Journal of Natural Science and Exploration

PRIME OPEN ACCESS

Volume 1, Issue 1

Research Article

Date of Submission: 05 May, 2025
Date of Acceptance: 06 June, 2025
Date of Publication: 16 June, 2025

Improving 7th Grade Students' Understanding of How Science Works

Wang Yuedong[1*] and Norbert Reich[2]

[1]Department of Chemistry and Biochemistry, University of California Santa Barbara, California, 93106,

[2]Department of Statistics and Applied Probability, University of California, Santa Barbara, California, 93106

***Corresponding Author:**
Wang Yuedong, Department of Chemistry and Biochemistry, University of California Santa Barbara, California, 93106.

Citation: Yuedong, W., Reich, N. (2025). Improving 7th Grade Students' Understanding of how Science Works. *J of Natural Science and Exploration, 1*(1), 01-14.

Abstract
We examined the effectiveness of the SciTrek outreach program involving two junior high teachers in the same school. SciTrek has run modules in 2nd-8th grade classrooms since 2010, with each classroom participating in two modules per academic year. For this study, students (206) in eight classrooms took part in one 7th grade module focused on conservation of mass lasting six days during the Fall, and a different module focused on cellular respiration and lasting eight days in the Winter. Students took a pre-assessment prior to the modules to determine their ability to identify a testable question, to plan and design an experimental investigation, to analyze and interpret data with computational thinking. Upon finishing both modules, students took the same assessment. Students improved significantly in all aspects, showing improvements of 10% to 45%. Students with lower academic performance (College Preparatory, three classrooms) improved more and showed more variability in their improvement than those with higher prior academic performance (Honors). The population included socioeconomically disadvantaged (57%) and Hispanic or Latino (64%) students. We discuss the importance of key features of the approach: the inclusion of four to six trained university volunteers in each classroom to guide small group discussions, and the active engagement of the 7th grade students in developing their own questions, designing their own experiments, data collection, and the defense of their conclusion to their peers.

Keywords: Critical Thinking, Science Outreach, Middle School and Science Outreach

Introduction
Two observations underpin the 2009 launch of the SciTrek outreach program from the Chemistry and Biochemistry Department at the University of California at Santa Barbara (UCSB): The broad consensus that science education at the K-12 as well as at the university levels needs to be improved for multiple reasons [1-8]. Also, in spite of the existence of creative, diverse and effective approaches to address this problem, there is room for improvement in student performance or retention at the national level [9-19].

What are the major, acknowledged challenges for K-12 science education in the US?
While some challenges are grade-level specific (e.g., elementary students typically only receive on average less than three hours per week of science instruction), most K-12 grades witness similar problems [20]. An emphasis on subject matter content has been historically seen as problematic although more recent developments such as Next Generation Science Standards (NGSS) and other inquiry-based instructional methods support an increased focus on science practices and aspects of critical thinking [21,22]. The focus on content is particularly prevalent in middle school and high school classes, and continues at the university level. Strong evidence now links this content focus to a second challenge, the attrition of K-12 students from STEM courses, with the result that insufficient students choose STEM related career paths which is particularly troubling for minorities [23-30]. In spite of diverse programs to address these problems, progress

on improving science education remains unacceptably slow [1-8]. Subject matter test performance at the national and international levels and test performance on science and engineering practices remain problematic.

In many cases, programs have provided convincing statistical analysis showing their effectiveness on student retention and performance [12,18,31-39]. These programs include professional development opportunities for teachers, outreach efforts, public participation in research, reform-based curricula changes, preservice teacher education among other efforts. These advances beg the question of why improvements in STEM performance and retention of students in STEM fields remains problematic in the US, which is particularly troubling for cohorts which have traditionally been underrepresented in STEM such as minorities and women [7,23,27,29,40-42]. While the use of standardized testing to gauge student achievement, both individually and collectively, has been criticized it remains one metric which clearly shows very little evidence of improvement, both within the US as well as when comparing US students to international students [34-36,43,44]. This is true both for subject matter content as well as assessments which gauge understanding of scientific practices. Similarly, the extensive literature on the need for US workers with a STEM background has not been addressed by any significant increase in the total numbers of K-12 students obtaining university training in STEM-related fields. Particularly troubling is the lack of significant increases in women and minorities obtaining university STEM related training [23,27,29,40,41].

Methodology

The SciTrek program has been run since 2010 and covers grades 2-8, and is now expanding into high school. This study is exclusively focused on modules run in the 7th grade. We describe the origin and organization of the SciTrek program, a description of the two 7th grade modules used for this study, and the content, administration, and analysis of the assessments.

Origin and Organization of Scitrek

Given the strong evidence that effective programs to improve science education for K-12 students exist, why is progress on making real change so slow? The effectiveness of the SciTrek program in helping elementary students improve in their attitudes towards science and secondary students improve in critical thinking as applied to science practices this work, is compelling [45]. The ability of the program to address the broader concerns, for example, of improving long term retention of minority and women in STEM studies, is an ongoing interest, not addressed here.

Four insights form the basis of the SciTrek approach, presented here for the 7th grade modules, guided in part by evidence from various studies, and the corresponding author's experience in working with university students at UCSB. First, obtaining personal guidance from an "expert" is essential for novices in STEM and non-STEM subjects [11, 37]. This process is labor intensive and best done in a small group setting which is extremely challenging for both elementary and secondary teachers [3,20,31,46,47]. The small groups of K-12 students (four to six, and for this study, 7th grade students) which are mentored by the university volunteers over a period of five to eight days, represents a distinct feature of the program. While these university volunteers have not received any formal teaching training, they do receive training in the modules as well as how to ask leading questions to encourage students to improve their ability to ask testable questions, experimental designs, interpretation of experimental results, and presentation of their findings [48, 49]. The multi-day discussions between volunteers and a small group of students are simply not possible by a single person (teacher) when 30 or more students are engaged in many different experiments. A typical comment from a participating 7th grade teacher supports this statement: "As a science teacher I have not seen another program like SciTrek. NGSS emphasize having students conduct experiments, but with over 200 students a day the reality of executing that is daunting. Being able to partner with UCSB students who guide students through the process makes it possible", JH, Science Department Head, Anacapa Middle School, Ventura California. Because the volunteers are 60-70% female and 30-40% minority, these interactions help students consider how females and minorities can engage in science [50-56].

Second, to the extent possible, students need to think that what they are working on is at least in part driven by their own interests [57]. This feature requires attention and is very age- and grade-driven. Unguided inquiry has been shown to be less effective than guided inquiry [17,18,37,38]. However, rote learning from following a detailed script is quite ineffective in building student interest and confidence [58]. Third, the possibility, indeed the likelihood of experimental "failure" and subsequent analysis by students, needs to be allowed for and encouraged [44]. Fourth, secondary school science teachers, and certainly elementary school educators, have minimal exposure to authentic science investigations and instruction during their own education and teacher training [59].

SciTrek typically reaches 10-15 schools, 60-70 teachers and approximately 2000 K-12 students each year. The program started at the second-grade level, and has expanded into secondary classrooms. As a result of district scheduling policy, no single school has implemented the entire SciTrek program. The elementary program (grades 2-6) and the secondary program are organized similarly, with important differences. Because elementary teachers rarely have a science background, the program seeks to empower the teachers to take a leadership role over a two to three-year period; this involves extensive consultation with SciTrek staff. At the secondary level, although the teachers generally have science undergraduate degrees, they often lack any meaningful experience in carrying out research of any type [44]. Moreover, they rarely have the time and resources to implement a program like SciTrek; interviews and testimonials from teachers involved with SciTrek at all levels strongly support this view. At the secondary level, SciTrek engages these teachers in the design of module content. The long-term goal is to have teachers run two SciTrek modules each year with minimal involvement of SciTrek staff, with the help of trained university students, experimental materials, and workbooks.

