens, based on four primary rationales: personal, economic, democratic, and cultural (NASEM 2016).

It is common in large institutions to have high-enrollment, lecture-based introductory science courses for nonmajors, or similar courses for both majors and nonmajors with separate lab courses that are smaller and primarily accessible to the majors. How can we provide authentic learning experiences to help all students in these courses develop the understanding of the process of science and STEM skills? This may be particularly important for nonmajors for many of whom the introductory science courses may be their only exposure to science during their college career.

Undergraduate research experiences (URE), both traditional URE and course-based undergraduate research experiences (CURE) (Harrison et al. 2011), and open inquiries (Buck et al. 2008; Rowland et al. 2016) are effective learning experiences that can help undergraduate students develop understanding of the scientific process and STEM skills (Handelsman et al. 2007; Lopatto 2009; O'Donnell et al. 2015). Providing these authentic inquiry experiences in high-enrollment, lecture-based introductory science courses, however, can be logistically difficult or impractical. Web-based approaches have the potential to make authentic inquiry experiences in such settings feasible and effective. Given the dramatic and potentially lasting pedagogical changes in undergraduate education triggered by the current pandemic, there is a heightened need to implement web-based authentic inquiry experiences and assess their efficacy in developing critical science literacy and skills.

Literature review

Inquiry-based learning has a long history in K-12 science education (Buchanan et al. 2016; Khalaf 2018; Schwab 1962) and there has been a growing interest in undergraduate science education over the recent decades (AAAS 2011; Beck et al. 2017; Handelsman et al. 2007; Koretsky et al. 2018; Sundberg et al. 2005). In inquiry-based learning, students engage in many activities and thinking processes as scientists do (NRC 2000). These activities may include making observations, framing questions and hypotheses, designing and conducting scientific investigations, formulating scientific explanations and models based on evidence and logic, communicating results, and revising the explanations or revisiting the investigations based on feedback and critique from peers (NRC 2000; Weaver et al. 2008). Practices of inquiry-based learning can be grouped into four types, or an inquiry continuum, based on the degree of independence of students in the inquiry process (Bell et al. 2005; Fay et al. 2007; Weaver et al. 2008; Wheeler and Bell 2012). The traditional confirmation labs are at one extreme of the continuum where instructors provide the research question and procedure and ask students to confirm a known outcome. In structured inquiry, the research question and procedure are provided, but students are asked to explore the unknown outcome through the inquiry. In guided inquiry, only the research question is provided, and students are asked to design and conduct the investigation to answer the question. In open inquiry, also referred to as authentic inquiry (Bielik and Yarden 2016; Buck et al. 2008; Rowland et al. 2016), students are given only a raw phenomenon and asked to formulate the research question and design and conduct the investigation to answer the question.

Numerous studies have shown that inquiry-based learning leads to significant learning gains in performance, attitudes, and inquiry-related skills (Gehring and Eastman 2007; Goldey et al. 2012; Howard & Miskowski 2005; Rissing and Cogan 2009; Russell and Weaver 2010; Webb et al. 2014). Although most of the studies focused on courses for science majors, nonmajors in inquiry-based introductory biology courses also showed greater improvements in science literacy and research skills (Gormally et al. 2009) and greater positive changes in attitudes toward science than those in a traditional course (Kiernan and Lotter 2019). In a study of the effects of inquiry-based ecology laboratory courses at a historically Black college and a private research university, students in both institutions had significant learning gains in confidence and scientific reasoning skills, with no institution effect, indicating positive impact of inquiry-based learning on the learning of both student populations (Beck and Blumer 2012). A longitudinal study also showed long-term improvements in learning through learner-centered and inquiry-based courses taken by biology majors early in a curriculum (Derting and Ebert-May 2010). A review of published studies on inquiry-based teaching in undergraduate biology laboratory courses showed that the

majority of them used guided rather than open inquiries (Beck et al. 2017). Open inquiries, however, are likely more effective for developing understanding of scientific practices and science as a social process. Most open inquiries are implemented in laboratory classes (Beck et al. 2017; Sundberg et al. 2005) with few in lecture-based introductory courses that typically serve a large number of nonmajor students for whom such a course may be their only exposure to science in their college career.

Inquiry-based learning and undergraduate research experiences exist along a continuum (Beck et al. 2017; D'Avanzo 1996; Weaver et al. 2008), where intellectual autonomy and responsibility of the students increase from confirmation labs to apprenticeship in a research lab (Weaver et al. 2008). A defining feature of undergraduate research experiences that separates it from inquiry-based learning is discovery of new knowledge (Auchincloss et al. 2014; Weaver et al. 2008). Undergraduate research has been well-documented as a high-impact learning experience that contributes positively to students' understanding of the scientific process (Kuh et al. 2010; Lopatto 2009; O'Donnell et al. 2015). In their 2003–2005 web-based survey study of students who participated in undergraduate research, Russell et al. (2007) reported that 88% of their respondents claimed that their understanding of how to conduct research increased a fair amount or a great deal. A 3-year study of the benefits to students of participating in undergraduate research by Seymour et al. (2004) showed multiple gains, including "thinking and working like a scientist" (p. 511). Using the SURE (Survey of Undergraduate Research Experiences) to examine the benefits of undergraduate research, Lopatto (2004, 2007, 2009) surveyed undergraduate students from multiple institutions. Results from his cross-institutional, multiyear study show student gains not only in understanding the research process but also in planning for postgraduate education, and in personal development, e.g., working independently.

Participation in undergraduate research also correlates to retention and persistence for all students, including underrepresented minority students (Bauer and Bennett 2003; Campbell and Skoog 2008; Finley and McNair 2013; Nagda et al. 1998). Using longitudinal data from a large research university, Jones et al. (2010), in particular, highlighted the importance of underrepresented minority student participation in undergraduate research. Their study showed that students who participated in research during their undergraduate career were more likely to complete a baccalaureate degree, persist in biology, and perform well in biology, which could impact interest in admission to graduate school. Chang et al. (2014) conducted a longitudinal study analyzing data from the Cooperative Institutional Research Program's 2004 Freshman Survey and 2008 College Senior Survey,

which included a sample of 3670 students at 217 institutions. Key findings from their study revealed that by increasing student involvement in key academic experiences, including undergraduate research, higher education institutions can play a significant role in reducing racial disparities in STEM, which appear to be largely a result of unequal preparation and access to academic opportunities.

Much of the literature related to undergraduate research focuses on UREs (undergraduate research experiences), which is an apprenticeship or internship model that typically involves a few selected students who invest time outside of class working with a faculty mentor on a research project. The challenge of this model is that it is difficult to scale up and make it available to all STEM students during their first 2 years of college studies (Desai et al. 2008; Harrison et al. 2011; Sadler et al. 2010; Wei & Woodin 2011). Researchers have also noted that the focus of most studies of undergraduate research is on upper-level students who are a self-selected population already interested in research.

Recently, there has been a greater emphasis on CUREs (course-based undergraduate research experiences) in which whole classes are involved in scientific research (Harrison et al. 2011). The aim of these experiences is to reach students who might not have identified research as a career or education at an earlier stage in their academic careers by integrating research into lower-level undergraduate courses (Harrison et al. 2011). Focusing on course-based research at the introductory level, Harrison et al. (2011) discussed their experiences using the Phage Genomics course as the model. Evidence from their study shows that first-year students doing research learn the process of science, as well as how scientists practice science. Their study also showed that students' career interests, such as pursuing graduate education, changed. Shaffer et al. (2014) share the results of their multi-campus research based on a bioinformatics project (a course-based research experience) that was incorporated into the curriculum of diverse institutions. They found that the bioinformatics project within a biology curriculum provides a mechanism for successfully engaging large numbers of students in undergraduate research; benefits to students are achievable at a wide variety of academic institutions; and successful implementation of course-based research experiences requires significant investment of instructional time for students to gain full benefit. Auchincloss et al. (2014) provide a framework that characterizes CUREs and outlines research possibilities and recommendations for educators who include research experiences in their courses. Although one of the benefits of CUREs is that all students enrolled in a course are involved in research, typically CUREs have focused on science majors (Bakshi et al. 2016; Harrison et al. 2011). Additionally, along with requiring a significant investment in time

(Shaffer et al. 2014), other barriers to offering CUREs exist, such as class size and a lack of resources to assist faculty with developing and effectively teaching CUREs (Bakshi et al., 2016). Whereas the examples above highlight continued efforts to engage college students in scientific inquiry and research, these opportunities, as well as research on the benefits of participation in experiences such as CUREs that are open to all students, remain focused on science majors (Ballen et al. 2017).

Current study and research questions

It is important to provide authentic inquiry experiences in high-enrollment, lecture-based introductory science courses because such experiences can help students develop science literacy as defined in the NASEM report (2016), critical thinking and problem-solving skills (NSTC 2018), interpret and make sense of science in their daily lives, and engage with science in meaningful ways with issues that have personal and community relevance (Feinstein et al. 2013). It is challenging to provide authentic inquiry experiences in these courses, however, because implementing open inquiry experiences or course-based undergraduate research experiences in these settings is logistically difficult or impractical.

A web-based authentic inquiry project, with a focus on understanding the process of science and developing skills rather than content knowledge, was developed for a high-enrollment introductory ecology course (without a lab component) and successfully implemented for over a decade in a Research 1 University (Wu et al. 2016). Although related to the work of a research lab, the authentic inquiry experiences focused on science as a process and not a product (Hester et al. 2018; Spell et al. 2014) and modeled the processes used by practicing scientists (Peffer et al. 2015; Peffer and Ramezani 2019), but were not part of the research that scientists actually carried out (Chinn and Malhotra 2002). These authentic inquiry experiences focused on developing science literacy and STEM skills (NASEM 2016), to help students learn ideas and skills that are of value and use to their respective futures (Rowland et al. 2016).

The current study attempted to explore the effect of the authentic inquiry project on student learning through analyses of the data from assessments built into the course, including data on student self-perception of learning from 2007 to 2016 and student performance in evaluating a scientific report and quality of their reports from 2013 to 2016. We asked the following research questions: (1) What were the self-perceptions of learning gains of students in their interest, ability, and understanding of the scientific process through the authentic inquiry experiences, for all and different groups of students—majors and nonmajors, lower- and upper-division students, women and men, and underrepresented minorities (URM) and other ethnicities (non-URM)? (2) Did the self-perceptions of learning gains differ between each pair of the groups? (3) What was the impact of the experiences on students' performance in evaluating the quality of scientific reports for all and different groups of students? (4) Did the changes in performance differ between these groups?

