Grounded Relevance: How Exploring Soil Data Can Promote Perceptions of Data Relevance, Data Literacy, Learning, and Motivation

Ian Thacker, Ph.D.

Rebecca Schroeder, Ph.D.

Sara Shields-Menard, Ph.D.

Sandrina Ramirez

Corina Lopez

University of Texas at San Antonio

Author Note

Ian Thacker https://orcid.org/0000-0002-2492-2929

Correspondence concerning this article should be addressed to Ian Thacker, Department of Educational Psychology, University of Texas at San Antonio, San Antonio, TX 78207. Email: ian.thacker@utsa.edu.

We have registered this study with the University of Texas at San Antonio Institutional Review Board (FY22-23-260).

We have no conflicts of interests to disclose.

Acknowledgements

This work is supported by the Hispanic Serving Institution Education Grant [2024-77040-43179/1032913], from the U.S. Department of Agriculture, National Institute of Food and Agriculture.

For Submission to the *Journal of Microbiology & Biology Education (JMBE)*

Note: This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. The final article will be available, upon publication, via its DOI (TBD).

Social Media Blurb.

What happens when students analyze authentic soil data? In our study with 298 undergraduates at an HSI, we found they gained microbiology knowledge, saw data science as offering insight and utility, and explained their ideas with greater confidence.

Abstract.

The ability to analyze, interpret, and draw insights from data and data visualizations is quickly becoming a necessary skill for success across multiple disciplines and careers. However, people struggle to make meaning from data, and traditional data science curriculum falls short of emphasizing its relevance to underrepresented students. To create opportunities for meaningful applications of data science for racially/ethically diverse students, we developed and implemented an online learning module focused on engaging N = 298 undergraduate students at a Hispanic Serving Institution (HSI) in an analysis of place-based soil data focused on identifying factors related to antibiotic producing bacteria. Using a pretest posttest study design, we found that students' perceptions of data science relevance (Cohen's *d*=.33) & microbiology knowledge (*d*=.39) significantly improved. We also inductively coded qualitative survey responses to explore how students framed "relevance" and how their perceptions changed from pretest-to-posttest; revealing that students perceived data tools as offering new scientific insights and applications across multiple disciplines, though changes over time were not significant. Automated text analysis of students' analytical thinking using LIWC (Linguistic Inquiry and Word Count) revealed that students tended to use significantly less "discrepant" (*d*=-.22) and "tentative" (*d*=-.19) language over time, suggesting that they improved their level of certainty from pretest to posttest. We also found that students identified more relevant and specific environmental- and soil-specific variables for data investigations at posttest compared to pretest. Overall, our findings provide evidence that interdisciplinary and place-based approaches to microbiology and data-literacy instruction support learning and motivation for racially/ethically diverse groups of students.

Keywords. Conceptual Change, Data Literacy, Integrated STEM, Place-Based Education

Grounded Relevance: How Exploring Soil Data Can Promote Perceptions of Data Relevance, Data Literacy, Learning, and Motivation

In today's world, data is ubiquitous and data literacy—decision-making through statistical methods and techniques—is essential across multiple industries (Ben-Zvi & Garfield, 2008; Gould, 2017; Steen, 2001). In microbiology and the life sciences, data literacy skills, such as the ability to interpret data visualizations, are foundational for making sense of evidence pertaining to complex biological systems, such as microbiological properties of soil, microbial ecology factors, high-throughput sequencing data, and mass spectrometry outputs (Dill-McFarland et al., 2021; Hahn et al., 2016; Hey et al., 2009, Rybarczyc et al., 2014). With the increasing amount of data available, the ability to analyze, interpret, and draw insights from data and data visualizations is becoming a necessary skill for many careers, likened to the importance of reading and mathematical literacy (Börner et al., 2019; Gal, 2002).

Despite the current high demand for data science education, traditional courses offered by statistics and computer science departments are not meeting the needs of those seeking training (Baumer, 2015), and statistics education, while firmly grounded in mathematics curriculum (Ben-Zvi & Garfield, 2008; National Council of Teachers of Mathematics, 2000), does not traditionally tap into topics that students find relevant, such as interdisciplinary and sociopolitical applications (Kokka, 2019; Weiland, 2017). With so few relevant applications that appeal to diverse audiences,[1] it is no surprise that there is a lack of diversity with regards to race/ethnicity and gender in math-intensive science, technology, engineering, and mathematics (STEM) fields (National Science Foundation [NSF], 2015). Hispanic students, in particular, are underrepresented in STEM fields, particularly those related to statistics and data science (Fry et al., 2021).

This interdisciplinary project aimed to enhance student data literacy, perceptions of data science relevance, and science learning through place-based soil data analysis at a Hispanic Serving Institution (HSI). We developed an online learning module that offers students opportunities to explore and interpret data visualizations of microbial content of soil data that they collected. We then investigated whether the intervention improved a racially/ethically diverse group of students' science learning and changes in students' qualitative descriptions of why data science might or might not be relevant for microbiology.

[1]We use the term "diverse" to refer to racial/ethnic and gender groups that have been historically underrepresented in STEM fields, including but not limited to Hispanic/Latino, Black/African American, Native American/Alaska Native, and women. In the context of this study, the term primarily reflects the racially and ethnically diverse student population served by the institution, particularly Hispanic/Latino students.

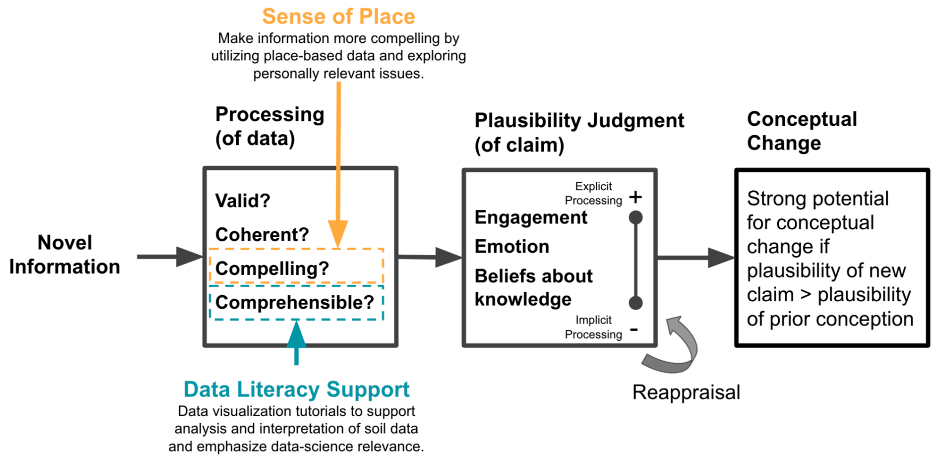
Theoretical Framework

To frame how soil data collection and analysis can support science learning and perceptions of data science relevance for underrepresented students, we integrate frameworks of *Conceptual Change, Data Literacy*, and *Place-Based Education*.