The secondary program is further distinguished by the involvement of undergraduate and graduate students (SciTrek Research Group, SRG, coauthors of this work) who obtain research credit for assisting in all aspects of the program, including training university volunteers, module development, student recruitment, leading modules in classrooms, developing and administering assessments along with their statistical analysis and publication in peer reviewed journals. SRG members typically have at least one year of extensive module experience prior to joining and work 10-15 hours per week on the program. Many former SRG members are now enrolled in post graduate programs (e.g., Ph.D.) and training in diverse health science programs.

The SciTrek program

Recruitment, Makeup and Training of University Students

While this study is focused on the 7th grade program, SciTrek runs similar programs in grades 2-8, which is summarized here. University undergraduate and graduate students are recruited from all disciplines, but 95% are STEM majors or have an interest in education. Undergraduates enrolled in the School of Education Minor in Science and Mathematics Education as well as the Minor in Education program participate in SciTrek. Students enrolled in Chemistry 102 (Teaching Chemistry at the High School/Junior High level), 193 (Internship in Chemistry), 198 (Research in Science Education) and Education 129 (CalTeach) participate in SciTrek modules. UCSB has 23,349 undergraduates, 46 % are STEM majors, and 2,965 are graduate students.

UCSB volunteers (four to six undergraduate and graduate students per K-12 classroom) who participate in running the two modules featured in this study are provided volunteer notebooks, which contain the same activities outlined in the student notebooks in addition to supplementary materials that further explain concepts covered in the module (See Supplemental Files 1-6). Answers to the questions in the student notebook are included in the volunteer notebook. All UCSB volunteers are required to attend a module orientation lasting one to two hours, specific to each individual module (Best Bread, and Conservation of Mass), where SciTrek staff or SRG members thoroughly explain the activities, procedures, and goals of the module to volunteers. The classroom teachers attend these orientations to become familiar with the module and advise the university volunteers about their students. Volunteers are introduced to the activities covered in each module as well as informed as to what is expected of them when speaking to and interacting with the students. Volunteers are provided extensive guidance in asking students leading questions to improve the quality of the questions the 7th grade students want to experiment on (48,49), on their experimental designs, on their data analysis, and how best to present their data. The volunteer training and subsequent interactions with the students on this topic are key features of the program. All SciTrek volunteers are offered transportation to local schools in order to help run the module. Secondary teachers run the module in their classrooms with the help of the university volunteers. SciTrek staff are also present in the classroom to help and look for ways of improving the modules.

Module description (Figures 1-3 and Supplemental Files 1-6).

Day 1: (45 Minutes)

- SRG members enter the classroom for the first time wearing lab coats and bring all module equipment. The lead introduces themselves, the SciTrek program and each volunteer provides their name, major and their interests. **(5 minutes)**
- Students are introduced to the concept of Cellular Respiration and its life sustaining role in biology in a class discussion facilitated by the SciTrek lead. Students will also answer questions about cellular respiration in the student workbook. All references to the student workbook in this figure pertain to Supplementary File 1 in the supplementary materials. **(5 Minutes)**
- Students assemble in groups of four to six and one university volunteer to carry out the "Activating Yeast," activity (Figure 2A,B) involving two flasks (Figure 2C,D); each flask contains three grams of yeast and 50 mL of water, and with only one flask having a gram of white cane sugar. Students briefly stir the two flasks using a provided stir plate. Students are prompted to closely observe the two flasks and record any noticeable visual and olfactory differences between the two flasks. The yeast flask containing sugar has been 'activated,' and should produce bubbles of carbon dioxide, creating a froth whereas the yeast flask without sugar should remain unchanged. The students record their initial observations (Figure 2A) and answer questions in their SciTrek workbook (Figure 2E) **(20 minutes)**
- To illustrate the similarity between human cells and yeast cells, students observe UCSB volunteers as they blow into bottles filled with Bromothymol Blue indicator. The indicator turns yellow, indicating how humans and yeast cells both produce carbon dioxide as waste. **(5 minutes)**
- As practice for the following days, students assemble the water displacement apparatus that will be used throughout the module (Figure 2C, D). **(10 Minutes)**

Day 2: (45 Minutes)

- The entire class engages in a discussion with the SciTrek lead and volunteers about the importance of a control trial in order to prepare for the activities for the rest of the day. Students understand that control trials contain variables set to baselines in order to compare and determine effectiveness of their experimental trials on Day 3 and Day 4. **(10 Minutes)**
- Students are tasked to perform their own control trial using the water displacement apparatus (Figure 2C, D). Three grams of yeast and one gram of white cane sugar is added to the first flask at thirty-seven degrees Celsius, stirring at low speed for a total of ten minutes. Students are provided extra time to gather materials and clean up after the trial. **(15 minutes)**
- Each group of students then perform the "Maximize!" activity, where they design and run a trial where their goal is to produce the most amount of carbon dioxide. They are allowed to change the amount and type of sugar, water temperature, and stir speed. Students are allowed to change as many variables as they would like. This activity frequently results in students understanding better the importance of testing only one variable at a time as students see that using a protocol which alters multiple parameters, they are unable to determine how much a specific variable affects carbon dioxide production. Once finished with the activity, students are prompted to answer follow-up questions in their workbook (Supplemental File 1). **(20 Minutes)**

Days 3 and 4: (45 minutes)

- Each group decides what variable they would like to experiment with and determine its effect on yeast carbon dioxide production. They work with their chosen variable through both experimental trial days. **(5 minutes)**
- Students design and run at least four trials of ten minutes each with the water displacement apparatus (Figure 2D). Each group decides, with input from the volunteer (if sought), how to vary the variable of interest. Each group records and graphs their results (Figure 2E). **(40 minutes)**

Day 5: (45 minutes)