Methods
Design of the inquiry project

A web-based ecological inquiry project using archived BearCam photos (see below for a description of the BearCam photos) was developed in 2006 to provide an authentic inquiry experience for students in a large introductory ecology course (Wu et al. 2016), through the activities of the NSF-funded Information Technology in Science (ITS) Center for Teaching and Learning at Texas A&M University (Schielack and Knight 2012). The learning goals of the inquiry project were to help students develop a better understanding of the nature and process of science and enhance their critical thinking and communication skills.

The activities of the inquiry project were structured over a 5-week period. Students conducted web-based individual research projects outside of class with ongoing peer feedback through online group discussions as well as instructor-facilitated discussions in class. Following is an outline of the activities through the 5 weeks and more detailed descriptions and discussions are available in Wu et al. (2016).

Week 1: form peer group and conduct background study Students (a) participate in an online ice-breaker activity to get to know each other in their inquiry project discussion group (of ~ 10 students) in Blackboard Learn, (b) conduct an online search of background information on the biology and behavior of grizzly bears, and (c) share one interesting piece of information on bear biology or behavior, which has not been mentioned by other group members. We also dedicate the first class period of week 1 to explore with students the rationale and relevance of the project, and to discuss ecological background and the process of the inquiry project. Finally, we discuss the rubric, with 30 items organized in 10 categories (Fig. 1), which is used throughout the project to guide the inquiry process and the writing and peer review of the inquiry report.

Week 2: develop hypothesis and design investigation Students (a) study the BearCam photos and observe patterns in bear behavior and spatial distribution, (b) formulate a hypothesis in terms of specific predictions of the pattern, (c) design the procedure for collecting relevant data, with sufficient sample size and appropriate sampling regimen for testing the hypothesis, and (d)

Rubric for Reviewing Inquiry Report

Each item needs to be answered with a "1" (positive) or "0" (negative).

Objective (If not present, all 3 scores=0)
1. Is the objective clear?
2. Is the objective reasonably specific?
3. Does the objective explain the purpose of the study?

Hypothesis (If not present, all 3 scores=0)
4. Is the hypothesis logical?
5. Is the hypothesis testable?
6. Is the hypothesis nontrivial?

Sampling (If not present, all 3 scores=0)
7. Is the number of samples reported?
8. Is the number of samples sufficient?
9. Is there sufficient description for sample selection?

Data Collection (If not present, all 3 scores=0)
10. Are the variables collected appropriate for testing the hypothesis?
11. Is there sufficient description for data collection (variables and how collected)?
12. Is there sufficient description for data analysis (i.e., frequency, count, average, etc.)?

Data Display (If not present, all 3 scores=0)
13. Are data displayed in an appropriate form?
14. Does the data display represent appropriate variables?
15. Does the data display address the hypothesis?

Results (If not present, all 3 scores=0)
16. Is there a presentation in the text of the results- without interpretations?
17. Are the results specific?
18. Do the results correspond to the variables specified in the methods section?

Conclusions (If not present, all 3 scores=0)
19. Are the conclusions based solely on the results?
20. Are the conclusions sufficiently developed based on the results?
21. Do the conclusions correspond to the hypothesis?

Discussion (If not present, all 3 scores=0)
22. Does the discussion include interpretation(s) of the results (possible mechanisms or explanations) that go beyond merely restating them?
23. Does the discussion discuss the limitations of the study?
24. Does the discussion suggest future studies/new questions?

Organization
25. Is the Introduction section free of content belonging to other sections?
26. Is the Methods section free of content belonging to other sections?
27. Is the Results & Discussion section free of content belonging to other sections?

Report
28. Is the report written in grammatically correct sentences?
29. Is the report written in a succinct, technical style?
30. Is the report free of unnecessary repetition?

Fig. 1 Rubric for evaluating inquiry report (from Wu et al. 2016)

share the hypothesis and design with the discussion group and provide feedback to the hypothesis and design of investigation of at least two other group members. A portion of one class period of week 2 was used to discuss scientific hypotheses, bias, random and stratified sampling, and effect of sample size, through mini-lectures and active learning activities. The archive photos used in the inquiry project were collected by Lawrence Griffing and his students who archived the pictures and video from a remote-controlled (at Texas A&M University) pair of video cameras at McNeil Falls in the McNeil River State Game Sanctuary for animal behavioral studies. Over 1200 archived photos organized by hours of a day were made available for the inquiry project, as well as an aerial photo of the McNeil River Fall and a scaled map with 1 m × 1 m grid to identify the location of the bears and estimate distances between bears (http://bearcaminquiryproject.weebly.com/).

Week 3: collect data, analyze data, and develop inquiry report Students (a) select a set of random/stratified random samples (photos with time and ID) appropriate for testing the hypothesis, (b) collect data from each photo selected and record the data in an Excel file, (c) conduct data analysis and generate figure(s) or table(s) to represent the results, (d) write an inquiry report following the guidelines and rubric provided and submit completed inquiry report to the Calibrated Peer Review (CPR) system, and (e) make at least one posting to the discussion group to discuss one's data collection, analysis, and interpretations and respond to at least two postings of other group members and provide feedback. CPR (cpr.molsci.ucla.edu/Home.aspx) is used as an online instructional tool to help students understand the process and value of peer review in scientific inquiries, enhance their skills in critically evaluating scientific writing, and provide and receive feedback on their inquiry reports.

Week 4: conduct calibrated peer review Students (a) complete three calibrations, (b) review three reports of peers, and (c) conduct self-assessment of one's own report. Reviews are completed in the CPR system using the rubric. The three "calibration" reports, of high, medium, and low quality, were set up and evaluated using the rubric by the instructor. Students can make up to two attempts to evaluate each calibration report and receive feedback to learn how to evaluate reports using the rubric.

Week 5: revise inquiry report based on peer review feedback Students revise the inquiry report based on the peer review feedback and self-assessment and submit a revised report online.

Implementation and assessment
The inquiry project using BearCam was implemented in a high-enrollment introductory ecology course, Fundamentals of Ecology, at Texas A&M University in the Fall semesters of 2007–2016. This course had two sections with 400–500 students combined, from 30 to 50 different majors for some of whom this course was required and for others an elective. It was a sophomore level course; however, the typical distribution of classifications had approximately 65% lower-division (freshman and sophomore) students and 35% upper-division (junior and senior) students. In the Fall semesters of 2007–2011, all students were required to conduct an inquiry project using BearCam; in Fall semesters of 2012–2016, about 60–80% of the students conducted an inquiry project using BearCam and the others conducted a different inquiry project with similar guidelines.

Throughout the 5-week period for the inquiry project, we typically used the first 5 to 15 min at the beginning of each class (MWF) to discuss the inquiry project, which provided opportunities for formative assessment of the project and of student learning. We asked and answered student questions, had targeted exercises and discussions coupled with clicker questions, and gave mini-lectures as needed. Whenever possible, we tried to relate what students were doing in their inquiry project to the experiences and examples of what we do in our own research. This seemed attractive and engaging to students, possibly because they saw that they were engaging in an authentic inquiry process similar to what researchers did, as well as learning from our experiences.

Self-perception of learning gains
An online survey was conducted at the end of the inquiry project each semester and students were asked to provide feedback on the inquiry project and reflect on their learning through the project (see Supplementary Information). Most of the questions remained the same over the years. With a subset of eight questions in the survey, we asked students to rate their interest in ecology, ability to formulate testable hypotheses, understanding how ecologists conduct research, and ability to evaluate the quality of scientific reports before and after the inquiry project. Student responses to these questions from Fall semesters of 2007–2016, in a five-level scale, rated from 1 (*very low*) to 5 (*very high*), were used to assess self-perception of learning through the inquiry project.

These four pairs of questions were part of the post-project online survey that was designed for formative assessment for the inquiry project, created by the group of faculty members (in ecology, educational psychology, and math education) who designed and implemented this inquiry project. These four pairs of questions were created based on considerations of (1) the focus of the inquiry project on developing understanding and skills related to the scientific process and (2) the anticipated potential learning gains given the specific design of the inquiry project.

For the questions on interest in ecology before and after the inquiry project, we did not specifically define what interest is (Renninger and Hidi 2011; Rowland et al. 2019). There were likely substantial variations among students in what interest in ecology means to them. We were interested in, and interpreted the results as, their perceptions of the change through the inquiry project in their various interpretations of interest in ecology collectively, without being able to interpret the specific nature of the gains. Similarly, we did not provide specific definitions in the three pairs of questions on rating their ability to formulate testable hypotheses,

understanding how ecologists conduct research, and ability to evaluate the quality of a scientific report before and after the inquiry project, respectively. We did, however, discuss in class the definitions of and difference between a scientific hypothesis and a statistical or research hypothesis and, given the limitation of the BearCam data, asked students to focus on a research hypothesis and explicitly discuss the potential mechanisms or explanations in the inquiry report. During the discussions on the inquiry project at the beginning of each class, we deliberately drew parallels between what students were doing at the time and what we as researchers do in our ongoing or past ecological research projects, the reasons for the practices, the insights we gained, and lessons learned. It was common to see comments in student feedback that they appreciated the sense of relevance they gained through these discussions. This mapping to the real world of ecological research could contribute to authenticity of the experience (Radinsky et al. 2001; Rowland et al. 2016) and students' understanding of the research process. We discussed the rubric at the start of the inquiry project and used it throughout the process, with in class discussions of sample student work based on the rubric and peer reviews using the rubric, which could influence to a degree, students' interpretation of what evaluating the quality of a scientific report means.

One of the concerns in survey studies is if the respondents can correctly interpret and comprehend the survey questions (Drennan 2003). Researchers have used cognitive interviews, as a "way of assessing respondents' understanding of questionnaire items" (Park et al. 2017, p. 2), to address this concern (Drennan 2003; Kaplan et al. 2018). We conducted a series of cognitive interviews with students to examine their interpretations of the survey questions. Student volunteers who completed the ecology course in the Fall 2019 semester were recruited to participate in a cognitive interview. We chose to interview this group of students because they would likely be able to recall the inquiry project and survey more easily compared to past groups. During each interview, the students were asked to read each survey question and explain the meaning of the question in their own words.