Conceptual Change

Conceptual change is a process where individuals restructure their conceptual knowledge to be more aligned with experts after engaging with novel information (Dole & Sinatra, 1998; Lombardi et al., 2016). For example, Dole & Sinatra's (1998) model of conceptual change takes into account information characteristics (e.g., whether soil data is *comprehensible*, *compelling*, and *relevant*) and learner characteristics (e.g., their *beliefs*, *motivation*, and *emotions*), which interact to determine students' levels of cognitive engagement. Higher levels of engagement, in concert with people's shifting motivational and emotional states, then predicts more serious consideration (or reconsiderations) of whether scientific ideas are *plausible* (Thacker & Herrick, 2025; Lombardi et al., 2016). More explicit plausibility considerations then predict a higher likelihood that people will exchange scientifically accepted ideas with their prior conceptions (i.e., experience conceptual change). Of these many factors involved in conceptual change, evidence suggests that data *comprehensibility* and *compellingness* can bolster student engagement and conceptual change (Thacker, 2023; 2024; Thacker et al., 2025; Thacker & Sinatra, 2022). In this project, we thus aimed to facilitate perceptions of engagement and conceptual change by promoting (a) compellingness of data by highlighting its relevance through place-based learning, and (b) comprehensibility through data literacy support (see Figure 1). We elaborate on these mechanisms below.

Figure 1.
Logic Model Synthesizing Conceptual Change, Place-Based Learning, and Data Literacy.



Data Literacy

Though there is no consensus on a definition, the term “data literacy” can be defined as the statistical competencies, methods, and techniques that facilitate decision-making (Gould, 2017). Because decision-making with data can be complex, data literacy frameworks often include many interrelated core competencies that range from about 5 to 20 different categories. Such data literacy frameworks often include competencies such as *understanding*, *acquiring, reading, interpreting, evaluating, managing, visualizing,* and *using* data (Börner et al., 2019; Carlson & Johnston, 2014; Kim et al., 2023; Prado & Marzal, 2013; Ridsdale et al., 2015); dimensions that are also critical for driving insight and advancement in fields of science more broadly (Qiao et al., 2024).

Of the many data literacy competencies identified in the literature, this study largely tries to understand how students frame problems and identify variables prior to acquiring data then interpret data visualizations in ways that address that problem. More specifically, a central data literacy process is *translating* relevant problems of interest into problems of data (Börner et al. 2019). That is, before even acquiring data, individuals must first understand how the data relates to a relevant situation or problem, and think analytically about how they might operationalize and measure variables of interest. As such, to improve students’ interpretation of data visualizations, they may need support in thinking analytically to operationalize real-world problems. A goal of this project was therefore to leverage students’ knowledge and experiences to help them contextualize, identify, and make meaning of relevant data and data visualizations.

Place-Based Education

A useful framework for contextualizing data and creating personally relevant data experiences for students is to connect to their “sense of place” (Gruenewald, 2003; Semken et al., 2017). Place-based learning is all about grounding science knowledge in an individual’s localized sense of place and the related systems of meaning that are defined by aesthetics, culture, history, and community orientation. Place-based learning is ideal for teaching topics such as microbiology, geoscience, and agricultural education because of their direct relation to one’s lived environment. Place-based learning can also be used as a means of culturally informed practices for inclusivity of racially/ethically diverse students and can be leveraged to foster pro-environmental attitudes (NGSS, 2013; Semken et al., 2017).

Although there are useful examples of place-based STEM education at the undergraduate level (Gosselin et al., 2016), leveraging place-based education to support racially/ethically diverse learners remains understudied. For example, in a review of the literature, Semken (2017) identified the need for researchers to (a) test the utility of Place-Based Education for racially/ethically diverse students, (b) study the role of emotion and motivation in place-based education, and (b) support student interpretation of data by teaching to reveal meaning and relationships rather than facts as analogous to place. And while there is long tradition of research focused on promoting inclusion of underrepresented students in STEM (e.g., Ladson-Billings; 1995; Borba, 1990), few if any research taps into students' sense of place as a means of promoting interdisciplinary STEM learning and data-literacy. In this project, we aim to address these research gaps by supporting place-based microbiology learning, motivation, and data literacy for racially/ethically diverse learners. We addressed the following research questions (RQs):

- **RQ1.** To what extent do student's perceptions of data science relevance and microbiology knowledge change after exploring place-based soil data and data visualizations?
- **RQ2.** How do students describe the relevance of data science tools for exploring the microbial content of soil? What levels of analytical thinking are evident in those descriptions? And how do these dimensions change after students explore place-based soil data and data visualizations?
- **RQ3.** What soil-related variables do students identify as being relevant for data exploration? And how do these perceptions change from pre- to post-intervention?

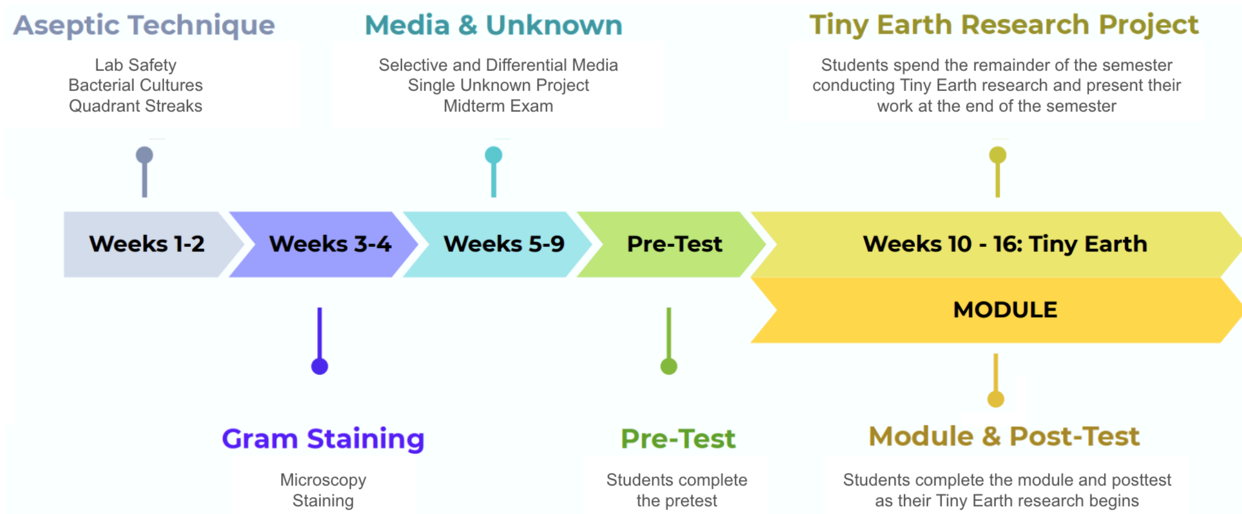
Methods

To address these research questions, we designed a learning intervention for place-based soil data exploration. We then used a pretest posttest research design that investigated changes in students' perceived data relevance and microbiology conceptions.