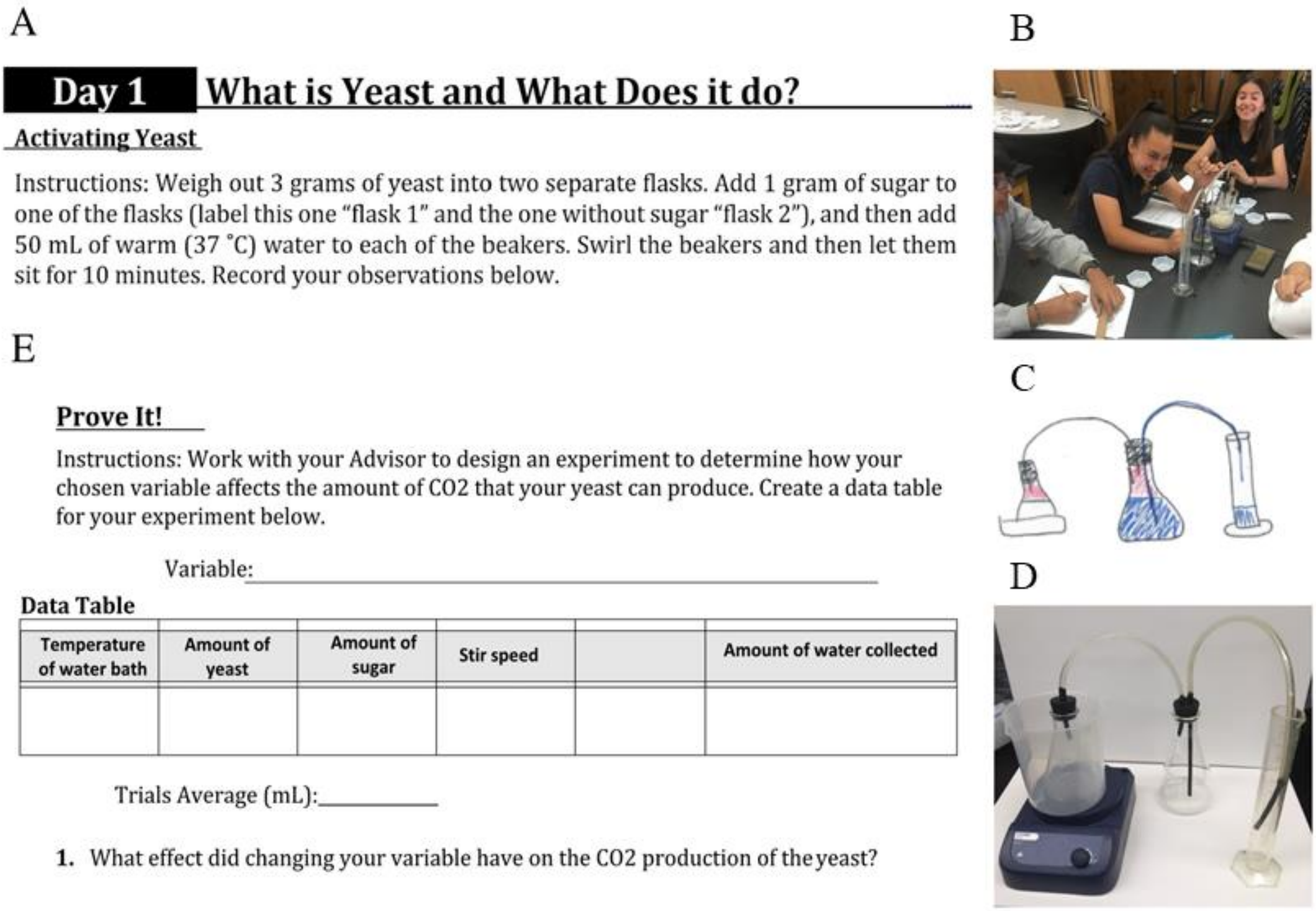
- The entire class then discusses what they believe to be the optimal condition for carbon dioxide production; volunteers are encouraged to prompt students to present evidence to support their suggestions. The entire class votes to determine the protocol to be used by all groups to generate optimized variables with which they will run their optimized trial. Students then answer questions in their student workbook based on the class results. This process is inherently complex as it involves optimizing multiple parameters; this concept is brought up and discussed in small groups and in the entire class **(10 minutes)**
- Using the optimized variables, each group runs a final trial using the optimized values and collects their results. The class average is determined, and a discussion of outliers, and variability in data is led by the teacher, SRG members and university volunteers. To incentivize good performance, the class that receives the highest average carbon dioxide produced wins a pizza party. **(20 minutes)**
- Students then answer several questions pertaining to the optimized trial in their workbook and then proceed to designing group posters that encapsulate the data gathered on days 3 and 4 (Supplemental File 1). **(15 minutes)**

Day 6: (45 minutes)

- Students finish working on their posters. Posters include their original hypothesis as well as tables and graphs displaying the variables that they have tested as well and the results from their experiments. **(20 minutes)**
- Each group presents their poster at the front of the class and the entire class is involved in questions directed to the presenters. **(25 minutes)**

Figure 1: Day to Day Schedule of the Best Bread Module

Note, the schedule for the Conservation of Mass module is presented in the supplementary materials (Supplemental File 8). The corresponding student, university volunteer and lead notebooks are also included in the Supplementary files (Supplemental Files 1-6).



Day 1 What is Yeast and What Does it do?

Activating Yeast

Instructions: Weigh out 3 grams of yeast into two separate flasks. Add 1 gram of sugar to one of the flasks (label this one "flask 1" and the one without sugar "flask 2"), and then add 50 mL of warm (37 ˚C) water to each of the beakers. Swirl the beakers and then let them sit for 10 minutes. Record your observations below.

Prove It!

Instructions: Work with your Advisor to design an experiment to determine how your chosen variable affects the amount of CO2 that your yeast can produce. Create a data table for your experiment below.

Variable:____

Data Table

Temperature of water bath	Amount of yeast	Amount of sugar	Stir speed		Amount of water collected

Trials Average (mL):__________

1. What effect did changing your variable have on the CO2 production of the yeast?

Figure 2: Best Bread Module

(see Supplemental Files 1-3 for corresponding complete notebooks). (A) Day 1 module activity. Students are tasked with using their senses, excluding taste or touch, to observe the two different flasks, one containing a solution of yeast, sugar, and water, and the other containing a solution of just yeast and water. These observations become the basis of the questions they propose to investigate in the following days and activities. (B) Photograph of students running the control trial using 3 grams of yeast and 1 gram of cane sugar in 100 mL of 37-degree water on Day 2. Students are asked to take notes and observe while the trial is running. (C) Students are asked to draw their rendition of the apparatus in the notebook on Day 1. A sample drawing is shown. Students can use their drawings as a quick reference in later days. (D) The correctly assembled device is shown. (E) Students are asked to set up the experiment apparatus consisting of two beakers, one containing the yeast mixture, the other containing water, connected with plastic tubing. They are first asked to run the control trial with three grams of sugar and one gram of yeast and record the amount of carbon dioxide gas produced from the trial. The students are then asked to perform a second trial to 'maximize' the amount of carbon dioxide gas by changing one or more variables, from a list of potential variables.