A total of fifteen cognitive interviews were conducted. We analyzed the responses by searching for instances where a student might have interpreted the question incorrectly or misunderstood a specific word in the question. We did not encounter any situations in which a student did not comprehend the question, and all students were able to successfully explain the meaning of each survey question. Additionally, students made positive statements regarding the clarity of the questions such as, "Each statement is very clear and to the point; students would not have a difficult time understanding the meaning of each statement."

Performance

In Fall semesters of 2013–2016, we added an assessment of students' performance in evaluating the quality of scientific reports. Students were assigned to evaluate an inquiry report (the medium quality calibration report used in CPR) using the rubric (Fig. 1) before the start of the inquiry project. The performance of a student was determined as the percent of the items that they scored the same as the instructor. This assignment served as the pretest, and the evaluation of the same report during calibrations in CPR as the posttest, for students' ability to evaluate inquiry reports using a rubric. Given that additional learning gains are likely through the CPR process and the revision of the report, this assessment provided a conservative measure of learning gains.

Student performance in their revised inquiry reports was used to assess their learning. A total of 1222 inquiry reports were submitted and graded by the teaching assistants (TAs) using the rubric during the Fall semesters of 2013–2016. The frequency distribution of both the overall scores and the scores for individual categories were examined.

Data analyses

Each student was designated to a "major" group, majors or nonmajors, a "classification" group, lower-division (freshman or sophomore) or upper-division (junior or senior), and an "ethnicity" group, underrepresented minorities (URM) or other ethnicities (non-URM). This course was a required course for nine majors (Bioenvironmental Sciences, Environmental Studies, Horticulture, Landscape Architecture and Urban Planning, Renewable Natural Resources, Rangeland Ecology and Management, Recreation, Park and Tourism Sciences, Spatial Sciences, and Wildlife and Fisheries Sciences) and students in these majors were considered "majors." Students in all other majors, for which this course was an elective, were considered as "nonmajors" for this study. The underrepresented minority group included African Americans, American Indians/Alaska Natives, and Latinos. All statistical analyses were conducted using JMP-Pro 14 (SAS Institute Inc. 2018).

Self-perception of learning gains

Student self-reported scores, in a five-level scale, rated from 1 (*very low*) to 5 (*very high*), were treated as ordinal data. Wilcoxon signed-rank test with matched samples was used to test the pre-IP and post-IP differences in student self-reported interest in ecology, ability to formulate testable hypotheses, understanding how ecologists conduct research, and ability to evaluate quality of scientific report, for all students in all semesters (Falls 2007–2016) and by semester, major, classification, gender, and ethnicity. Cliff's delta (δ) was used as the effect

size measure and δ = 0.11, 0.28, and 0.43 were interpreted as small, medium, and large effect sizes, respectively (Vargha and Delaney 2000).

Cochran-Mantel-Haenszel test, using semester as the grouping variable, was used to test the differences between the groups of majors and nonmajors, lower- and upper-division, women and men, and URM and non-URM, in the gains of student self-reported interest in ecology, ability to formulate testable hypotheses, understanding how ecologists conduct research, and ability to evaluate quality of scientific report. Cramer's V was used as the effect size measure and V = 0.10, 0.30 and 0.50 were interpreted as small, medium and large effect sizes, respectively with df = 1.

Performance

Comparisons were made between the scores for student evaluation of inquiry reports before the inquiry project (pre-IP) and after the completion of the first submission of the inquiry project ("post-IP," but before the completion of the CPR and revision of the report), for all students combined, majors, nonmajors, lower-division, upper-division, women, men, URM, and non-URM. The score was determined as the percent of the rubric items that a student rated correctly. Students were also grouped based on levels of pre-IP performance—grouped by B, C, D, or F (using 90, 80, 70, and 60% break points) in scores for pre-IP evaluation of inquiry report (no A scores). A mixed model ANOVA, with subject (unique student code) nested within semester as random effects, was used to compare the pre-IP and post-IP scores for evaluation of inquiry reports for all students and by major, classification, gender, ethnicity, and pre-IP performance groups, respectively. The distribution of the scores was right skewed, and the variable was log-transformed for the analyses. A Bonferroni correction was used to adjust the critical α-value (0.05/13 = 0.0038). Cohen's d was used to evaluate the effect size of the differences between the pre-IP and post-IP scores. The effects of major, classification, gender, and ethnicity on the gains (post-IP score minus pre-IP score) were evaluated using a mixed model ANOVA, with subject (unique student code) nested within semester as random effects.

The grade distributions of the overall scores for the revised report, as well as the scores for the individual categories, were generated. Potential effects of major, classification, gender, ethnicity on overall score, and scores for the individual categories were evaluated using mixed model ANOVA, with subject (unique student code) nested within semester as random effects. The distributions of the response variables were right skewed, and the variables were log-transformed for the analyses.

Results

Self-perception of learning gains

With respect to the first research question on the self-perceptions of learning gains of students in their interest, ability, and understanding of the scientific process through the authentic inquiry experiences, for all and different groups of students, the results showed significant differences in the pre-IP and post-IP scores of self-reported interest in ecology, ability to formulate testable hypotheses, understanding how ecologists conduct research, and ability to evaluate the quality of scientific reports, for all students from Fall semesters of 2007–2016 (Fig. 2). When evaluated separately, the differences were also significant for each semester. Although the self-reported pre-IP and post-IP scores were significantly different in all categories, the effect size differed among the categories. The effect size was highest in understanding how ecologists conduct research, followed by ability to evaluate quality of scientific report, ability to formulate testable hypotheses, and interest in ecology (Fig. 2).

When evaluated by individual student groups from Fall semesters of 2007–2016, combined as well as for each semester, the self-reported gains in interest in ecology, ability to formulate testable hypotheses, understanding how ecologists conduct research, and ability to evaluate the quality of scientific reports were significant for every group—majors, nonmajors, lower-division students, upper-division students, women, men, URM students, and non-URM students (Fig. 3).

When comparing the self-reported gains between each pair of the student groups to explore the second research question on differences between groups, there was no statistically significant difference between majors and nonmajors in any of the categories (Fig. 4a). There were significant differences between lower- and upper-division students in all categories except for interest in ecology, with the lower-division students having higher level of gains (Fig. 4b). No significant differences were found between men and women, except for the ability to evaluate the quality of scientific reports with women having greater gains (Fig. 4c). There were significant differences between URM and non-URM students, with higher level of gains by URM students, except for ability to evaluate the quality of scientific reports (Fig. 4d). The effect sizes for these significant differences were small (Fig. 4b–d).