The intervention was embedded within a 16-week microbiology lab course that culminated in a research project aligned with Tiny Earth curriculum (Hurley et al., 2021; Miller et al., 2025). A timeline of course activities is presented in Figure 2. During weeks 1 and 2, students learned aseptic techniques, including lab safety procedures, bacterial culture handling, and quadrant streaking. In weeks 3 and 4, instruction focused on Gram staining, with students practicing microscopy and bacterial staining procedures. Weeks 5 through 9 were devoted to media and unknown identification, during which students learned to use selective and differential media and completed a single-unknown identification project alongside a midterm assessment. After this phase, students completed the study pretest prior to the start of week 10. Beginning in approximately week 10, students completed the focal intervention and embedded posttest (described in the next section) prior to initiating a Tiny Earth research project in which they collected soil samples, cultured and isolated microorganisms, screened isolates for antibiotic-producing activity, and characterized strains. Students presented their research projects and findings at the end of the course.

Figure 2.

Timeline of Activities in Traditional Microbiology Lab Course Leading Up to Module Implementation

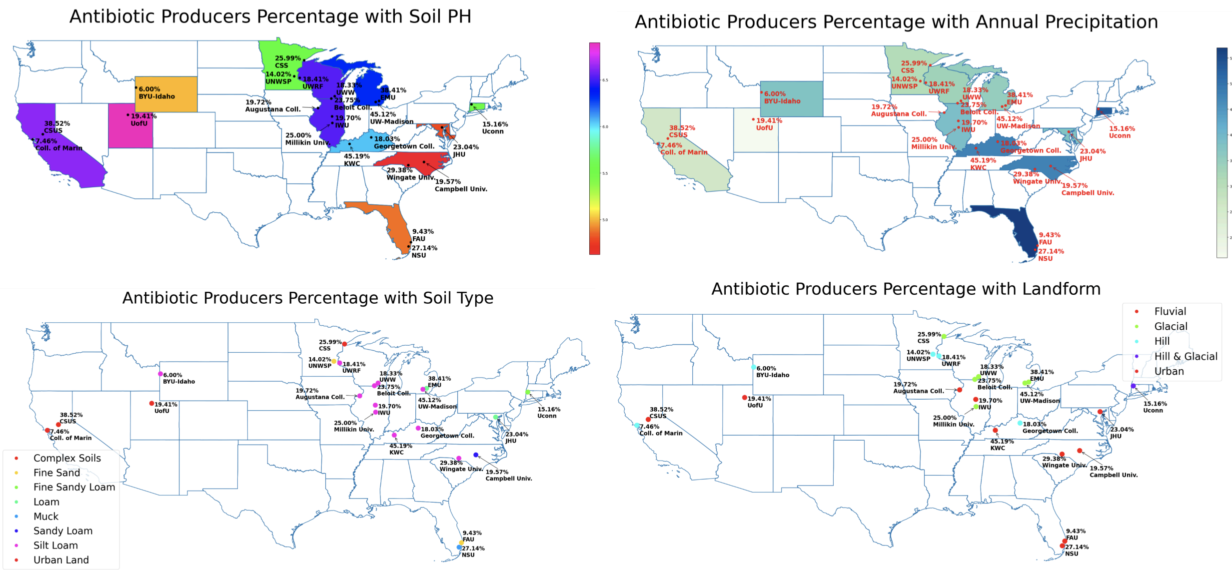


Intervention Developed

Prior to conducting the current study, we developed and tested an interdisciplinary data-literacy/microbiology learning intervention (see Thacker et al., 2025). The intervention is an asynchronous, open-access learning module developed in SoftChalk Cloud that introduces undergraduate students enrolled in a microbiology course to the Tiny Earth Initiative (Hurley et al., 2021; Miller et al., 2025), with a particular focus on introducing students to data science applications. Tiny Earth is a national initiative concentrated on identifying new antibiotic producing bacteria in the soil to combat the escalating antibiotic resistance crisis by encouraging undergraduate students to collect soil from the place where they live, study the bacteria in that soil, analyze the bacteria for antibiotic activity, and add their data to an online repository. In addition to introducing students to the goals and public data repository of Tiny Earth, the module introduced students to relevant information related to the antibiotic resistance crisis (the problem of diminishing effects of common antibiotics to stop bacterial infection), discussed microbial ecology factors that are important for creating soil conditions that harbor new antibiotic producing bacteria, prompted students to generate personally relevant research questions, provided a short tutorial on data visualization tools, and immersed students in soil data visualization interpretation. Students explored whether a selection of variables (landform type, soil type, annual precipitation, pH, and annual air temperature) were related to the percentage of antibiotic producers in a selection of map visualizations generated from the TinyEarth national repository data (see Figure 2 for examples).

Figure 2.

Examples of Data Visualizations that Students Used to Address Questions Students Generated



A course shell of the full module can be accessed using the following open access SoftChalk cloud link (https://www.softchalkcloud.com/lesson/serve/IEmc5RhkOGrA9n/html), and a summary of the modules and course content can be found in the supplemental materials (Supplemental Materials, Appendix A, Table A1).

Participants and Procedures

To answer our research questions, we recruited *N* = 298 undergraduate students from an HSI in a southwestern state of the USA over the course of three semesters. Students reported their year of study (1% first year, 13% second year, 38% third year, 38% fourth year, and 10% other), gender (75% Female, 22% Male, 1% Nonbinary, 2% prefer not to say), ethnicity (55% Hispanic), race (1% American Indian/Alaska Native, 12% Asian, 7% Black/African-American, 9% Two or more races, 62% White/Caucasian, 9% Other race), and whether they were enrolled in a STEM major (79% STEM major, 15% not STEM, 1% plan to enroll in a STEM major, 4% Other).

Survey Measures

At pretest, all participants completed a measure of their microbiology knowledge, data-literacy, data relevance perceptions. After the pretest, learners completed the ~60 minute module and then completed a posttest with identical measures of their microbiology knowledge and data literacy, data relevance perceptions, and a demographics questionnaire. All items can be found in Appendix B-D of the Supplemental Materials.

Microbiology Knowledge. The microbiology knowledge measure was created by our research team (Thacker et al., 2025) and consisted of a 12-item multiple choice questionnaire (e.g., "What is the source of most of the antibiotics in current use? (a) Chemical labs, (b) Plants, (c) Soil bacteria, (d) Water bacteria"). Each item was coded as either correct or incorrect, and the number of correct answers was averaged and converted to percentage points for all analyses. Cronbach's alpha for the microbiology knowledge scale was 0.65 at pretest and 0.71 at posttest.