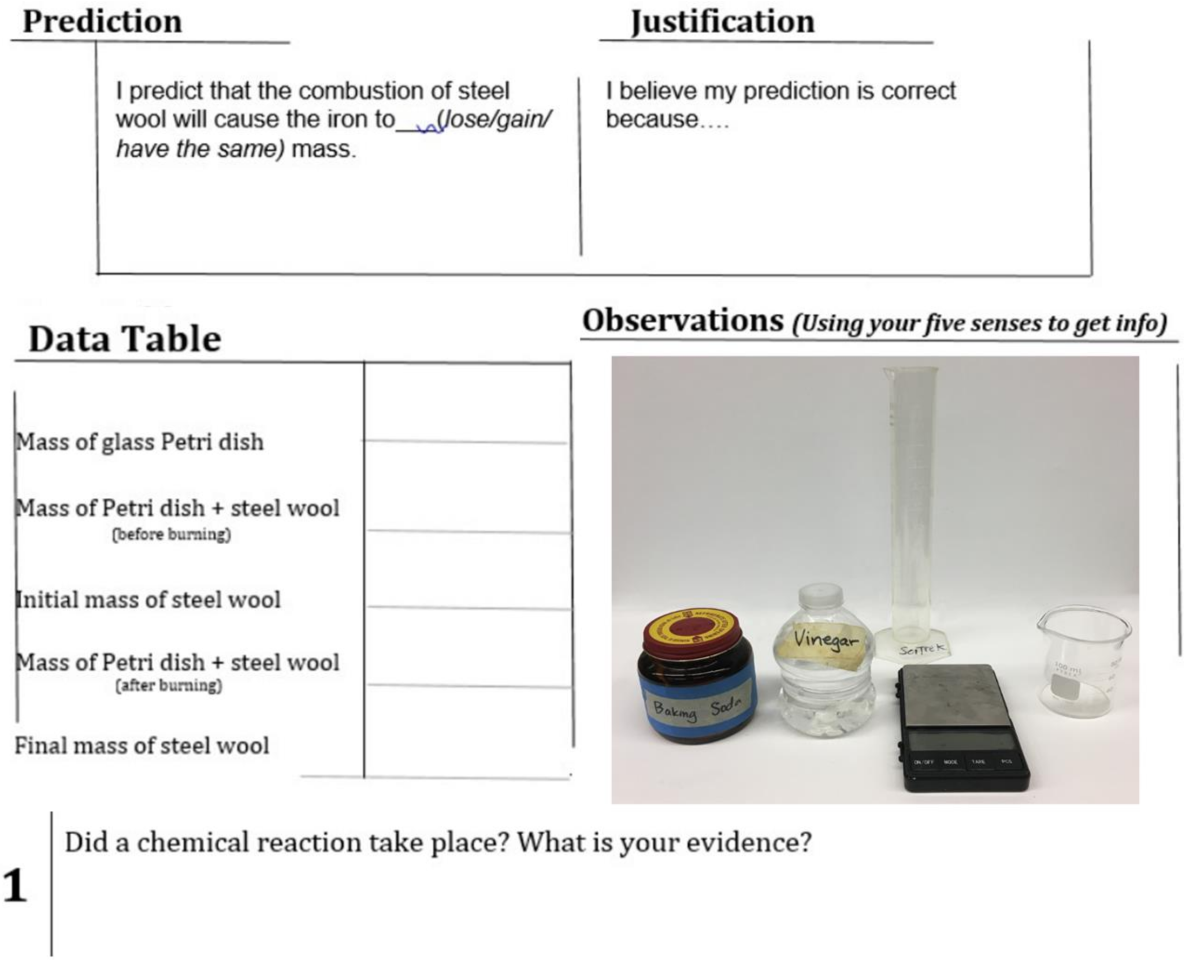


Figure 3: Conservation of Mass Module

(see Supplemental Files 4-6 for corresponding complete notebooks). On Day 1, after observing the burning marshmallow activity, students are provided steel wool and asked to formulate and justify a hypothesis on whether the steel wool will gain or lose mass, or remain unchanged after being ignited. Students are asked to mix baking soda and vinegar together on Day 3 of the Conservation of Mass module. Students will be tasked with constructing a closed system around this reaction on Days 4 and 5.

The organization of the two 7¬th grade modules that make up the study (Best Bread and Conservation of Mass) are representative of other secondary modules (e.g., the 8th grade modules Waves and Germs). The Best Bread module aims to cover Next Generation Science Standards, MS-LS1-5 and MS-LS1-1 [22]. MS-LS-5 covers the construction of scientific explanations based on evidence for how environmental and genetic factors influence the growth of organisms. In particular, it asks students to construct a scientific explanation based on evidence obtained from (among other sources) their own experiments. MS-LS-1 performance expectations ask students to investigate to provide evidence that living things are made of cells. MS-LS1 and 5 cover cellular respiration as well as understanding how environmental changes may affect an organism's ability to function. The module introduces students to the concept of cellular respiration and has them change environmental variables in a flask containing yeast, sugar, and water (Figures 2A- 2E). Students record quantitative results by measuring the amount of gas produced via water displacement. Students participating in this module learn the fundamentals of experimental design, which include forming a hypothesis as well as only changing one variable at a time (Figure 2A, 2E). The second module, Conservation of Mass, addresses NGSS MS-PS1-5 "Matter and its Interactions". It introduces students to a series of reactions that explore whether mass is lost, gained, or conserved during various physical changes and chemical reactions. Students ultimately are taught the concept of open and closed systems, and are asked to design a closed system where a baking soda and vinegar reaction takes place without any loss of mass (Conservation of Mass Notebooks available as Supplemental Files 4-6). Both modules encourage students to explore the interactions of changing variables and the value of experiments that either don't lead to a logical conclusion or disagree with a student's expected outcome; small group discussions, led by the university volunteers are used to emphasize the importance of multiple trials.

Due to scheduling constraints, SciTrek runs the modules either on Tuesday and Thursday or Monday, Wednesday and Fridays for six days. Teachers are provided "off day" activities focused on improving the subjects covered during the module such as graphing and mathematics if they choose to use them. To fit into the 7th grade curriculum, the Best Bread module is run in the fall and the Conservation of Mass module is run in the winter/spring. As in all SciTrek modules, students are provided equipment and workbooks to complete their activities and record data during each module.

The day-to-day activities for the Best Bread module (Figure 1) and for the Conservation of Mass Module (Supplemental Files 6,7) highlight diverse formats used in these and other modules. These include presentations by a "lead" (classroom Teacher, or a SRG member). More often, the activities occur at one table composed of four-six 7th grade students and one university volunteer. The students are guided by the volunteer at their table or the lead. Each student has their own notebook which has questions they need to answer related to the activities they either observe or carry out themselves. The volunteers are trained to be a resource for the students [48, 49]. Groups of students are frequently asked to report to the entire class as to their results, or answers to questions posed in their workbooks (Figure 1). At the end of the module, each student group summarizes their results by creating a poster that highlights their hypothesis and includes relevant graphs, tables, and a conclusion. The presentation to the class involves questions and response from the entire class of the presenting group.

Assessments (See Supplemental File 8)

The goal of the assessments is to test students' understanding of critical thinking concepts stressed by the SciTrek 7th grade module curriculum—Experimental Design, Observation/Inference, and Testable Questions, within the context of the activities covered during the modules. While each of these concepts has multiple components, for the purpose of these two 7th grade modules Experimental design focused on experimental error analysis, and observation/inference focused on the distinction between simple observations and making logical conclusions from evidence. Testable questions focus on differentiating between what can or cannot be addressed by collecting evidence. The assessment design also coincides with the Three-Dimensional Learning concepts promoted in the Next Generation Science Standards [22].

The data collected in this study comes from assessments administered in the 2018-2019 7th grade academic school year between two participating teachers at Santa Barbara Junior High School in Santa Barbara, California. The teachers both have over fifteen years' experience in teaching the 7th grade. The assessments consisted of seven questions, some with multiple parts, worth a total of 15 points. Two questions were not counted due to technical issues, resulting in a total of 13 points possible. Students were provided 10 minutes to complete the assessment and were informed that their score will not affect their academic grade. The same assessment was distributed twice during the school year, once prior to attending any SciTrek modules, and once after both modules have been completed.

The assessments were individually coded and graded separately by SRG members who had no knowledge of the student identity, whether the assessment was a pre- or post-module assessment, and if the student was in an Honors or College Prep class. The final scores for each question and sub-question (4a, 4b, etc) were recorded and the average pre and post assessment scores for each question were generated. The odds ratio, 95% confidence intervals, and p-values were generated by McNemar's test, using the function mcnemar.exact in R package exact 2x2. The results and analysis from each individual question are provided in Table 2.

Results

Our study involved two experienced 7th grade teachers at Santa Barbara Junior High School, both of whom had participated with SciTrek previously. The Santa Barbara Junior High School student demographics are representative of schools in the Santa Barbara School District, with two thirds of the students being socio-economically disadvantaged, and nearly 70% minority. Students are placed in College Preparatory (CP) or Honors (H) classrooms by the school based on teachers' recommendations and reassessed during the school year by the teacher. The SciTrek program was run the same way in CP and H classrooms. The distribution of College Preparatory and Honors students between the two teachers is summarized in Table I.