Performance

With respect to the third research question on the impact of the experiences on students' performance in evaluating the quality of scientific reports, the results showed that the post-IP average score for report evaluation was significantly higher than the pre-IP average score, for all students, both majors and nonmajors, both
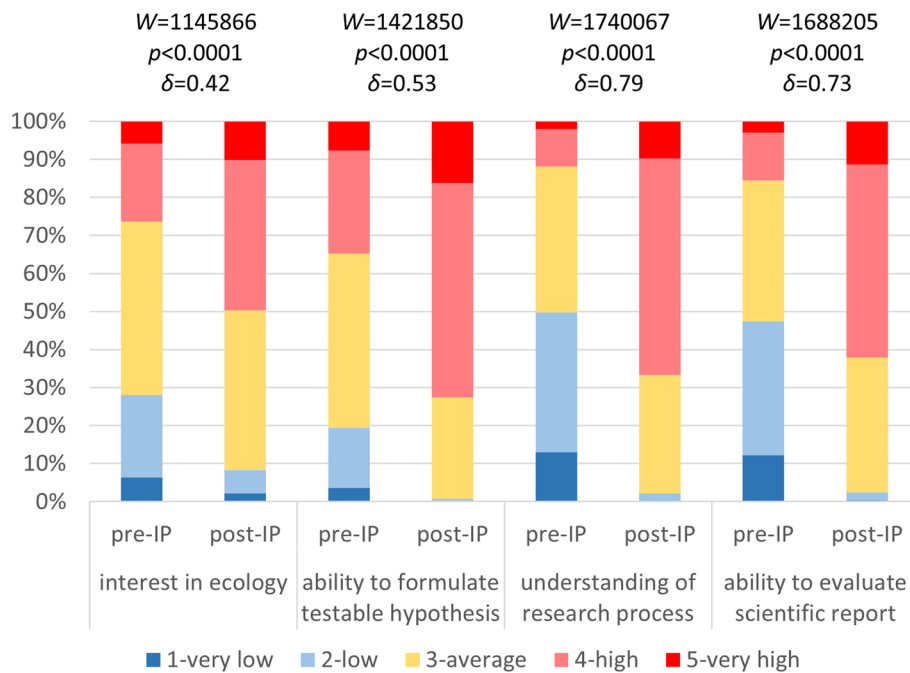
Fig. 2 Students' self-perception of learning. Distributions of self-reported levels of interest in ecology, ability to formulate testable hypotheses, understanding how ecologists conduct research, and ability to evaluate quality of scientific report, before (pre-IP) and after (post-IP) the inquiry project ($n = 2714$). The W statistic and p value for Wilcoxon signed-rank test and the effect size (Cliff's delta, δ) are shown above each category
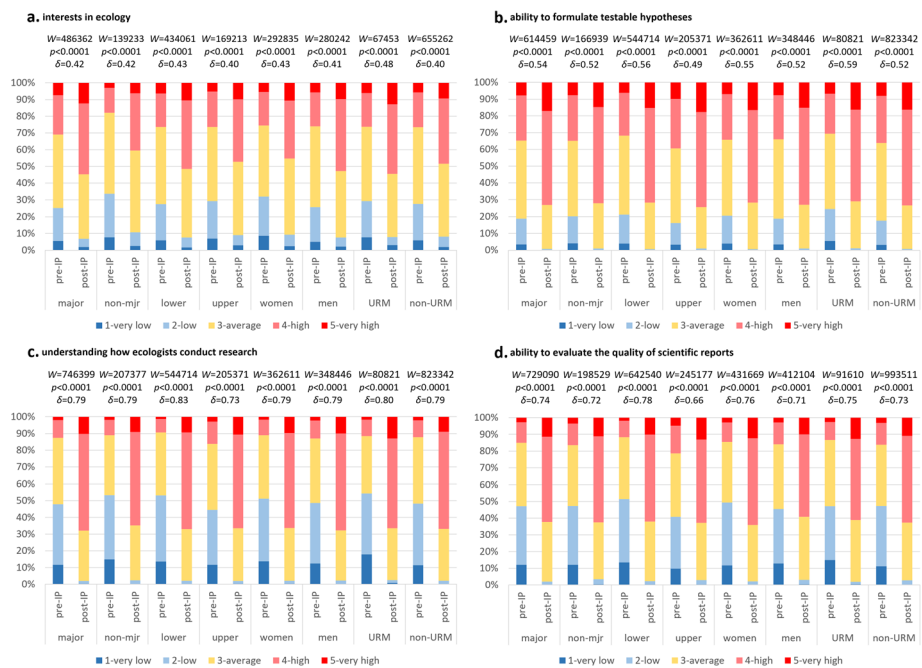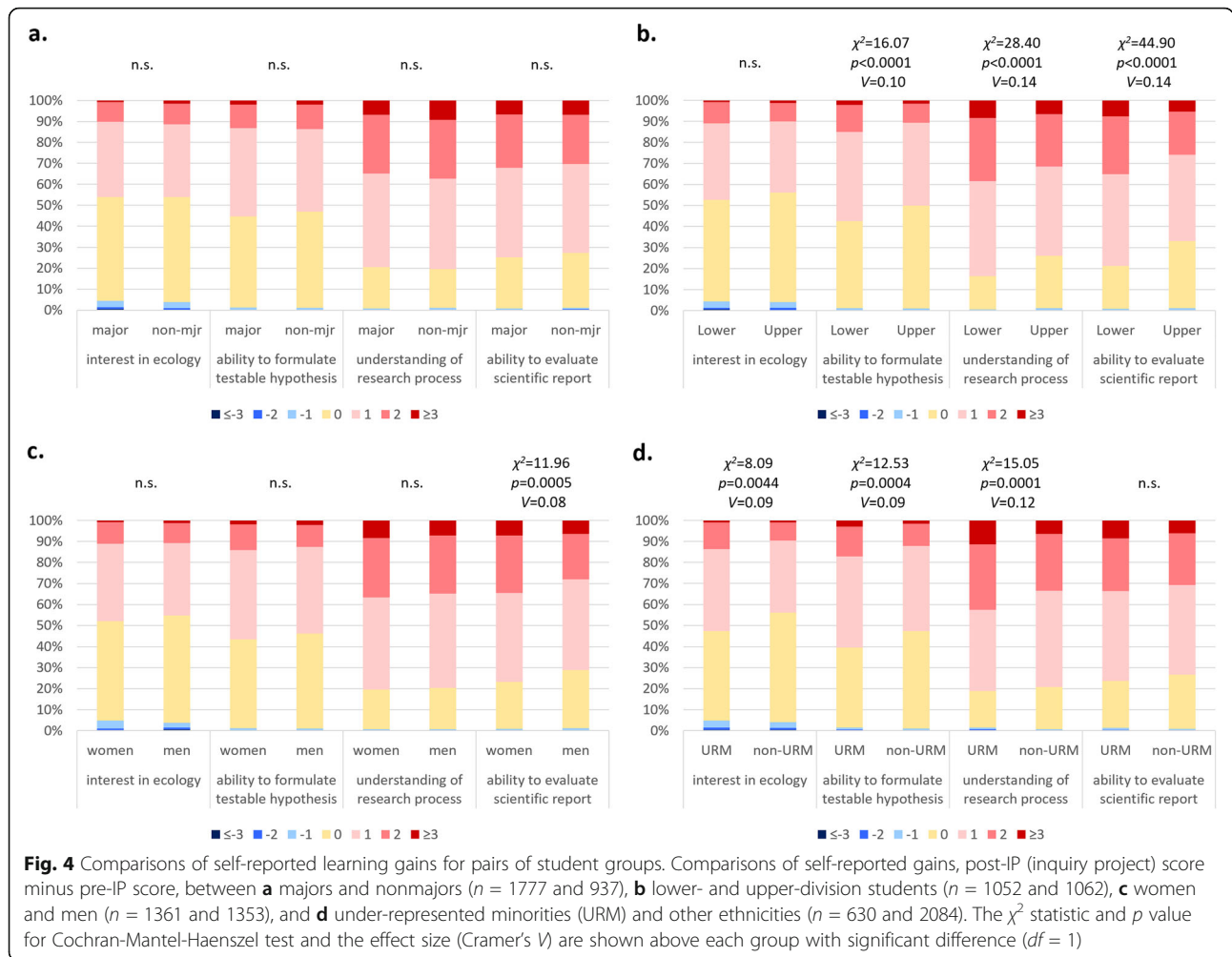

Fig. 3 Self-perception of learning by student groups. Distributions of self-reported levels of **a** interest in ecology, **b** ability to formulate testable hypotheses, **c** understanding how ecologists conduct research, and **d** ability to evaluate quality of scientific report, before (pre-IP) and after (post-IP) the inquiry project, for majors and nonmajors ($n = 1777$ and 937), lower- and upper-division students ($n = 1052$ and 1062), women and men ($n = 1361$ and 1353), and under-represented minorities (URM) and other ethnicities ($n = 630$ and 2084). The W statistic and p value for Wilcoxon signed-rank test and the effect size (Cliff's delta, δ) are shown above each group

Fig. 4 Comparisons of self-reported learning gains for pairs of student groups. Comparisons of self-reported gains, post-IP (inquiry project) score minus pre-IP score, between **a** majors and nonmajors (*n* = 1777 and 937), **b** lower- and upper-division students (*n* = 1052 and 1062), **c** women and men (*n* = 1361 and 1353), and **d** under-represented minorities (URM) and other ethnicities (*n* = 630 and 2084). The χ^2 statistic and *p* value for Cochran-Mantel-Haenszel test and the effect size (Cramer's *V*) are shown above each group with significant difference (*df* = 1)

lower- and upper-division students, both women and men, and both URM and non-URM students (Fig. 5).

When examined by pre-IP performance groups (B, C, D, and F), there was no statistically significant gain for pre-IP B group, but there were significant gains for the pre-IP C, D, and F groups with increasing effect size (Fig. 6). These changes in the performances of pre-IP performance groups could be attributed to regression to the mean since we did not employ a control group. The significant change (8.82%) in the overall mean, however, suggests that there was a non-random change. The change in the pre-B group was non-significant but negative (− 0.54%) which is likely due to regression to the mean. The change in pre-C group is significant (4.07%). This positive change is likely underestimated due to the effect of regression to the mean, given that the pre-IP mean for this group (74.26%) was higher than the overall pre-IP mean (68.41%). The changes for the pre-D and pre-F group are large (11.29 and 23.41%, respectively) which likely overestimate the change due to regression to the mean. If we can assume the change in pre-B

indicates the general magnitude of the effect of regression to the mean, the impact is likely relatively small. Given these considerations, the general pattern of these results is likely valid.

The findings related to the fourth research question on differences in gains in performance between each pair of the groups showed no significant difference between the gains of majors and nonmajors, lower- and upper-division students, or women and men. However, the gain of URM students (11.43 ± 14.05%) was significantly (*p* = 0.0220) greater than that of non-URM students (7.96 ± 11.59%). The performance of URM was substantially lower than that of non-URM in pre-IP evaluation, but became close to that of non-URM in post-IP evaluation.

About 46.5% of the students received an A for their revised inquiry reports overall and the percent of students received B, C, D, and F were 29.7, 14.4, 5.8, and 3.6%, respectively. Each of the 10 categories of the rubric has three items and the possible scores for each category range from 0 (meeting none of the three criteria) to 3
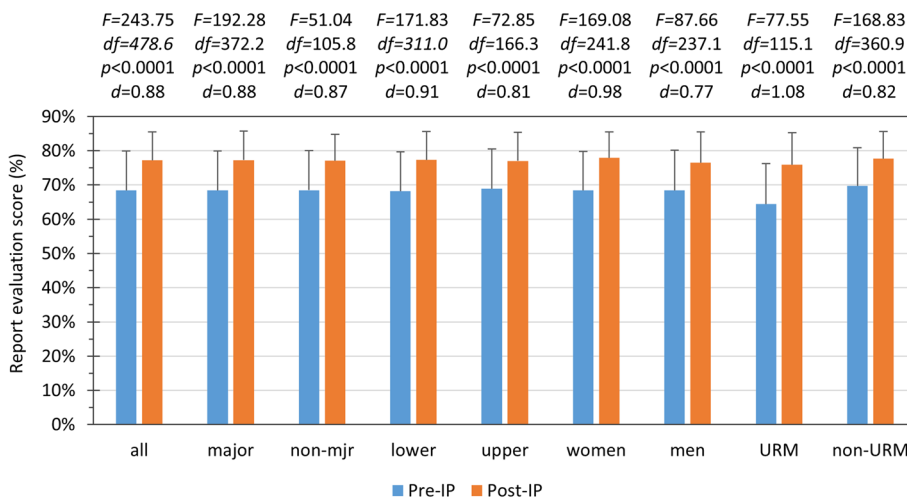
Fig. 5 Student performance in report evaluation pre- and post-inquiry project. Comparison of the pre-inquiry project (pre-IP) and post-inquiry project (post-IP) scores for evaluation of an inquiry report using a rubric, for all students (*n* = 467), majors and nonmajors (*n* = 336 and 101), lower- and upper-division students (*n* = 301 and 166), women and men (*n* = 239 and 228), and under-represented minorities (URM) and other ethnicities (*n* = 116 and 352). The *F* ratio, *df*, and *p* value for mixed model ANOVA and the effect size (Cohen's *d*) are shown above each group

(meeting all three criteria). The percent of students making each of the scores (0, 1, 2, or 3) in each of the 10 categories were shown in Fig. 7. Between 59.5% (for Discussion) and 86.7% (for Hypothesis) of the students received full scores (3) in individual categories.