The microbiology knowledge quiz questions were intended to assess the students' foundational understanding of antibiotic resistance, microbiological origins of antibiotics, and the research purpose and methodology of the Tiny Earth Initiative. These core concepts were explicitly introduced in the Module and are only later reinforced throughout the lab-based Tiny Earth research experience. We used identical questions in the post-test to evaluate student learning gains of these core concepts and increased familiarity of the scientific language central to the Tiny Earth curriculum after engaging with the module.

Data Literacy / Data Relevance Perceptions. This study focuses on specific, theory-informed components of data literacy that are especially relevant in introductory biology courses. In particular, we examine students' perceived relevance of data science, their ability to explain analytic reasoning when discussing data use, and their capacity to identify meaningful variables before engaging in data exploration. Collectively, these measures reflect proximal indicators of data-literate thinking that emerge prior to more advanced statistical or computational skills and align with developmental models of data literacy (Börner et al., 2019; Gould, 2017).

Namely, students completed a three-item data literacy measure. Two of the data literacy questions were open-ended and were used in qualitative analyses. The first open-ended item prompted students to "*Explain why you think that data science is or is not relevant for the field of soil microbiology. Provide examples if possible.*" The second item prompted students to, "*Please make a list of any variables that might be relevant for data scientists to investigate when exploring information about soil.*" The third item was a single item assessing students' perceptions of data science relevance *"In your opinion, are data science techniques relevant for addressing questions in soil microbiology?"* The response scale ranged from 1 = *Not at all relevant* to 5 = *Very relevant*.

The data relevance perceptions questions were designed to capture students' applied understanding of data science for generating insight in microbiology. In contrast, the microbiology knowledge / multiple choice questions assessed students' foundational microbiology knowledge independent of data science.

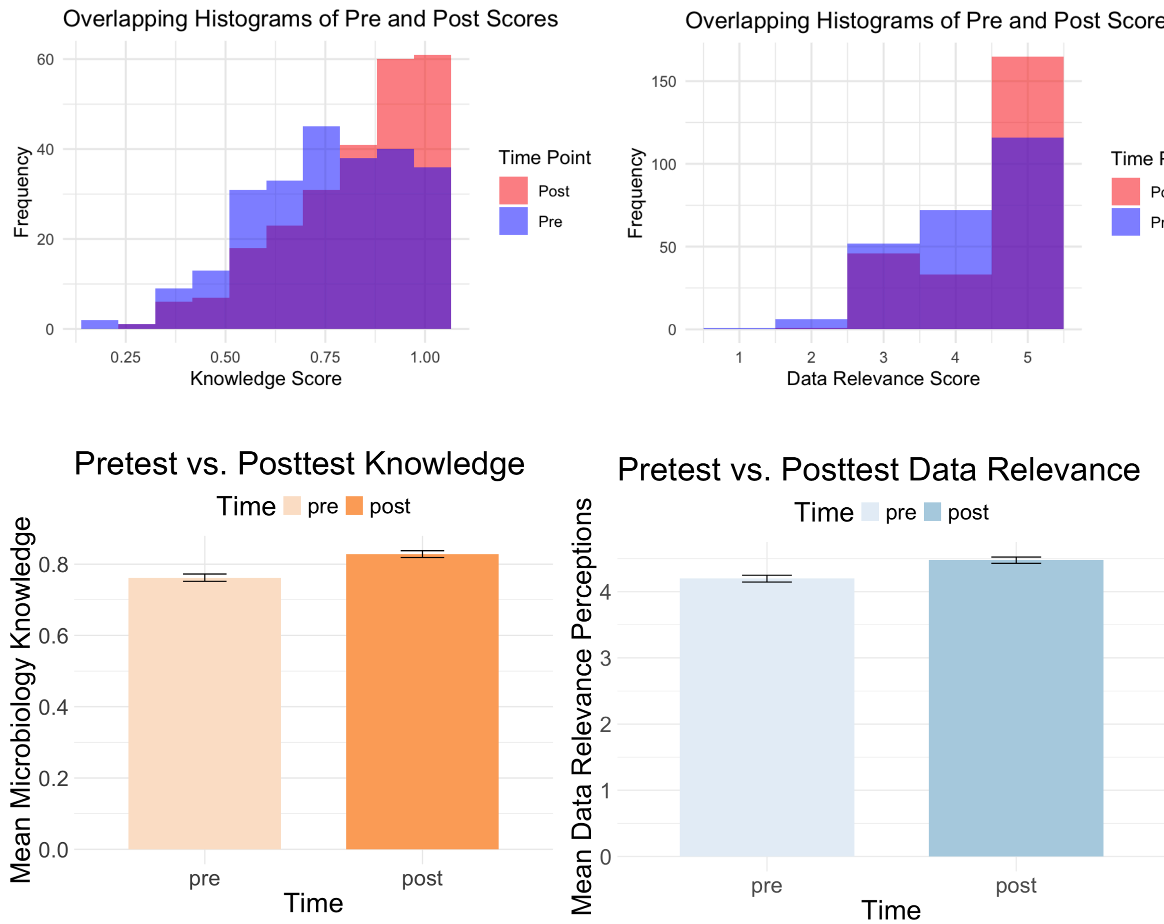
Findings

RQ1: Pre-Post Improvements in Data Science Relevance and Microbiology Knowledge

Prior to addressing RQ1, investigating whether there were significant changes in mean microbiology knowledge and data relevance perceptions, we first inspected histograms of continuous variables to assess the normality of the distributions (see Figure 3). Because the distributions were skewed, we used nonparametric Wilcoxon signed-rank tests. We found that students' microbiology knowledge significantly improved from pretest to posttest (M_{pre} = 76%, SD_{pre} = 17, M_{post} = 83%, SD_{post} = 16; W = 4326, $p < .001$, $d = .39$), as did their perceptions of data science relevance (M_{pre} = 4.2, SD_{pre} = 0.88, M_{post} = 4.5, SD_{post} = 0.81; W = 2425, $p < .001$, Cohen's $d = .33$). That is, we found that both knowledge and data science relevance perceptions increased by about a third of a standard deviation from pretest to posttest. Visualizations of distributions and means at pretest and posttest are shown in Figure 3.

Figure 3.

Pretest and Posttest Data Relevance and Microbiology Distributions and Means (N =298).



Note. Whiskers indicated in bar plots represent 1 standard error (SE) intervals.

RQ2: Qualitative Analyses of Student Perceptions of Data Science Relevance

To explore how students perceived the relevance of data science in soil microbiology analysis, we analyzed their open-ended responses in two ways. The first was by using inductive qualitative analysis, and the second involved an automated text analysis using the Linguistic Inquiry and Word Count (LIWC) software. We discuss these analyses and results next.