Teacher	Teacher # 1	Teacher # 2
Total Students	110	96
Total in College Preparation	55	29
Total in Honors	55	67

Table 1: Summary Table of Participants in this Study

Data includes the number of assessed students. Assignment of students is done by the school and guided by performance relative to others in the class at the discretion of the teacher.

Questions	Pre correct probability	Post correct probability	odds ratios	95% confidence intervals	p-values
1a	0.44	0.57	4.88	(2.10, 13.36)	<0.0001
1b	0.46	0.54	2.13	(1.12, 4.24)	0.0186
2	0.55	0.81	5.57	(2.95, 11.50)	<0.0001
3	0.53	0.89	13.17	(5.79, 36.96)	<0.0001
4a	0.91	0.97	3.20	(1.12, 11.17)	0.0266
4b	0.70	0.82	2.33	(1.31, 4.31)	0.0027
4d	0.75	0.85	2.50	(1.31, 5.03)	0.0038

5a	0.80	0.92	3.18	(1.58, 6.95)	0.0005
5b	0.54	0.77	4.00	(2.29, 7.41)	<0.0001
5c	0.80	0.89	2.46	(1.26, 5.11)	0.0066
5d	0.86	0.90	1.57	(0.77, 3.32)	0.2433
6	0.70	0.77	1.66	(0.96, 2.92)	0.0815
7	0.54	0.65	1.87	(1.16, 3.10)	0.0242

Table 2: Students Show Significant Improvements in all but One Question in all Areas of Science Process Skills Targeted in these Modules

Each type of question is labeled with their corresponding questions in the table. Experimental design is labeled with red, observation/inference is labeled with blue, and testable questions are labeled with green. The assessment and rubric are provided (Supplemental File 7). Odds ratios were determined by comparing the proportion of students who had answered correctly on the pre-assessment with the proportion of students who had answered correctly on the post assessment. 95% confidence intervals and p values were calculated based on the exact McNemar's test. Note, questions 4C and 5E were removed from the study due to technical problems.

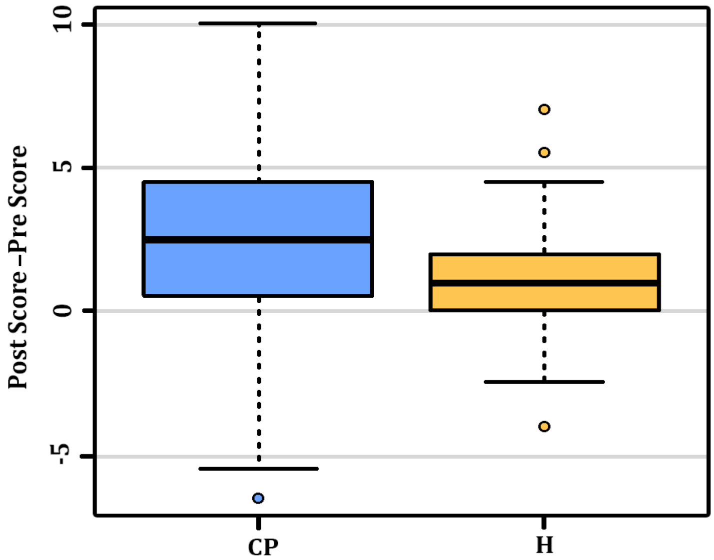


Figure 4: College Prep (Cp) Students show Greater Improvements in the Scitrek Critical Thinking Assessment Compared to Honors (H) Students

Values are derived from the mean of student performance, with bolded lines representing median values; the assessments had a total of 13 points possible. 1st and 3rd quartiles are represented by the extent of the colored boxes. The horizontal lines at the extreme ends of the plot represent the range of the sample, while the isolated circles represent outliers, determined by multiplying the Interquartile Range (IQR) by 1.5 above or below the 3rd and 1st Quartile, respectively. Median improvements were observed in post assessment scores compared to pre-assessment scores for both CP and H students, with greater median improvement in CP students compared to H students.

Figure 4 shows that students in both CP and H classrooms improved significantly on the post assessment compared to the pre-assessment. CP students improved 36% in overall performance while H students also improved, albeit only 12.8% (Figure 4). However, this trend is not reflected with both Teacher 1 and Teacher 2 students (Figure 5 and Tables 2,3).

Questions	Post mean scores	Pre mean scores	Mean difference (post-pre)	% improvement	95% confidence intervals	p-values
All	10.35	8.59	1.76	20.5	(1.43, 2.09)	<0.0001
All-Honors (H)	10.89	9.66	1.23	12.77	(0.92, 1.54)	<0.0001
All-College Prep (CP)	9.57	7.04	2.53	35.96	(1.89, 3.17)	<0.0001
All-Teacher 1	8.60	7.61	0.99	13.07	(0.53, 1.46)	<0.0001
All-Teacher 2	11.88	9.45	2.43	25.75	(2.01, 2.86)	<0.0001
Teacher 1 H	12.1	10.83	1.27	11.76	(0.84, 1.71)	<0.0001
Teacher 1 CP	11.65	8.06	3.59	44.53	(2.99, 4.19)	<0.0001

Teacher 2 H	9.90	8.70	1.20	12.13	(0.75, 1.65)	<0.0001
Teacher 2 CP	5.60	5.09	0.52	10.17	(-0.67, 1.70)	0.3784
Experiment Design	2.54	2.14	0.40	18.62	(0.24, 0.56)	<0.0001
Distinguishing observations from inferences	4.25	3.70	0.55	14.95	(0.36, 0.75)	<0.0001
Identifying testable questions	2.64	2.37	0.27	11.47	(0.15, 0.40)	0.0003

Table 3: Students in Cp Classrooms Improved More than those in Honors Classrooms, and all Students Show Significant Improvements in Experimental Design, Distinguishing Between Observations and Inferences, and Identifying Testable Questions

Average scores of post and pretests, and differences between post and pre-scores with 95% confidence intervals. The percentage of improvements over the pre-scores are also shown.

In fact, CP students with Teacher 2 showed smaller improvements (10.17%) than the H students (12.13%). This contrasts with the students in the classroom of Teacher 1, which showed 44.53% (CP) and 11.76% (H) improvements. The similar improvements with both teachers in the five H classrooms suggest this level of improvement is a reliable representation of what can be expected. In contrast, the CP students show much higher variability and teacher (or classroom)-dependent variation in performance. Regardless, the data in Figures 4 and 5 provide compelling data showing significant improvement in student understanding of the critical thinking parameters (experimental design, distinguishing between observations and inference, and identifying testable questions, Figures 5 and 6). Moreover, Figure 5 clearly shows that students of both teachers improved.