There was no significant difference in the average inquiry report scores between majors and nonmajors,

lower- and upper-division students, or URM and non-URM students. However, the average score for women (89.10 ± 11.10%) was significantly (*p* = 0.0002) higher than that of men (86.32 ± 12.80%). When comparing the scores for the individual categories of the rubric, women scored significantly higher than men for Sampling (*p* = 0.0007), Discussion (*p* = 0.0038), and Writing (*p* =
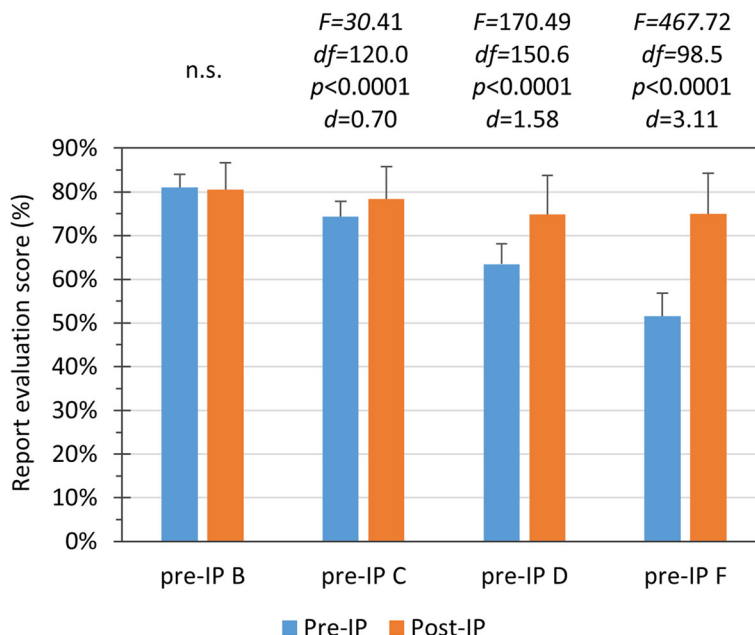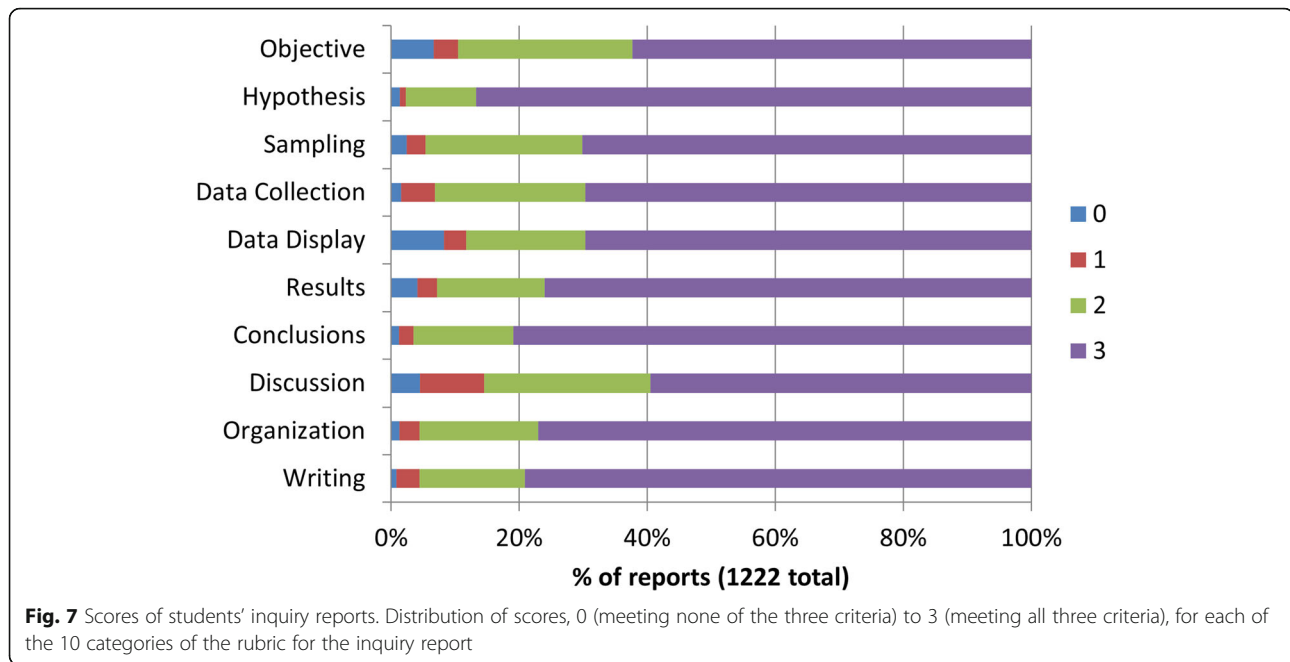

Fig. 6 Improvements in report evaluation score for students at different levels of pre-inquiry project performance. Comparison of the pre-inquiry project (pre-IP) and post-inquiry project (post-IP) scores for evaluation of an inquiry report using a rubric, for each of the pre-IP performance groups (made B, C, D, or F, respectively, in pre-IP evaluation of inquiry report; *n* = 123, 121, 126, 97). The *F*-Ratio, *df*, and *p* value for mixed model ANOVA and the effect size (Cohen's *d*) are shown above each group

Fig. 7 Scores of students' inquiry reports. Distribution of scores, 0 (meeting none of the three criteria) to 3 (meeting all three criteria), for each of the 10 categories of the rubric for the inquiry report

0.0093) (higher in other categories as well but the differences were not significant). There was no significant difference between majors and nonmajors, lower- and upper-division students, or URM and non-URM students, with the exceptions that majors scored higher for Data Collection, lower-division students scored higher for Objective, and non-URM students scored higher for Conclusions.

Discussion

Authentic inquiry experiences consistently associated with significant learning gains for all students

Although studies show that providing lab-based open inquiry and research experiences offers many benefits to undergraduate students, at large institutions the high-enrollment of many undergraduate science courses creates challenges to offering these opportunities. The results of this study show that web-based authentic inquiry projects in high-enrollment introductory courses such as the Authentic Inquiry Project have the potential to address these challenges and provide *all* students with valuable inquiry experience that models scientific research in the field. These authentic inquiry experiences were consistently associated with significant learning gains in both student self-perception of learning and their performance, for all students regardless of major, classification, gender, or ethnicity (Figs. 2 and 3).

According to Linn and colleagues (2015), much of the evidence related to the benefits of undergraduate research is based on student self-report surveys or interviews. They reviewed 60 empirical studies published in the last 5 years and found that only four directly

measured gains in research capabilities or conceptual understanding and over half solely relied on self-report surveys, which they claim has serious limitations. The long-term (2007–2016) data on student self-perception of learning in this study were used in conjunction with student performance data from fall semesters 2013–2016, which consistently show significant learning gains for all students (Fig 5).

Student performance in evaluating the quality of scientific reports using a rubric pre- and post-IP showed significant gains (Fig. 5). These gains represent not only improved understanding of the process and communication of science but also the improved skills of evaluation based on criteria—higher level thinking skills and general competencies that are important goals of undergraduate STEM education (NSTC 2018). Furthermore, it is noteworthy that the authentic inquiry experiences were particularly helpful to students who were less prepared with regard to these skills. The students with lower initial performance levels achieved greater gains, which narrowed the performance gap (Fig. 6). This is significant because traditional undergraduate research experiences are often provided to selected and self-selected students with better preparation and higher levels of interest (Eagan et al. 2011; Linn et al. 2015). Additionally, although there are some examples of CUREs being offered in high-enrollment introductory lab courses (Brownell et al. 2013), they are mostly offered to smaller classes of majors (Bangera and Brownell 2014). These authentic inquiry experiences can facilitate learning gains for a large number of students with various prior experiences (e.g., majors and nonmajors) and levels of preparation.

A high percentage of students were able to acquire the inquiry skills identified in the rubric and received high scores on their inquiry reports (Fig. 7), which reflected attainment of their understanding of the research process. The high scores for the hypothesis category corroborated the self-reported gains in ability to formulate testable hypothesis. Student performance on inquiry reports also provide some evidence for attainment of written communication skills (Fig. 7). It is important to note, however, that the actual performance of students may not be as high as reflected by these scores. This is a result of TAs' concern about being able to adequately justify taking points off if questioned by students.

Strong learning gains for lower-division students

Research shows that students are more likely to leave STEM majors during the first 2 years of college (Seymour and Hewitt 1997) and their experiences in their first-year science courses have a strong influence in students' decision to leave a science major (Strenta et al. 1994). The "Engage to Excel" report called for focused efforts to improve the lower-division undergraduate STEM education and engage students in STEM experiences during the first 2 years of college (PCAST 2012). In light of these research and call for actions, the significant gain in the interest in ecology of lower-division students, and especially the difference of interest in ecology between lower- and upper-division students post-IP showing lower-division students having significantly ($p < 0.0084$) higher levels self-perception of interest is important to note. If students maintain interest in their major, they are more likely to persist in their major. Learning gains in this area could have a potential impact on retention and persistence in STEM, as well as on future career and/or graduate school considerations (Lopatto 2009). According PCAST (2012), fewer than 40% of students who have an interest in science graduate with a STEM degree. Capturing and maintaining students' interest in science early on in their studies could help address this concern.

It is also important to note that both lower- and upper-division students had significant gains in self-perception of ability to formulate hypothesis, understanding of how research is conducted, and ability to evaluate a scientific report. Furthermore, the gains for lower-division students were significantly higher than those for upper-division students in all three categories (Fig. 4b), with lower pre-IP levels for lower-division students and similar post-IP levels between lower- and upper-division students. Interestingly, the lower- and upper-division students actually had similar levels of performance in evaluating a scientific report, both pre-IP and post-IP, with similar gains (Fig. 5), which was inconsistent with the pattern of their self-reported abilities

to evaluate a scientific report. It appears either the perceptions of lower-division students underestimated their abilities, or the perceptions of upper-division students over-estimated their abilities to evaluate a scientific report. It would be worthwhile to explore how well the upper-division students were able to gain these abilities through upper-level courses and whether the authentic inquiry experiences for the lower-division students would enable them to gain more through upper-level courses.