Inductive Analysis

The first analyses used traditional inductive coding processes. Namely, two graduate student researchers (fourth and fifth authors on this submission) openly-coded (Saldaña, 2021) all student responses to the question: "*Explain why you think that data science is or is not relevant for the field of soil microbiology…*" The coders read each student's response at pretest and posttest and categorized them freely. We then met to discuss common codes, emerging themes related to all student responses, and collectively developed a codebook (see Appendix B) which we used to systematically recode all responses (Charmaz, 2015; Saldaña, 2021). The two coders independently applied the codebook to the full dataset. Interrater reliability analysis revealed high percent agreement (average: 91%) but the kappa values were low due to infrequent code use and data sparsity. As such, coders met to reconcile discrepancies and arrived at consensus-codes which are reported for all analyses.

Four dimensions emerged: students perceived that data science offers: (a) *understanding*, it illuminates, explains, or provides insight into properties of soil by offering comparing and contrasting information or showing trends; (b) *utility* to address real-world problems in specific fields by using specific analytic techniques, helping to visualize problems and save time; (c) is *vaguely useful* or relevant with unclear reasoning for why, and (d) that data science is *not relevant* in their perspective. A summary of the results are presented in Table 1, and examples of student responses for each code are presented within the codebook (Appendix B).

Understanding is defined as the perception that data helps to illuminate, explain, or provide insight into properties of soil. This was identified among 100 students (33.6% of the full sample) at pretest and 103 (34.6%) at posttest. This was a parent code that was refined and split into two subcodes based on differences in how students described how data science enhances understanding; (a) that it offered new ways to *compare and contrast* information (e.g., "Data science is relevant to soil microbiology because it allows us to compare soil samples from various locations and soil over time"), and (b) that it offers *insight through analysis* by describing, analyzing, and identifying trends and patterns in data that offer researchers new insights (e.g., a student said that "Data science techniques, including statistical analysis, machine learning, and data visualization, can be used to analyze and interpret these datasets to extract meaningful insights about microbial communities, their interactions, and their impact on soil health"). We found that, from pretest to posttest, "compare/contrast" decreased in frequency (10.1% pre, 5.4% post) while "insight through analysis" increased (23.5% pre, 29.2% post).

Utility Perceptions is defined as the perception that data science can be useful for specific real-world tasks and professions. This was identified among 150 (50.3%) of students at pretest and 157 (52.7%) at posttest. This parent code was broken into five subcodes based on the different applications that students identified. First, students provided thoughts on (a) *field specific applications* (e.g., a student shared that "...data science is highly relevant in soil microbiology as it empowers researchers to make sense of large and complex datasets, model soil ecosystem processes, and apply findings to various fields like agriculture, environmental management, and climate change research"). Second, students noted (b) *specific data science techniques* that might be applied to study microbial processes (e.g., "Techniques like machine learning, predictive modeling, and data visualization enable researchers to identify patterns in microbial communities and predict factors influencing soil health or antibiotic production"). Third, students indicated (c) *efficiency* as a major perceived benefit of data science techniques (e.g., "Makes organizing vast amounts of data more efficient"). Fourth, students noted that *data presentations,* such as data summaries, visualizations, or written reports—are useful for supporting interpretation and communication of information (e.g., "Data science is very relevant to microbiology, data science allows us to test the bacteria and show people the percentage of antibiotic resistance"). Fifth, students noted that data science techniques are useful for *organizing data*, which includes managing, saving, identifying data (e.g., "Data science can organize and store important data such as the NCBI GenBank.data science …allows for the analysis of large samples which help identify patterns and microbial responses to environmental changes").

Vague was a code defined as instances where students identified data science as useful but without providing much explanation, information, or reasoning for why. This code was identified among 55 participants (18.5%) at pretest and 47 (15.8%) at posttest. This code was not broken into subcodes. Examples of vague descriptions include "I think that data can be used as a beginning point for the field of soil microbiology," or "[data science] is extremely relevant and is needed in the field of soil microbiology due to the amount of data and complexity of the data."

Not relevant is the final, and least frequent code, which was identified among 8 (2.7%) of students at pretest and 6 (2.0%) at posttest. This code is defined as instances where students did not identify data science techniques that are relevant for soil microbiology (e.g., student responses coded as not relevant include: "IDK," "N/A," or "Not relevant").

Table 1.

Frequency of Inductive Codes Indicating Students' Explanation for Why Data Science Is Relevant for Soil Microbiology, If At All.

		Pretest (*n* = 251)		**Posttest (*n* = 249)**	
Code	Subcode	# of students	% of students	# of students	% of students
Understanding	Total Across All Subcodes of "Understanding"	100	33.6%	103	34.6%
	Helps to compare/contrast	30	10.1%	16	5.4%
	Insight through analysis (by describing, showing trends, and patterns)	70	23.5%	87	29.2%
Utility Perceptions	Total Across All Subcodes of "Utility Perceptions"	150	50.3%	157	52.7%
	Field-specific uses	27	9.1%	15	5.0%
	Data-specific uses/techniques	36	12.1%	35	11.7%
	Efficiency (e.g., time saving)	15	5.0%	11	3.7%
	Data presentation	21	7.0%	32	10.7%
	Data-organization uses	51	17.1%	64	21.5%
Vague	Vague (but positive) response	55	18.5%	47	15.8%
Not Relevant	Does not see relevance	8	2.7%	6	2.0%

Note. For detailed code definitions and descriptions, see the codebook, Appendix B. Percentages calculated with a denominator of N = 298, the total students who completed the study. Categories are not mutually exclusive.

Automated Text Analysis

The second way in which we coded the qualitative data was using LIWC 2022 (Linguistic Inquiry and Word Count; Pennebaker, 2001) an automated text analysis program that assesses student responses in terms of percentages of total words in a text that map onto sets of validated "dictionaries" of words that represent psychological constructs (Boyd et al., 2022). Specifically, we assessed students' use of words pertaining to *analytical thinking* dimensions, which demonstrate cognitive skill and engagement (see, e.g., Markowitz, 2023; Pennebaker et al., 2014) at pretest and posttest and compared them. Across the five dimensions of analytic cognitive processes (insight, causation, discrepancy, tentativeness, certitude, and differentiation), Wilcoxon signed rank tests revealed significantly lower levels of "discrepancy" language at posttest ($W = 5295$; $p = .006$, $d = -.22$), discrepancy being words that indicate a difference between what is and what ought to be, such as "should," would," or "could." We also found significantly lower levels of "tentativeness" at posttest ($W = 2051$, $p = .006$, $d = -.19$), tentativeness being words that indicate uncertainty and hesitation such as “maybe,” “perhaps,” and “guess.” See Table 2 for means and SDs across analytic thinking.

Table 2.
Descriptive Statistics Related to Analytical Dimensions of Thinking Pretest and Posttest from LIWC.