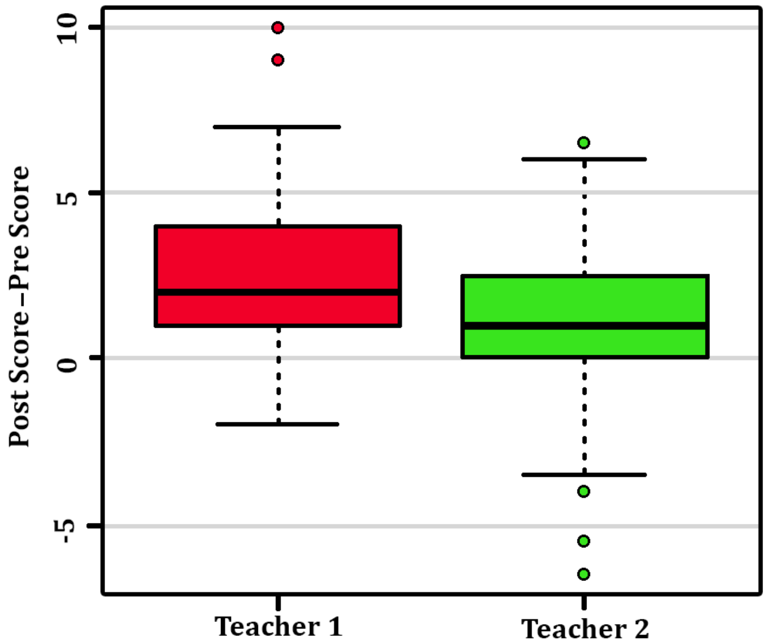


Figure 5: Students Show Improvements in the SciTrek Critical Thinking Assessment Independent of Instructor

Values are derived from the mean of student performance, with bolded lines representing median values; the assessment had a total of 13 points. 1st and 3rd quartiles are represented by the extent of the colored boxes. The horizontal lines at the extreme ends of the plot represent the range of the sample, while the isolated circles represent outliers, determined by multiplying the Interquartile Range (IQR) by 1.5 above or below the 3rd and 1st Quartile, respectively. The classes of the two teachers who participated in this study both showed statistically significant improvement in post assessment scores compared to pre-assessment scores.

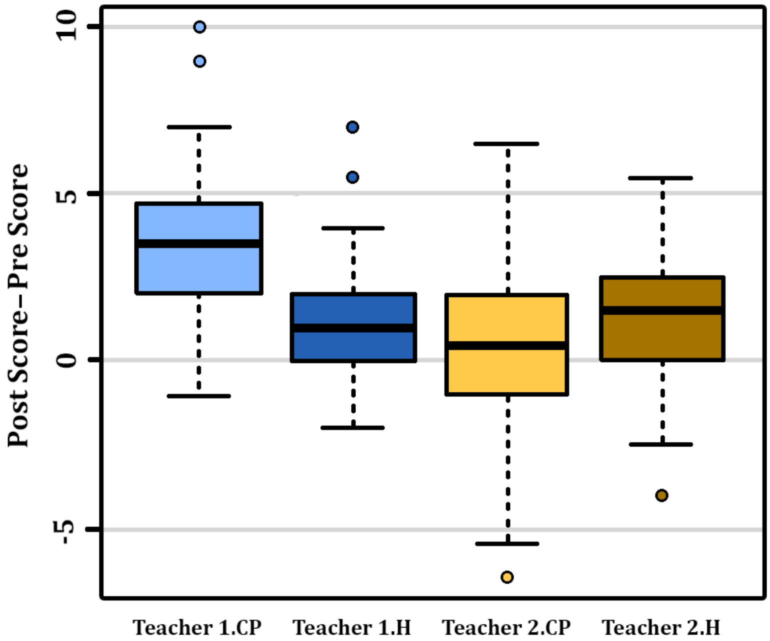


Figure 6: Overall Improvements in SciTrek Critical Thinking Assessment Specific to Instructor and College Prep (CP) and Honors (H) Standing

Values are derived from the mean of student performance, with bolded lines representing median values (total points possible on the assessment is 13). 1st and 3rd quartiles are represented by the extent of the colored boxes. The horizontal lines at the extreme ends of the plot represent the range of the sample, while the isolated circles represent outliers, determined by 1.5 * Interquartile Range (IQR) above or below the 3rd and 1st Quartile, respectively. Median improvements were observed in all post assessment scores among CP and H students separated by teacher. Only one category, Teacher 2 CP observed a 3rd quartile post-pre-assessment score <0.

Figure 6 shows that the high variability of student performance (CP, Figure 4) is observed with both teachers. And, that the lower variability for H students (Figure 4) is reflected in the performance of these students with both teachers. Importantly the latter reflect data from five classrooms (122 students). Clearly, the high variability (Figure 6) and very different improvements between teachers (T1, 3.59 out of 13, T2, 0.52 out of 13, Table 3) implicate teacher and student composition parameters. Interestingly, the pre-module performance of the students in T2 (CP, 5.09 mean, Figure 6) is by far the lowest mean of all groups, including the pre-module performance of students in T1 (CP, 8.06 mean, Figure 6). This may suggest variations in the ability of students in the CP classrooms which we are not aware of that impact both the initial performance, as well as potential for improvement resulting from participating in the SciTrek program.

Table 2 provides details about student performance on individual questions (Supplemental file 7) and the three critical thinking parameters which the two modules focus on (Experimental design, observation/inference, and testable questions). Questions 1a and b, 2, 6, and 7 are focused on experimental design, questions 3, 5A-D (5E was removed due to problems with question design) are focused on observations versus inference, and questions 4A, B and D are focused on testable questions (note, 4C was removed for technical reasons). With the exception of questions 5D and 6, all improvements are statistically significant. Questions 1 and 6 address a common problem for students, that in spite of being told about controls and limiting the number of changing variables, in practice they often do not display such understanding. Question 1 asks students to identify a poorly designed experiment and to redesign this to improve it. Question 2 asks students to determine which of two experimental outcomes is better; one has much greater variability than the other. The concepts of accuracy and precision, so core to science, are challenging to convey to students without relying on statistical analysis. Encouragingly, students improved dramatically on this question (mean pre-score of 0.55 and post score of 0.81 out of a possible score of 1), showing that this concept can be addressed without complex analysis. The design aspects of question 7 (picking the best basketball players), while somewhat open ended, asks students to first determine the flaw in the proposed method, followed by the more open ended second part, to propose an improved approach. Again, the overall mean improvement of 11% is significant.

The distinction between observations, inferences and opinions is important for students to understand, whether presented in the context of a SciTrek science module, or more broadly. Questions 3 and 5A-D address this, and with the exception of question 5D, student performance improved significantly on these questions. Question 3, which showed a very large improvement of 68% (Table 2), asks students to determine what types of procedural problems could lead to a flawed experimental result. In the context of the SciTrek program, improving students' ability to identify a flawed experimental design is arguably more important and perhaps teachable than being able to design an appropriate experiment. The latter requires both understanding as well as the imagination to create the needed experiment, whereas the former simply requires an understanding of the correct practices.

It is important for students to appreciate that not all questions can be addressed simply by collecting data, a concept referred to in SciTrek documents as "testable questions" [60]. Question 4 (parts A, B and D) tries to address this by posing various situations ranging from opinions (A, preferred color) to a quantifiable question (D, are you above average in height)? The improvements for this topic, although statistically significant (Table 2) are relatively small, ranging from 8.8% (4A) to 17% (4D). This topic is inherently subtle, perhaps even subjective. However, understanding what is testable is clearly linked to envisioning how something could be tested. Thus, while a challenge to devise suitable questions for this topic, we remain committed to both improving the module activities focused on this, as well as the assessments that measure how students perform.

Discussion

Critical thinking in the context of science practices, which is the focus of our study, can be challenging to teach as well as to assess [8,14,34,37-39,46]. Even the idea of a validated means to assess critical thinking has generated diverse opinions [61]. Similarly, how to best teach such skills remains unclear [8,14,34,37-39,46]. The SciTrek 7th grade program described here attempts to address both of these, but is clearly a work in progress. Notably, there was no attempt to "teach to the test" for these skills, so improvements in our assessments of these skills should reflect real changes. The value of this program to teach the content covered by these two modules was not assessed in our study.