These authentic inquiry experiences, although limited in sophistication and nuances, could serve as an early intervention for the majors to promote their interest and enhance their understanding and skills of the scientific process, which can potentially support their motivation and persistence and improve their readiness for engagement of more in-depth research experiences in the upper-level courses (Laursen et al. 2010; Lopatto 2007; Seymour et al. 2004). In the same process, these authentic inquiry experiences facilitate the development of science literacy and STEM skills of the nonmajors and the interactions among the majors and nonmajors with diverse backgrounds and perspectives likely enrich the learning for all of students.

Significant learning gains for nonmajors

To respond to the need to develop science literacy for all, which includes understanding scientific practices and science as a social process, in addition to content knowledge (NASEM 2016), we must provide opportunities for all students, both science/STEM majors and nonmajors, to engage in authentic experiences that help them develop science literacy. Authentic inquiry projects, such as this one, show promise in helping all students develop science literacy as evidenced by the significant learning gains for both majors and nonmajors. The authentic inquiry experiences were consistently associated with significant increases in self-perception of interest in ecology for both majors and nonmajors (Fig. 3a). The majors had higher levels of self-perception of interest in ecology than nonmajors both pre- and post-IP (Fig. 3a), which is not surprising given that interest is one of the factors on which students base their choice of major. The gains in self-perception of interest in ecology, however, were similar between the two groups (Fig. 4a), suggesting that the authentic inquiry experiences were equally effective in stimulating interest in ecology for both the majors and nonmajors.

The self-perception of learning gains was highly significant for both majors and nonmajors and similar between the two groups, in ability to formulate testable hypotheses, understanding how ecologists conduct research, and ability to evaluate the quality of scientific reports (Figs. 3 and 4a). Consistent with the pattern of

their self-perception of learning gains, students' performance in evaluating the quality of scientific reports using a rubric pre- and post-IP also showed significant and similar gains for both majors and nonmajors (Fig. 5). Research has shown that there is a relationship between student attitudes toward science, interest being one measure of attitude, and achievement in science courses (Cook & Mulvihill 2008; Osborne et al. 2003). These findings suggest that authentic inquiry experiences through these inquiry projects are likely effective in facilitating the development of key elements of science literacy and STEM skills for the nonmajors.

A recent report highlights the importance of developing STEM skills for all Americans because they enable individuals to succeed in personal, professional, and civic life and are vital to the innovations and prosperity of the Nation (NSTC 2018). For most nonmajors, the few general education lower-division science elective courses are their only exposure to science. Therefore, it is critical to provide effective learning experiences of the scientific process, such as these authentic inquiry projects, for nonmajors in these introductory science courses.

Strong learning gains for women and URM students

Given the continued challenge of recruitment, retention, and graduation rates of women and URM students in STEM disciplines and the lower graduation rates of URM students across all majors, the similar gains between women and men and the greater gains of the URM than non-URM students in self-perception of interest in ecology through the authentic inquiry project (Fig. 4c, d) may be particularly salient. Both women and men self-report significant learning gains in addition to interest in ecology. Whereas women had similar levels of gains as men in self-perception of ability to formulate hypothesis and understanding of how research is conducted, they self-reported significantly greater gains than men in ability to evaluate a scientific report (Fig. 4c), although the gain in their actual performance in evaluating a scientific report in the direct assessment was similar (higher but not statistically different) to that of men. In terms of the performance in the culminating inquiry report, women had a significantly higher level of overall performance, and higher or similar level of performance in all evaluation subcategories for the inquiry reports, than men. These results reveal the importance of further investigation into differences in outcomes related to gender and self-reported gains compared to direct measures. Research findings on gender differences in students' self-evaluation of academic ability in a specific academic area or domain—referred to as academic self-concept—continue to suggest that male students typically report higher self-concepts than female students, in male stereotyped fields, such as mathematics and science, despite a lack of differences in performance (Nagy et al. 2010; Sáinz and Eccles 2012). While we cannot make a similar claim here without further investigation, the fact that women self-reported similar gains as men in some categories yet performed better than men in the culminating report is important to note. Academic self-concept has been identified as an important contributor to academic achievement (Eccles et al. 1993; Marsh et al. 2005), motivation (Eccles et al. 1983), and career choice (Luzzo and McWhirter 2001). Given the importance of self-concept in learning, future pre-IP self-reports, which were designed as formative assessments in this study, can be used more intentionally to guide pedagogical decisions throughout the course related to student self-concept.

Both URM and non-URM students reported significant learning gains in addition to self-perception of interest in ecology. URM students had greater gains in self-perception of ability to formulate hypotheses and understanding of how research is conducted than non-URM students (Fig. 4d). Although URM students reported similar levels of gains as non-URM students in self-perception of ability to evaluate a scientific report (Fig. 4d), the gain in their actual performance in evaluating a scientific report in direct assessment was significantly greater than that of non-URM students. The pre-IP performance of URM was lower, but their post-IP performance was similar to that of non-URM students. The overall performance in the inquiry report was similar between the URM and non-URM students which suggests that the IP helped close the gap between URM students' initial preparation and performance compared to non-URM students.

The learning gains in both self-perception of learning and actual performance for women and URM students may have implications for academic success, retention in general, and retention and success in STEM. Research shows that students' belief in their ability to successfully accomplish academic tasks (i.e., their sense of self-efficacy) can impact performance and persistence in their educational pursuits (Bandura 1993; Lent et al., 2003). MacPhee et al. (2013) identify the importance of self-efficacy to student performance for women and URM students in particular. Students' confidence in their abilities can also impact their sense of belonging in their chosen field and/or in college in general. Women and URM STEM students are especially vulnerable to feeling as if they do not fit in (Stout et al. 2011) and integration and sense of belonging has been shown to be a predictor of student academic success in STEM disciplines (Cole and Espinoza 2008). Written responses from student surveys regarding their experiences with the inquiry project consistently highlight the value of working with peers throughout the project. For future studies

related to this project, it would be beneficial to investigate the specific benefits of peer learning and whether it contributed positively to students' sense of belonging and self-efficacy. Chen and Soldner (2013) reported that an increase in the probability of undergraduate college students leaving STEM majors was associated with poor performance in STEM courses relative to non-STEM courses. They also reported that less success in STEM courses was also associated with an increased probability of dropping out of college altogether. The findings in this study related to women and URM students make a strong case for providing authentic inquiry experiences in lower-division science courses, where attitudes about learning and ability in the sciences are shaped.

Limitations

A limitation of this study is that it is not a true experiment with controls. It is conducted in one large introductory course over time. It was a diverse class with 400–550 students from 30 to 50 different majors each semester (in two sections). The situational factors for the class varied over the decade as well, with changes in student demographics, TAs and instructors, textbook, classroom space and learning management system, and instructor experience and teaching practices. The consistent, significant changes in both student self-perception and student performance, despite this variability over the decade, provide some confidence in the meaningful impact of the authentic inquiry experiences. The design of the study resembled, to a degree, the Recurrent Institutional Cycle Design (Campbell and Stanley 1963) where the pretests act as a quasi-control group and the repetition of the course shows replication. If a consistent effect is found despite the uncontrolled situational factors, such as in this study, then there is an argument for a treatment effect.

Although the eight survey questions related to self-perceptions of learning gains were evaluated positively using retroactive cognitive interviews, they were part of an online survey that was designed for formative assessment for the inquiry project and was not a validated instrument. Given the potential role these authentic inquiry projects can play in filling the gap to provide STEM experiences for all students, further work needs to be done to develop or adapt validated instruments to assess the impact of these authentic inquiry experiences on student learning.

The survey questions related to self-perceptions of learning gains used a retrospective pretest approach, which likely overestimate the gains (Taylor et al. 2009) due to motivational biases (Hill and Betz 2005). These survey questions were designed with a focus on facilitating student self-reflection on their learning (as a learning process), as well as assessing their self-perception of change through the inquiry project, for which the use of

retrospective pretest was more appropriate (Hill and Betz 2005). Use of a prospective pretest, administered at the beginning of the project, would result in lower gains, likely underestimated gains due to response-shift biases (Howard and Dailey 1979), which are more appropriate for estimating program effects (Hill and Betz 2005). The true effect sizes for the changes in student self-perceptions are likely lower than the estimated ones. If the true effect sizes were only half of the estimated ones, we would have small to medium effect size for changes in self-perceptions of interest in ecology and medium to large effect sizes for changes in self-perceptions of ability to formulate testable hypotheses, understanding how ecologists conduct research, and ability to evaluate the quality of scientific reports.

Conclusions

Both scholarly research and national policies identify critical needs for reform in undergraduate science education to provide STEM experiences in lower-division courses to engage and improve the retention of students for developing the STEM workforce. Reforms in undergraduate science education are also critically needed to develop science literacy and STEM skills for all students, which help individuals succeed in personal, professional, and civic life and are vital to the innovations and prosperity of the Nation. Open or authentic inquiry and undergraduate research are effective science experiences leading to multiple gains in student learning and development. Most inquiry-based learning activities, however, are implemented in laboratory classes and the majority of them are guided inquiries. Although highly effective, the traditional apprentice approach of undergraduate research can only engage a limited number of selected students; and the development in course-based undergraduate research experiences (CURE) has significantly expanded the reach. It is still a great challenge, however, to provide STEM research experiences to all students, especially to nonmajors and in large introductory courses. It is productive to explore other effective approaches to provide STEM experiences that allow students to engage in the scientific process.

We examined a web-based authentic inquiry project implemented in a high-enrollment introductory ecology course for over a decade. Our results in both student self-perception of learning and their actual performance show that the authentic inquiry experiences were consistently associated with significant learning gains in terms of interest, understanding, and skills of the scientific process for all students. The authentic inquiry experiences proved effective for lower-division students, nonmajors, and women and URM students. Lower-division students self-reported significant gains in interest in ecology and understanding and skills of the

scientific process and had significant gains in performance in evaluating a scientific report, at levels similar or greater than those of upper division students. Nonmajors had significant learning gains at a similar level as majors, based on both self-perception of learning and actual performance. The significant learning gains of women and URM students were similar or greater than those of their counterparts. Furthermore, the authentic inquiry experiences were particularly helpful to students who were less prepared with regard to the actual ability to evaluate a scientific report and narrowed the performance gap.