	M	SD	Min	Median	Max	Skew	Kurtosis
insight.pre	9.42	7.46	0	8.2	50	1.87	5.93
insight.post	9.56	6.43	0	9.09	50	1.8	8.02
cause.pre	5.1	4.32	0	4.55	22.22	0.84	0.74
cause.post	4.81	4.45	0	4.5	26.67	1.3	3.28
discrepancy.pre	**2.53**	2.95	0	1.67	14.29	1.03	0.58
discrepancy.post	**1.9**	2.69	0	0	10	1.21	0.42
tentative.pre	**1.79**	3.54	0	0	33.33	3.87	25.39
tentative.post	**1.16**	2.94	0	0	25	3.82	20.21
certitude.pre	0.3	1.24	0	0	10	4.88	25.85
certitude.post	0.12	0.69	0	0	6.25	6.09	39.3

Note. Values are mean percentages of total words in student responses. Bolded pairs of means have significant pre-post differences at the .05 alpha level.

Synthesizing across our inductive qualitative analyses and automated qualitative analysis related to RQ1, we found that students largely viewed data science to be relevant, and used significantly more confident language in their explanations at posttest compared to pretest. Specifically, students wrote that data science can be useful for addressing pressing problems in microbiology and can offer new insights through organizing and communicating trends and comparisons in soil data. The language they used also demonstrated significantly lower levels of “discrepancy” and “tentativeness,” at posttest suggesting that they may have felt more confident in their responses.

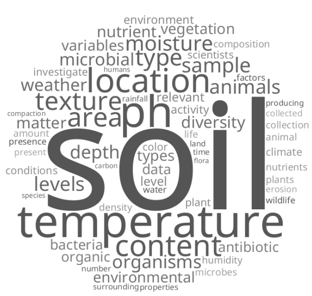
RQ3: Relevant Soil-Variables Identified

We also used LIWC to analyze the list of variables that students provided at pretest and posttest that they perceived would be relevant in data-exploration. Pretest and posttest word counts of listed variables are shown in Table 3, along with word clouds demonstrating frequencies, and differences in frequencies. Comparing pre and post, it is evident that students became more specific in their language: there was improved salience of the word “moisture,” “temperature, “pH,” and other terms at posttest compared to pretest, while “soil” remains prominent across both. In a comparative analysis between pretest-and-posttest that highlights words that are more associated with pre vs post, we see that “animals” and “organisms” are more prominently related to the pretest, whereas more chemical and landscape factors are more prominent in the posttest.

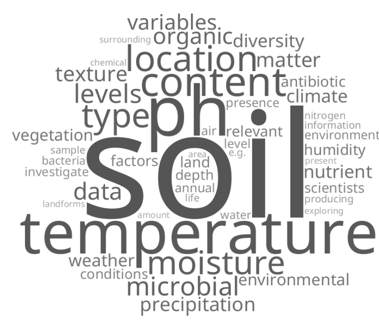
Table 3. *Pretest Posttest Comparison of Words Used When Students Listed Relevant Soil Variables*

PRETEST					POSTTEST				
Word	**Freq**	**Freq %**	**Students w/ word**	**% of Students with word**	**Word**	**Freq**	**Freq %**	**Students w/ word**	**% of Students w/ word**
soil	228	14.3	96	81.4	soil	161	12.8	66	74.2
temperature	56	3.5	56	47.5	pH	66	5.3	66	74.2
pH	54	3.3	54	45.8	temperature	56	4.5	56	62.9
location	39	2.4	37	31.4	content	35	2.8	19	21.3
content	34	2.1	20	16.9	location	30	2.4	29	32.6
area	29	1.8	20	16.9	type	28	2.2	27	30.3
texture	28	1.8	28	23.7	moisture	27	2.1	26	29.2
type	25	1.6	22	18.6	microbial	20	1.6	13	14.6
moisture	22	1.4	22	18.6	levels	20	1.6	16	18.0

Word Count Visualization

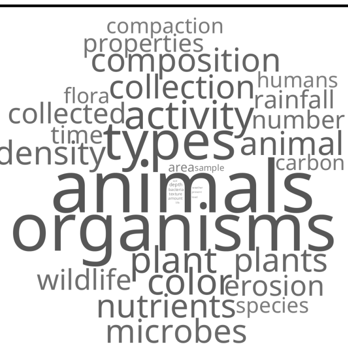


Pretest



Posttest

Word Count Differences Visualization
(i.e., Terms Associated More with the Pretest vs Posttest)



Pretest

Posttest

Note. Analyses in this table were conducted using LIWC 2022 (Linguistic Inquiry and Word Count). "Students w/ word" = The number of students who used the given word. "% of Students with word" = The percentage of students who used the given word.

Discussion

We sought to develop a learning intervention that promotes data-literacy skills as they apply to place-based soil microbiology learning experiences. Namely, undergraduate students engaged with an asynchronous learning module created to introduce students to information about the antibiotic resistance crisis, the goals of the Tiny Earth initiative, and offer opportunities to reason with place-based data from the Tiny Earth national database as a means of testing their own predictions about antibiotic producing bacteria in soil. The design drew from principles of conceptual change and data-literacy research (e.g., Thacker & Herrick, 2025, Thacker et al., 2025) by promoting compellingness of data by highlighting its relevance through place-based learning, and comprehensibility through data literacy support.

We found that students significantly improved their perceptions of data science relevance and microbiology knowledge from pretest to posttest. This is consistent with research showing that interdisciplinary and place-based applications of data science techniques can boost student motivation and learning (Thacker et al., 2025). Findings also support models of conceptual change indicating that motivational factors are key processes in learning (Lombardi et al., 2016; Dole & Sinatra, 1998). Importantly, we found that these gains extend beyond students' attitudes toward data science. Shifts in language use and variable identification indicate growth in foundational data science practices that are often overlooked in introductory STEM courses. The ability to reason about which variables matter, and why, precedes statistical modeling and is essential for meaningful engagement with data. These results point to the instructional value of situating data science within authentic scientific contexts that leverage students' disciplinary knowledge.