As importantly, it is challenging (unresolved) to identify which features of this study are responsible for the observed improvements in critical thinking as applied to science practices. Thus, is the involvement of small groups led by university students over the ten to sixteen hours of in class activities most important? The other features, namely the 7th grade students' active engagement in developing their own questions and experimental design, or the arguing from evidence, would seem to be a less distinctive feature since many programs of varying levels of inquiry emphasis, make use of such activities. Similarly, one could argue that it is the distinctive features of the module activities themselves that play an important role. This study cannot evaluate which part of the SciTrek approach is most important, which is the subject of ongoing studies in our program.

Overall, the results are very encouraging, showing significant improvements in the three critical thinking parameters. Improvements in experimental design, observation/inference, and testable questions are observed and in all cases the improvement is supported by more than one question in the assessment. Rather than developing lesson plans that attempt to directly teach the students the critical thinking skills, the SciTrek program relies heavily on repeatedly showing students examples and encouraging discussions on those topics. For example, for the Best Bread module, students are presented at the start with the "big question" to be investigated, "How can we maximize the amount of carbon dioxide produced by yeast"? Students discuss alternatives to this question which may or may not be testable. Students are reminded throughout the module how their efforts relate to this testable question [46].

Observations of the 7th grade students involved in this study suggests that the active, guided participation of these students in carrying out experiments, many of which fail to generate a definitive conclusion, is important to their understanding of what is often presented as the "scientific method". While this anecdotal information is not as compelling as the statistical evidence provided in this study, the frequent comments from these students such as "I learned that failure can be helpful" will hopefully have longer term consequences than simply memorizing the scientific method.

Students' ability to identify what can be addressed through experimentation (Testable Question, Questions 4A, B and D, Table II) is a foundational feature of science investigations [60]. For example, a long-term goal of the entire SciTrek program is to improve the understanding of K-12 students' of when a study of any kind is done well (NGSS, MS-LS-1, 5). Significant improvement in all three questions is thus encouraging. Similarly, experimental design (Questions 1A, 1B, 2, 6 and 7) naturally follows from understanding if a question can be addressed experimentally. The ability of students to both judge an experimental design, and to propose a design, feature prominently in many aspects of NGSS literature (e.g., MS-LS-1, MS-LS-1, and MS PS-5). Students showed significant improvements in four of the five questions covering this topic. The third parameter covered in the assessment, distinguishing between observations and inferences, while challenging for students, is clearly an essential aspect of doing science. Five questions covered this topic (3, 5A-D), and all but 5D showed significant improvements.

While encouraging, our results leave many unanswered questions. The scope of the program was developed in consultation with participating teachers, which was important in the decision to have two modules per grade level, which is currently done in all participating classrooms throughout the SciTrek program. The in-class format, relying on five to eight hours per module is a compromise between the teachers' curriculum plans, the time needed for 7th grade students to carry out multiple rounds of experimentation along with demonstrations to initiate each module, and the recruitment and training of university students (ranging from four to six per classroom, or 300-500 per year for the entire program). Certainly, a critical part of this was whether students' understanding of the critical thinking concepts can be changed with this format and level of engagement. A primary focus of this study was to address this latter point. Ongoing efforts are focused on determining the impact of student participation in multiple years of the SciTrek program, on critical thinking performance, as well as choices these students make regarding continuing in STEM courses in their secondary education.

Of primary interest is whether the SciTrek approach provides a practical and cost-effective means to impact science education beyond the program at the University of California at Santa Barbara, a concern raised in the introduction regarding many effective programs. The engagement of university students is an important feature of the program, and we are currently working with schools and colleges/universities in regions (counties) outside of Santa Barbara County to determine if the program can work "at a distance". This was in part motivated by the ongoing Covid19 epidemic, but this too remains of keen interest.

Room for Improvement

The involvement of control groups (students not participating in SciTrek) is important for future studies, but can be challenging as many schools do not provide such an opportunity; using students in other schools in the district is possible, but has its own challenges. The development of assessments that are better aligned with the content and activities of the modules is also challenging; for example, the performance of students in the pre-module assessments on questions 4A, 5C and 5D are 80% or higher, making it problematic to identify improvements in the post-module assessments (Table 2). Another area requiring further work is to remove or minimize issues related to language, both in the running of the module (written notebooks, instructions from those running the module etc.), as well as in the assessments.

Ethics Statement

The UCSB IRB has authorized this study in accordance with the UC guidelines.

Informed Consent

A Consent to Participate and Consent to Publish) were obtained from the parent or legal guardians.

References

1. Alberts, B. (2009). Redefining science education. *Science, 323*(5913), 437-437.
2. Atkinson, R. D., & Mayo, M. J. (2010). Refueling the US innovation economy: Fresh approaches to science, technology, engineering and mathematics (STEM) education. *The Information Technology & Innovation Foundation, Forthcoming*.
3. Blank, R. K. (2013). Science instructional time is declining in elementary schools: What are the implications for student achievement and closing the gap?. *Science Education, 97*(6), 830-847.
4. DeSilver, D. (2017). US students' academic achievement still lags that of their peers in many other countries. *Pew Research Center, 15*.
5. Foltz, L. G., Gannon, S., & Kirschmann, S. L. (2014). Factors that contribute to the persistence of minority students in STEM Fields. *Planning for Higher Education, 42*(4), 1-13.
6. Hossain, M. M., & G Robinson, M. (2012). How to motivate US students to pursue STEM (science, technology, engineering and mathematics) careers. *Online Submission*.
7. Hughes, B. E. (2018). Coming out in STEM: Factors affecting retention of sexual minority STEM students. *Science advances, 4*(3), eaao6373.
8. Kafka, A. L., Ebel, J. E., Barnett, M., Macherides-Moulis, A., Campbell, L., & Gordon, S. E. (2006). Classroom seismographs and the challenge of encouraging a culture of scientific inquiry in K–12 schools. *Seismological Research Letters, 77*(6), 734-738.
9. Clementson, J. J. (1991). *A qualitative study of elementary teachers' and students' interactions with Full Option Science System: FOSS*. The University of Nebraska-Lincoln.
10. Foster, K. M., Bergin, K. B., McKenna, A. F., Millard, D. L., Perez, L. C., Prival, J. T., ... & Hamos, J. E. (2010). Partnerships for STEM education. *Science, 329*(5994), 906-907.
11. 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, 111*(23), 8410-8415.
12. Haywood, B. K., & Besley, J. C. (2014). Education, outreach, and inclusive engagement: Towards integrated indicators of successful program outcomes in participatory science. *Public understanding of science, 23*(1), 92-106.
13. Illingworth, S. M., & Roop, H. A. (2015). Developing key skills as a science communicator: Case studies of two scientist-led outreach programmes. *Geosciences, 5*(1), 2-14.
14. Isik-Ercan, Z. (2020). 'You have 25 kids playing around!': learning to implement inquiry-based science learning in an urban second-grade classroom. *International Journal of Science Education, 42*(3), 329-349.
15. Koehler, B. G., Park, L. Y., & Kaplan, L. J. (1999). Science for kids outreach programs: College students teaching science to elementary students and their parents. *Journal of Chemical Education, 76*(11), 1505.
16. Laursen, S., Liston, C., Thiry, H., & Graf, J. (2007). What good is a scientist in the classroom? Participant outcomes and program design features for a short-duration science outreach intervention in K–12 classrooms. *CBE—Life Sciences Education, 6*(1), 49-64.
17. Robardey, C. P., Allard, D. W., & Brown, D. M. (1994). An assessment of the effectiveness of full option science system training for third-through sixth-grade teachers. *Journal of Elementary Science Education, 6*(1), 17-29.