These findings suggest that authentic inquiry experiences can serve as an effective approach for engaging students in high-enrollment, introductory science courses. They can be an early-intervention for the majors to promote their interest and enhance their understanding and skills of the scientific process and improve their readiness for engagement of more in-depth research experiences in the upper-level courses. In the same process, providing these authentic inquiry experiences to all students can facilitate the development of science literacy and STEM skills of nonmajors and help them succeed in personal, professional, and civic life and contribute to the innovations and prosperity of society.

Abbreviations
ANOVA: Analysis of variance; CPR: Calibrated Peer Review; CURE: Course-based undergraduate research experiences; IP: Inquiry project; STEM: Science, Technology, Engineering, and Mathematics; URE: Undergraduate research experiences

Authors' contributions
XBW, SK, and JFS conceived the design of the study. XBW implemented the inquiry and XJ contributed significantly to the implementation of the project and assessments of student performance. MM conducted and analyzed the cognitive interviews. XBW performed the statistical analyses. XBW and CS conducted literature review, interpreted the results, and drafted the manuscript. All authors edited and approved the final manuscript.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX 77843, USA. [2]Teaching + Learning Commons, University of California-San Diego, La Jolla, CA 92093, USA. [3]School of Education & Human Development, Southern Methodist University, Dallas, TX 75275, USA. [4]Department of Mathematics, Texas A&M University, College Station, TX 77843, USA.

References
American Association for the Advancement of Science (AAAS) (2011). *Vision and change in undergraduate biology education: a call to action*. American Association for the Advancement of Science.

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., … Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: a meeting report. *CBE—Life Sciences Education, 13*(1), 29–40. https://doi.org/10.1187/cbe.14-01-0004.

Bakshi, A., Patrick, L. E., & Wischusen, E. W. (2016). A framework for implementing course-based undergraduate research experiences (CUREs) in freshman biology labs. *The American Biology Teacher, 78*(6), 448–455. https://doi.org/10.1525/abt.2016.78.6.448.

Ballen, C. J., Blum, J. E., Brownell, S., Hebert, S., Hewlett, J., Klein, J. R., McDonald, E. A., Monti, D. L., Nold, S. C., Siemmons, K. E., Soneral, P., & Cotner, S. (2017). A call to develop course-based undergraduate research experiences (CUREs) for nonmajors courses. *CBE-Life Sciences Education, 16*(2). https://doi.org/10.1187/cbe.16-12-0352.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117–148. https://doi.org/10.1207/s15326985ep2802_3.

Bangera, G., & Brownell, S. E. (2014). Course-based undergraduate research experiences can make scientific research more inclusive. *CBE—Life Sciences Education, 13*(4), 602–606. https://doi.org/10.1187/cbe.14-06-0099.

Bauer, K. W., & Bennett, J. S. (2003). Alumni perceptions used to assess undergraduate research experience. *Journal of Higher Education, 74*(2), 210–230. https://doi.org/10.1353/jhe.2003.0011.

Beck, C., Butler, A., & Burke da Silva, K. (2017). Promoting inquiry-based teaching in laboratory courses: Are we meeting the grade? *CBE—Life Sciences Education, 13*(3), 444–452. https://doi.org/10.1187/cbe.13-12-0245.

Beck, C. W., & Blumer, L. S. (2012). Inquiry-based ecology laboratory courses improve student confidence and scientific reasoning skills. *Ecosphere, 3*(12), 1–11. https://doi.org/10.1890/ES12-00280.1.

Bell, R. L., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction. *The Science Teacher, 72*(7), 30–33.

Bielik, T., & Yarden, A. (2016). Promoting the asking of research questions in a high-school biotechnology inquiry-oriented program. *International Journal of STEM Education, 3*(1), 15. https://doi.org/10.1186/s40594-016-0048-x.

Brownell, S. E., Kloser, M. J., Fukami, T., & Shavelson, R. J. (2013). Context matters: volunteer bias, small sample size, and the value of comparison groups in the assessment of research-based undergraduate introductory biology lab courses. *Journal of Microbiology & Biology Education, 14*(2), 176–182. https://doi.org/10.1128/jmbe.v14i2.609.

Buchanan, S., Harlan, M. A., Bruce, C., & Edwards, S. (2016). Inquiry based learning models, information literacy, and student engagement: a literature review. *School Libraries Worldwide, 22*(2), 23–39.

Buck, L. B., Bretz, S. L., & Towns, M. H. (2008). Characterizing the level of inquiry in the undergraduate laboratory. *Journal of College Science Teaching, 38*, 52–58.

Campbell, A., & Skoog, G. D. (2008). Transcending deficits and differences through undergraduate research. In R. Taraban, & R. L. Blanton (Eds.), *Creating*

effective undergraduate research programs in science: the transformation from student to scientist, (pp. 206–214). Teachers College Press.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Chang, M. J., Sharkness, J., Hurtado, S., & Newman, C. B. (2014). What matter in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research in Science Teaching*, 51(5), 555–580. https://doi.org/10.1002/tea.21146.

Chen, X., & Soldner, M. (2013). *STEM attrition: college students' paths into and out of STEM fields, Statistical analysis report. Report NCES 2014–001* (). US Department of Education.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218. https://doi.org/10.1002/sce.10001.

Cole, D., & Espinoza, A. (2008). Examining the academic success of Latino students in science technology engineering and mathematics (STEM) majors. *Journal of College Student Development*, 49(4), 285–300. https://doi.org/10.1353/csd.0.0018.

Cook, M., & Mulvihill, T. M. (2008). Examining US college students' attitudes towards science: learning from non-majors. *Educational Research and Review*, 3, 38–47.

D'Avanzo, C. (1996). Three ways to teach ecology labs by inquiry: guided, open-ended, and teacher-collaborative. *Bulletin of the Ecological Society of America*, 77, 92–93. https://doi.org/10.2307/20168031.

Derting, T. L., & Ebert-May, D. (2010). Learner-centered inquiry in undergraduate biology: Positive relationships with long-term student achievement. *CBE—Life Sciences Education*, 9(4), 462–472. https://doi.org/10.1187/cbe.10-02-0011.

Desai, K. V., Gatson, S. N., Stiles, T. W., Stewart, R. H., Laine, G. A., & Quick, C. M. (2008). Integrating research and education at research extensive universities with research: intensive communities. *Advances in Physiology Education*, 32(2), 136–141. https://doi.org/10.1152/advan.90112.2008.

Drennan, J. (2003). Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing*, 42(1), 57–63. https://doi.org/10.1046/j.1365-2648.2003.02579.x.

Eagan, K., Herrera, F. A., Garibay, J. C., Hurtado, S., & Chang, M. (2011). *Becoming STEM proteges: Factors predicting the access and development of meaningful faculty-student relationships*. Paper presented at the Association for Institutional Research Annual Forum.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation*, (pp. 75–146). W. H. Freeman.

Eccles, J. S., Wigfield, A., Harold, R., & Blumfeld, P. (1993). Age and gender differences in children's achievement self-perceptions during the elementary school years. *Child Development*, 64(3), 830–847. https://doi.org/10.2307/1131221.

Fay, M. E., Grove, N. P., Towns, M. H., & Bretz, S. L. (2007). A rubric to characterize inquiry in the undergraduate chemistry laboratory. *Chemistry Education Research and Practice*, 8(2), 212–219. https://doi.org/10.1039/B6RP90031C.

Feinstein, N. W., Allen, S., & Jenkins, E. (2013). Outside the pipeline—reimagining science education for nonscientists. *Science*, 340(6130), 314–317. https://doi.org/10.1126/science.1230855.

Finley, A., & McNair, T. (2013). *Assessing underserved students' engagement in high-impact practices*. American Association of Colleges and Universities.

Gehring, K. M., & Eastman, D. A. (2007). Information fluency for undergraduate biology majors: applications of inquiry-based learning in a developmental biology course. *CBE—Life Sciences Education*, 7, 54–63. https://doi.org/10.1187/cbe.07-10-0091.

Goldey, E. S., Abercrombie, C. L., Ivy, T. M., Kusher, D. I., Moeller, J. F., Rayner, D. A., … Spivey, N. W. (2012). Biological inquiry: a new course and assessment plan in response to the call to transform undergraduate biology. *CBE—Life Sciences Education*, 11(4), 353–363. https://doi.org/10.1187/cbe.11-02-0017.

Gormally, C., Brickman, P., Hallar, B., & Armstrong, N. (2009). Effects of inquiry-based learning on students' science literacy skills and confidence. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 16. https://doi.org/10.20429/ijsotl.2009.030216.

Handelsman, J., Miller, S., & Pfund, C. (2007). *Scientific teaching*. W. H. Freeman and Company.

Harrison, M., Dunbar, D., Ratmansky, L., Boyd, K., & Lopatto, D. (2011). Classroom-based science research at the introductory level: changes in career choices and attitude. *CBE—Life Sciences Education*, 10(3), 279–286. https://doi.org/10.1187/cbe.10-12-0151.

Hester, S. D., Nadler, M., Katcher, J., Elfring, L. K., Dykstra, E., Rezende, L. F., & Bolger, M. S. (2018). Authentic inquiry through modeling in biology (AIM-Bio): an introductory laboratory curriculum that increases undergraduates' scientific agency and skills. *CBE—Life Sciences Education*, 17(4), ar63 1-23. https://doi.org/10.1187/cbe.18-06-0090.

Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501–517. https://doi.org/10.1177/1098214005281356.

Howard, D. R., & Miskowski, A. J. (2005). Using a module-based laboratory to incorporate inquiry into a large cell biology course. *Cell Biology Education*, 4(3), 249–260. https://doi.org/10.1187/cbe.04-09-0052.

Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: a source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144–150. https://doi.org/10.1037/0021-9010.64.2.144.

Jones, M. T., Barlow, A. E. L., & Villarejo. M. (2010). Importance of undergraduate research for minority persistence and achievement in biology. *The Journal of Higher Education*, 81(1), 82–115. https://doi.org/10.1353/jhe.0.0082.

Kaplan, S., Seabott, H., Cunningham, E., Helman, J., Calderon, A., Thirlby, R., & Schenarts, K. (2018). Resident wellness and social support: Development and cognitive validation of a resident social capital assessment tool. *Journal of Surgical Education*, 75(2), 313–320. https://doi.org/10.1016/j.jsurg.2018.02.002.

Khalaf, B. K. (2018). Traditional and inquiry-based learning pedagogy: A systematic critical review. *International Journal of Instruction*, 11(4), 545–564. https://doi.org/10.12973/iji.2018.11434a.

Kiernan, D. A., & Lotter, C. (2019). Inquiry-based teaching in the college classroom: the nontraditional student. *The American Biology Teacher*, 81(7), 479–484. https://doi.org/10.1525/abt.2019.81.7.479.

Koretsky, M., Keeler, J., Ivanovitch, J., & Cao, Y. (2018). The role of pedagogical tools in active learning: a case for sense-making. *International Journal of STEM Education*, 5(1), 18. https://doi.org/10.1186/s40594-018-0116-5.

Kuh, G. D., Kinzie, J., Schuh, J. H., Whitt, E. J., & Associates (2010). *Student success in college: creating conditions that matter*. Jossey-Bass.

Laursen, S., Hunter, A., Seymour, E., & Thiry, H. (2010). *Undergraduate research in the sciences: engaging students in real science*. Jossey-Bass.

Lent, R. W., Brown, S. D., Schmidt, J., Brenner, B., Lyons, H., & Treistman, D. (2003). Relation of contextual supports and barriers to choice behavior in engineering majors: Test of alternative social cognitive models. *Journal of Counseling Psychology*, 50(4), 458–465. https://doi.org/10.1037/0022-0167.50.4.458.

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: impacts and opportunities. *Science*, 347(6222), 1261757. https://doi.org/10.1126/science.1261757.

Lopatto, D. (2004). Survey of undergraduate research experiences (SURE): first findings. *CBE—Life Sciences Education*, 3(4), 270–277. https://doi.org/10.1187/cbe.04-07-0045.

Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. *CBE—Life Sciences Education*, 6(4), 297–306. https://doi.org/10.1187/cbe.07-06-0039.

Lopatto, D. (2009). *Science in solution: the impact of undergraduate research on student learning*. The Research Corporation for Science Advancement.

Luzzo, D. A., & McWhirter, E. H. (2001). Sex and ethnic differences in the perception of educational and career-related barriers and levels of coping efficacy. *Journal of Counseling & Development*, 79(1), 61–67. https://doi.org/10.1002/j.1556-6676.2001.tb01944.x.

MacPhee, D., Farro, S., & Canetto, S. (2013). Academic self-efficacy and performance of underrepresented STEM majors: gender, ethnic, and social class patterns. *Analyses of Social Issues and Public Policy*, 13(1), 347–369. https://doi.org/10.1111/asap.12033.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades and standardized test scores: reciprocal effects models of causal ordering. *Child Development*, 76(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x.

Nagda, B. A., Gregerman, S. R., Jonides, L., von Hippel, W., & Lerner, J. S. (1998). Undergraduate student-faculty partnerships affect student retention. *Review of Higher Education*, 22(1), 55–72. https://doi.org/10.1353/rhe.1998.0016.

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The development of students' mathematics self-concept in relation to gender: different countries, different trajectories? *Journal of Research on Adolescence*, 20(2), 482–506. https://doi.org/10.1111/j.1532-7795.2010.00644.x.

National Academies of Sciences, Engineering, and Medicine (2016). *Science literacy: concepts, contexts, and consequences*. The National Academies Press. https://doi.org/10.17226/23595.

National Research Council (NRC) (2000). *Inquiry and the national science education standards: a guide for teaching and learning*. The National Academies Press. https://doi.org/10.17226/9596.

National Science and Technology Council (NSTC) (2018). *Charting a course for success: America's strategy for STEM education*. The White House.

O'Donnell, K., Botelho, J., Brown, J., González, G. M., & Head, W. (2015). Undergraduate research and its impact on student success for underrepresented students. *New Directions for Higher Education*, *169*(169), 27–38. https://doi.org/10.1002/he.20120.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: a review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199.

Park, N., Park, C., McCreary, L., & Norr, K. (2017). Cognitive interviews for validating the family nutrition physical activity instrument for Korean-American families with young children. *Journal of Pediatric Nursing*, *36*, 1–6. https://doi.org/10.1016/j.pedn.2017.04.014.

Peffer, M. E., & Ramezani, N. (2019). Assessing epistemological beliefs of experts and novices via practices in authentic science inquiry. *International Journal of STEM Education*, *6*(1), 3. https://doi.org/10.1186/s40594-018-0157-9.

Peffer, M. E., Schunn, C., Renken, M., & Revak, A. (2015). Science classroom inquiry (SCI) simulations: a novel method to scaffold science learning. *PLoS ONE*, *10*(3), e0120638. https://doi.org/10.1371/journal.pone.0120638.

President's Council of Advisors on Science and Technology (2012). *Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Executive Office of the President.

Radinsky, J., Bouillion, L., Lento, E. M., & Gomez, L. M. (2001). Mutual benefit partnership: a curricular design for authenticity. *Journal of Curriculum Studies*, *33*(4), 405–430. https://doi.org/10.1080/00220270118862.

Renninger, K. A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist*, *46*(3), 168–184. https://doi.org/10.1080/00461520.2011.587723.

Rissing, S. W., & Cogan, J. G. (2009). Can an inquiry approach improve college student learning in a teaching laboratory? *CBE—Life Sciences Education*, *8*(1), 55–61. https://doi.org/10.1187/cbe.08-05-0023.

Rowland, A. A., Knekta, E., Eddy, S., & Corwin, L. A. (2019). Defining and measuring students' interest in biology: an analysis of the biology education literature. *CBE—Life Sciences Education*, *18*(3), 1–14. https://doi.org/10.1187/cbe.19-02-0037.

Rowland, S., Pedwell, R., Lawrie, G., Lovie-Toon, J., & Hung, Y. (2016). Do we need to design course-based undergraduate research experiences for authenticity. *CBE—Life Sciences Education*, *15*(4), 1–16. https://doi.org/10.1187/cbe.16-02-0102.

Russell, C. B., & Weaver, G. C. (2010). A comparative study of traditional, inquiry-based, and research-based laboratory curricula: impacts on understanding of the nature of science. *Chemistry Education Research and Practice*, *12*(1), 57–67. https://doi.org/10.1039/C1RP90008K.

Russell, S. H., Hancock, M. P., & McCullough, J. (2007). The pipeline: benefits of undergraduate research experiences. *Science*, *31*(5824), 548–549. https://doi.org/10.1126/science.1140384.

Sadler, T. D., Burgin, S., McKinney, L., & Ponjuan, L. (2010). Learning science through research apprenticeships: a critical review of the literature. *Journal of Research in Science Teaching*, *47*(3), 235–256. https://doi.org/10.1002/tea.20326.

Sáinz, M., & Eccles, J. (2012). Self-concept of computer and math ability: gender implications across time and within ICT studies. *Journal of Vocational Behavior*, *80*(2), 486–499. https://doi.org/10.1016/j.jvb.2011.08.005.

SAS Institute Inc (2018). *JMP® 14 basic analysis*. SAS Institute Inc.

Schielack, J. F., & Knight, S. L. (2012). *The new science education leadership: an IT-based learning ecology model*. Teachers College Press.

Schwab, J. (1962). The teaching of science as enquiry. In J. J. Schwab, & P. F. Brandwein (Eds.), *The teaching of science*, (pp. 1–103). Simon and Schuster.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: why undergraduates leave the sciences*. Westview Press.

Seymour, E., Hunter, A. B., Laursen, S. L., & Deantoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three-year study. *Science Education*, *88*(4), 493–534. https://doi.org/10.1002/sce.10131.

Shaffer, C. D., Alvarez, C. J., Bednarski, A. E., Dunbar, D., Goodman, A. L., Reinke, C., … Elgin, S. C. R. (2014). A course-based research experience: how benefits change with increased investment in instructional time. *CBE-Life Science Education*, *13*(1), 111–130. https://doi.org/10.1187/cbe-13-08-0152.

Spell, R. M., Guinan, J. A., Miller, K. R., & Beck, C. W. (2014). Redefining authentic research experiences in introductory biology laboratories and barriers to their implementation. *CBE-Life Science Education*, *13*(1), 102–110. https://doi.org/10.1187/cbe.13-08-0169.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *100*(2), 255–270. https://doi.org/10.1037/a0021385.

Strenta, A. C., Elliott, R., Adair, R., Matier, M., & Scott, J. (1994). Choosing and leaving science in highly selective institutions. *Research in Higher Education*, *35*(5), 513–547. https://doi.org/10.1007/BF02497086.

Sundberg, M. D., Armstrong, J. E., & Wischusen, E. W. (2005). A reappraisal of the status of introductory biology laboratory education in U.S. colleges & universities. *The American Biology Teacher*, *67*(9), 525–529. https://doi.org/10.2307/4451904.

Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Gilding the outcome by tarnishing the past: Inflationary biases in retrospective pretests. *American Journal of Evaluation*, *30*(1), 31–43. https://doi.org/10.1177/1098214008328517.

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, *25*(2), 101–132. https://doi.org/10.3102/10769986025002101.

Weaver, G., Russell, C., & Wink, D. (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.

Webb, A. M., Knight, S. L., Wu, X. B., & Schielack, J. F. (2014). Teaching science with web-based inquiry projects: an exploratory investigation. *International Journal of Virtual and Personal Learning Environments*, *5*(2), 57–68. https://doi.org/10.4018/ijvple.2014040105.

Wei, C. A., & Woodin, T. (2011). Undergraduate research experiences in biology: alternatives to the apprenticeship model. *CBE-Life Sciences Education*, *10*(2), 123–131. https://doi.org/10.1187/cbe.11-03-0028.

Wheeler, L., & Bell, R. (2012). Open-ended inquiry: practical ways of implementing inquiry in the chemistry classroom. *The Science Teacher*, *79*(6), 32–39.

Wu, X. B., Knight, S. L., Schielack, J. F., Robledo, D. C., Jaime, X. A., Peterson, C. A., & Griffing, L. R. (2016). Authentic ecological inquiries using BearCam archives. *CourseSource*. https://doi.org/10.24918/cs.2016.22.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.