18. Tarjan, L. M., Nesnera, K. D., & Hoffman, R. (2015). Authentic science investigation in the classroom: Tools for creating original, testable questions and graphical hypotheses. *Science Scope, 39*(4), 42-48.
19. Williams-Pierce, C. C. (2011). Five key ingredients for improving student motivation.
20. Hoyer, K. M., & Sparks, D. (2017). Instructional Time for Third-and Eighth-Graders in Public and Private Schools: School Year 2011-12. Stats in Brief. NCES 2017-076. *National Center for Education Statistics*.
21. Nowicki, B. L., Sullivan-Watts, B., Shim, M. K., Young, B., & Pockalny, R. (2013). Factors influencing science content accuracy in elementary inquiry science lessons. *Research in Science Education, 43*, 1135-1154.
22. Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of science teacher education, 25*(2), 211-221.
23. National Center for Science and Engineering Statistics. (2019). Women, Minorities, and Persons with Disabilities in Science and Engineering. Retrieved from.
24. Noonan, R. (2017). STEM Jobs: 2017 Update. U.S. Department of Commerce Economics and Statistics Administration Office of the Chief Economist. Retrieved from.
25. Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy insights from the behavioral and brain sciences, 3*(2), 220-227.
26. Painter, J. (2011). Autonomy, competence, and intrinsic motivation in science education: A self-determination theory perspective.
27. Palmer, R. T., Maramba, D. C., & Dancy II, T. E. (2011). A qualitative investigation of factors promoting the retention and persistence of students of color in STEM. *Journal of Negro Education, 80*(4), 491-504.
28. Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of research in science teaching, 49*(4), 515-537.
29. Syed, M., & Chemers, M. M. (2011). Ethnic minorities and women in STEM: Casting a wide net to address a persistent social problem. *Journal of Social Issues, 67*(3), 435-441.
30. Yanowitz, K. L., & Vanderpool, S. S. (2004). Assessing girls' reactions to science workshops. *Journal of Science Education and Technology, 13*, 353-359.
31. Ertmer, P. A., & Simons, K. D. (2006). Jumping the PBL implementation hurdle: Supporting the efforts of K–12 teachers. *Interdisciplinary Journal of Problem-based learning, 1*(1), 40-54.
32. Ann Haefner, L., & Zembal-Saul, C. (2004). Learning by doing? Prospective elementary teachers' developing understandings of scientific inquiry and science teaching and learning. *International Journal of Science Education, 26*(13), 1653-1674.
33. Markowitz, D. G. (2004). Evaluation of the long-term impact of a university high school summer science program on students' interest and perceived abilities in science. *Journal of science education and technology, 13*, 395-407.
34. Metz, C. J., Downes, S., & Metz, M. J. (2018). The in's and out's of science outreach: assessment of an engaging new program. *Advances in physiology education, 42*(3), 487-492.
35. Duschl, R. A., Schweingruber, H. A., Shouse, A. W., & National Research Council. (2007). Committee on Science Learning Kindergarten Through Eighth Grade. *National Research Council (US)*.
36. Beatty, A. S., Koenig, J. A., Wilson, M. R., & Pellegrino, J. W. (Eds.). (2014). Developing assessments for the next generation science standards.
37. Nold, H. (2017). Using critical thinking teaching methods to increase student success: An action research project. *International Journal of teaching and learning in Higher Education, 29*(1), 17-32.
38. Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education, 93*(3), 223-231.
39. Quigley, C., Marshall, J. C., Deaton, C., Cook, M. P., & Padilla, M. (2011). Challenges to inquiry teaching and suggestions for how to meet them. *Science Educator, 20*(1), 55-61.
40. Buffington, C., Cerf, B., Jones, C., & Weinberg, B. A. (2016). STEM training and early career outcomes of female and male graduate students: Evidence from UMETRICS data linked to the 2010 census. *American Economic Review, 106*(5), 333-338.
41. Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., ... & Zavala, M. (2016). Improving underrepresented minority student persistence in STEM. *CBE—Life Sciences Education, 15*(3), es5.
42. Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., ... & Dilig, R. (2020). The Condition of Education 2020. NCES 2020-144. *National Center for Education Statistics*.
43. Ryan, T. G., & Whitman, A. (2013). The inequity and effect of standardized literacy testing for First Nations students-an Ontario (Canadian) outlook. *Journal of Educational Thought/Revue de la Pensée Educative,* 163-181.
44. Steenhof, N., Woods, N. N., Van Gerven, P. W., & Mylopoulos, M. (2019). Productive failure as an instructional approach to promote future learning. *Advances in Health Sciences Education, 24*, 739-749.
45. Helsel, R. T., Lambert, S., Dickerson, L., Strelich, J., Woods, V., & Feldwinn, D. (2022). Design of a phenomenon-based science outreach program and its effects on elementary students' epistemological understanding of, and attitudes toward, science. *School Science and Mathematics, 122*(2), 74-85.
46. Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science education, 94*(4), 577-616.
47. Roehrig, G. H., & Luft, J. A. (2004). Constraints experienced by beginning secondary science teachers in implementing scientific inquiry lessons. *International Journal of Science Education, 26*(1), 3-24.

48. Soysal, Y. (2024). Science teachers' challenging questions for encouraging students to think and speak in novel ways. *Science & Education, 33*(4), 963-1003.
49. Goodman, L and Berntson, G., [2000] The American Biology Teacher, Vol 7 pp 473-476.
50. Kuchynka, S. L., Eaton, A., & Rivera, L. M. (2022). Understanding and addressing gender-based inequities in STEM: Research synthesis and recommendations for US K-12 education. *Social issues and policy review, 16*(1), 252-288.
51. DeWitt, J., & Archer, L. (2015). Who aspires to a science career? A comparison of survey responses from primary and secondary school students. *International Journal of Science Education, 37*(13), 2170-2192.
52. Kessels, U., Heyder, A., Latsch, M., & Hannover, B. (2018). How gender differences in academic engagement relate to students' gender identity. In *Gender and educational achievement* (pp. 116-125). Routledge.
53. Maltese, A. V., & Tai, R. H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education, 32*(5), 669-685.
54. Jenkins, E. W., & Pell, R. G. (2006). "Me and the environmental challenges": A survey of English secondary school students' attitudes towards the environment. *International Journal of science education, 28*(7), 765-780.
55. Potvin, P., & Hasni, A. (2014). Analysis of the decline in interest towards school science and technology from grades 5 through 11. *Journal of Science Education and Technology, 23*, 784-802.
56. Master, A. (2021). Gender stereotypes influence children's STEM motivation. *Child Development Perspectives, 15*(3), 203-210.
57. Clayton, C. D., & Ardito, G. (2009). Teaching for ownership in the middle school science classroom: Towards practical inquiry in an age of accountability. *Middle Grades Research Journal, 4*(4), 53-79.
58. Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into practice, 41*(4), 226-232.
59. Raphael, J., Tobias, S., & Greenberg, R. (1999). Research experience as a component of science and mathematics teacher preparation. *Journal of Science Teacher Education, 10*(2), 147-158.
60. Graves, C., & Rutherford, S. (2012). Writing a scientific research ("testable") question: The first step in using online data sets for guided inquiry assignments. *Journal of College Science Teaching, 41*(4), 46-51.
61. Verburgh, A., François, S., Elen, J., & Janssen, R. (2013). The assessment of critical thinking critically assessed in higher education: A validation study of the CCTT and the HCTA. *Education Research International, 2013*(1), 198920.