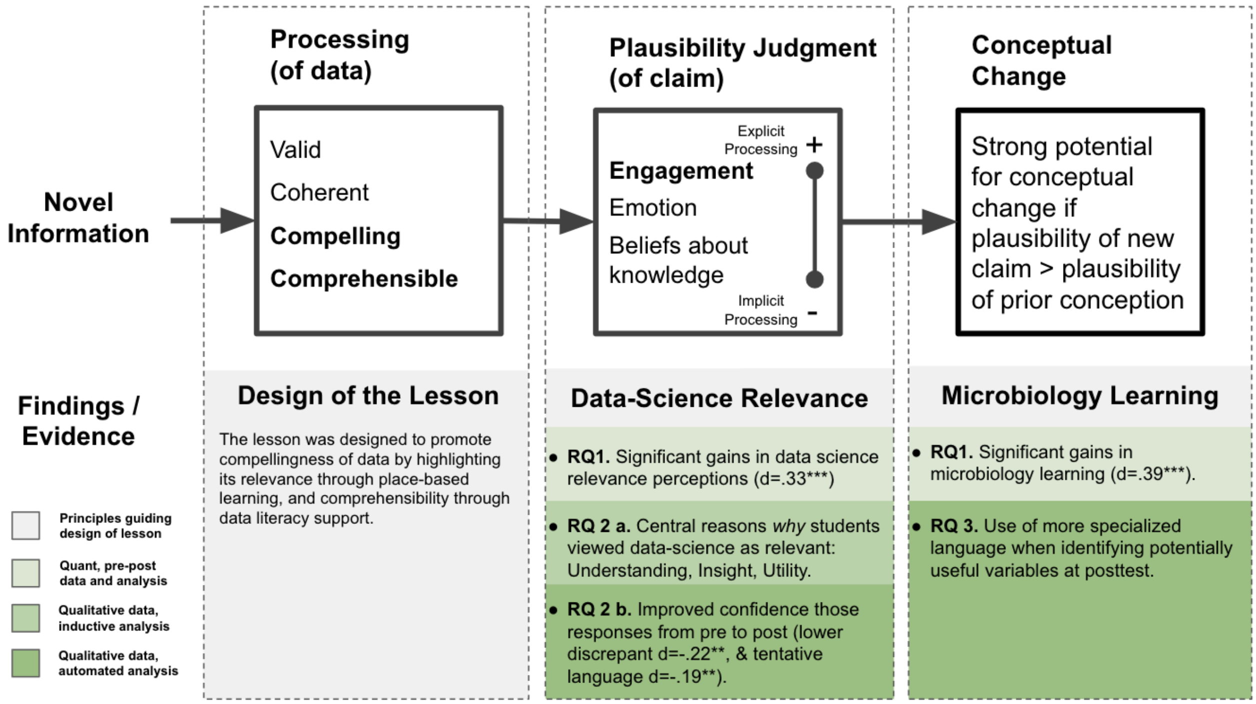
We also investigated student's open ended explanations for why data science may or may not be relevant, and found that they tended to explain data science relevance in terms of its utility across multiple disciplines (e.g., medical and STEM professions) and its capability to help scientists better understand properties of soil data. This suggests that students are predisposed to expect data tools to support their understanding of multidisciplinary topics, and appear to be eager to learn and apply what they know to problem-solve in multiple domains. We believe that this is an asset that might be leveraged to enhance student learning. Future researchers and practitioners might build from student's optimistic expectations of data science and emphasize the insight- developing benefits of data-analysis techniques when applied to soil analysis or other domains.

Automated text analysis revealed that students used less discrepant and tentative language at posttest compared to pretest when discussing data relevance, indicating improved levels of certainty. We also found that students identified lists of soil-related variables that became more specific and specialized over time. Terms such as "animals" and "organisms" became less emphasized in place of more relevant microbiological properties such as "pH", "nitrogen," and "chemical." This suggests that students demonstrated improvement in terms of confidence and identification of relevant variables, both of which are critical to persisting in data-literate professions. Further, the word "soil" was the most prominent term used across all aspects of student thinking, suggesting that there were concrete perceptual anchors guiding student learning throughout this experience.

As such, we provide both quantitative and qualitative evidence suggesting that place-based soil data investigations can enhance students' perceptions that data science is relevant and their microbiology knowledge. Specifically, our quantitative demonstrated significant gains in students' perceptions that data-science is relevant for the field, while the qualitative evidence illustrates the multiple reasons for why they perceived data science to be relevant and that their confidence in these statements improved. These data relevance perceptions, being key indicators of engagement and motivation, are theorized to drive science learning and conceptual change (Thacker & Herrick, 2025; Dole & Sinatra, 1998; Lombardi et al., 2016), which is another key finding. Namely, students' significant improvements in microbiology knowledge is consistent with theory (Figure 1) predicting that, when students engage with data that is perceived as valid, coherent, compelling, and comprehensible, they will more deeply engage with science ideas, and have higher potential for learning and conceptual change. This synthesis across our findings and theory are illustrated in Figure 4.

Figure 4.

Summary of Evidence and Findings in Relation to Theoretical Model Tested



Note: RQ = Research Question. Coefficients for *d* represent Cohen's d, differences between pretest and posttest in units of standard deviations. *** $p < .001$, ** $p < .01$.

Limitations

This study had necessary limitations. First, while our prior research utilized a control group to enhance causal inferences resulting from the implementation of our intervention (Thacker et al., 2025), the current study did not. As such, we recognize that changes from pretest to posttest may not necessarily have been caused by the intervention. Future research might incorporate experimental or quasi-experimental research designs to enhance the potential for causal inference. Second, the purpose of the study was to explore the perspectives and learning outcomes of racially/ethically groups of students, and thus our study sample had a high proportion of Hispanic students compared to the national average. While this may be considered a strength of the study in terms of its inclusion of traditionally understudied groups of students, we also recognize that this may limit the generalizability to teaching contexts in which the student makeup is more reflective of the national average.

Conclusions

Generally speaking, our findings demonstrate that data-literacy supports combined with place-based learning experiences have the potential to dually enhance students' knowledge, motivation, and data-literacy in terms of identifying analytical variables (Lombardi et al., 2016; Semken et al., 2017; Börner et al., 2019; Carlson & Johnston, 2014). We encourage researchers and practitioners to consider grounding data-analysis experiences in real world and place-based scenarios in order to support students in understanding the underlying chemical, biological, and socio-political forces that undergird what they observe.

References

Thacker, I., Herrick, I. (2025). Data Literacy for Conceptual Change: An Integrated Model With Special Applications in Climate Change Education. In R. M. Zbiek, X. Yao, A. McCloskey, & F. Arbaugh (Eds.), Proceedings of the forty-seventh annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 1722–1733). State College, PA. https://doi.org/10.51272/pmena.47.2025

Thacker, I., Shroeder, R., Shields-Menard, S., & Goforth, N. (2025). Dirt Don't Hurt: How relevant soil data can support learning and motivation at a Hispanic Serving Institution. International Journal of Science and Mathematics Education, 23, 803–826. https://doi.org/10.1007/s10763-024-10491-1

Thacker, I. (2024). Supporting secondary students' climate change learning and motivation using novel data and data visualizations. Contemporary Educational Psychology, 78, 102285. https://doi.org/10.1016/j.cedpsych.2024.102285

Thacker, I. (2023). Climate change by the numbers: Leveraging mathematical skills for science learning online. Learning & Instruction, 86, 101782. https://doi.org/10.1016/j.learninstruc.2023.101782

Thacker, I. & Sinatra, G. M. (2022). Supporting climate change understanding with novel data, estimation instruction, and epistemic prompts. Journal of Educational Psychology, 114(5), 910–927. https://doi.org/10.1037/edu0000729

Baumer, B. (2015). A data science course for undergraduates: Thinking with data. The *American Statistician, 69*(4), 334-342.

Ben‑Zvi, D., & Garfield, J. (2008). Introducing the emerging discipline of statistics education. *School Science and Mathematics, 108*(8), 355-361.

Borba, M. C. (1990). Ethnomathematics and education. For the learning of mathematics, 10(1), 39-43.

Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 1857–1864. https://doi.org/10.1073/pnas.1807180116

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*. Austin, TX: University of Texas at Austin.

Carlson, J., & Johnston, L. (2015). Data information literacy: Librarians, data, and the education of a new generation of researchers. Purdue University Press.

Charmaz, K. (2015). *Grounded theory. Qualitative psychology: A practical guide to research methods,* 3, 53-84.

Dill-McFarland, K. A., König, S. G., Mazel, F., Oliver, D. C., McEwen, L. M., Hong, K. Y., & Hallam, S. J. (2021). An integrated, modular approach to data science education in microbiology. *PLoS computational biology*, 17(2), e1008661.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist, 33*(2-3), 109-128.

Fry, R., Kennedy, B., & Funk, C. (2021). STEM jobs see uneven progress in increasing gender, racial and ethnic diversity. Pew Research Center Science & Society.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-25.

Gosselin, D., Burian, S., Lutz, T., & Maxson, J. (2016). Integrating geoscience into undergraduate education about environment, society, and sustainability using place-based learning: three examples. Journal of Environmental Studies and Sciences, 6, 531-540.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22–25. http://iase-web.org/Publications.php?p=SERJ

Gruenewald, D. A. (2003). The best of both worlds: A critical pedagogy of place. *Educational Researcher, 32*(4), 3-12.

Hahn, A. S., Konwar, K. M., Louca, S., Hanson, N. W., & Hallam, S. J. (2016). The information science of microbial ecology. Current opinion in microbiology, 31, 209-216.

Hey, T., Tansley, S., Tolle, K. & Gray, J (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Redmond, WA: Microsoft research. https://www.microsoft.com/en-us/research/publication/fourthparadigm-data-intensive-scientific-discovery/.

Hurley, A., Chevrette, M. G., Acharya, D. D., Lozano, G. L., Garavito, M., Heinritz, J., ... & Handelsman, J. (2021). Tiny earth: a big idea for STEM education and antibiotic discovery. MBio, 12(1), 10-1128.

Kim, J., Hong, L., Evans, S., Oyler‐Rice, E., & Ali, I. (2023). Development and validation of a data literacy assessment scale. Proceedings of the Association for Information Science and Technology, 60(1), 620-624.

Kokka, K. (2022). Toward a theory of affective pedagogical goals for social justice mathematics. Journal for Research in Mathematics Education, 53(2), 133-153.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465-491.

Lombardi, D., Nussbaum, E. M., & Sinatra, G. M. (2016). Plausibility judgments in conceptual change and epistemic cognition. *Educational Psychologist, 51*(1), 35-56.

Markowitz, D. M. (2023). Analytic thinking as revealed by function words: What does language really measure?. Applied Cognitive Psychology, 37(3), 643-650.

Miller, S., Hernandez, P. R., Du, W., Aldana, C. C., Lee, H., Maldonado, N., ... & Estrada, M. (2025). Tiny Earth CURE Demonstrates Equitable Benefits for US College Science Students. *CBE—Life Sciences Education, 24*(2), ar30.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Authors.

National Science Foundation (2015). Science and engineering degrees, by race/ethnicity of recipients: 2002-12. Arlington, VA.

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* Washington, DC: The National Academy Press.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. PLOS ONE, 9(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Pennebaker, J. W. (2001). *Linguistic inquiry and word count: LIWC 2001*. Austin, TX: Pennebaker Conglomerates.

Prado, J., & Marzal, M. A.´ (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri, 63*(2), 123–134.

Qiao, C., Chen, Y., Guo, Q., & Yu, Y. (2024). Understanding science data literacy: a conceptual framework and assessment tool for college students majoring in STEM. *International Journal of STEM Education*, *11*(1). https://doi.org/10.1186/s40594-024-00484-5

Rybarczyk, B. J., Walton, K. L., & Grillo, W. H. (2014). The development and implementation of an instrument to assess students' data analysis skills in molecular biology. *Journal of Microbiology & Biology Education, 15*(2), 259-267.

Ridsdale, C., Rothwell, J., Smith, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report.* Dalhousie University. https://dalspace. library.dal.ca/handle/10222/64578

Saldaña, J. (2021). *The coding manual for qualitative researchers.* Sage Publications

Semken, S., Ward, E. G., Moosavi, S., & Chinn, P. W. (2017). Place-based education in geoscience: Theory, research, practice, and assessment. Journal of Geoscience Education, 65(4), 542-562.

Steen, L. A. (2001). Mathematics and numeracy: Two literacies, one language. *The Mathematics Educator, 6*(1), 10-16.

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. Educational Studies in Mathematics, 96(1), 33-47.

Supplemental Materials

Appendix A: Supplemental Tables

Table A1 *Summary of Content in Focal Asynchronous Module*

Module Section	Core Topics	Key Content Emphases	Learning Activities
Tiny Earth Initiative	Course-based undergraduate research; crowdsourced discovery	Overview of the Tiny Earth Initiative as a global research network engaging students in a research experience towards antibiotic discovery	Student-sourcing Antibiotic Discovery overview video; Collaborative Padlet activity to elicit prior knowledge and questions
Antibiotic Resistance Crisis	Antibiotic resistance; ESKAPE pathogens	Global and U.S. burden of antibiotic-resistant infections; biological mechanisms of resistance (selection pressure, mutation, horizontal gene transfer); contributors to resistance and barriers to new antibiotic development	Instructional video; formative checks for understanding (multiple-choice, true/false)
Microbial Ecology of Soils	Soil microbiology; microbial competition; secondary metabolites	Role of soil bacteria in antibiotic production; ecological drivers of antibiotic synthesis; influence of soil properties (organic matter, pH, temperature, moisture, microbial diversity)	Formative checks for understanding (multiple-choice, true/false)
Antibiotic Discovery Pipeline	Soil sampling; bacterial isolation; screening; genomics	Stepwise description of the Tiny Earth discovery pipeline, from local soil collection to genomic and metabolomic characterization of promising isolates	Guided walkthrough of pipeline stages; embedded knowledge checks
Foundations of Data Science	Data science concepts; interdisciplinary applications	Definition of data science; role of data scientists in extracting insights from complex datasets; relevance to microbiology and public health	Overview of the data science process (get, explore, model); concept checks
Data Science Tools and Processes	Programming languages; analytic workflows	Common data science languages (e.g., Python, R, SQL); justification for Python and Jupyter Notebooks; data types and variable structures	Examples of data types; analytic tools (using Jupyter notebook) checks for understanding
Data Science in Tiny Earth Research	Machine learning; data mining; visualization; cloud computing	Applications of data science to accelerate antibiotic discovery, including pattern detection, prioritization of isolates, data sharing, and interdisciplinary collaboration	Formative checks for understanding (multiple-choice, true/false)
Exploratory Data Analysis and Visualization	Data cleaning; visualization; hypothesis evaluation	Preparation of real-world datasets; identification of outliers and data entry errors; use of visualizations to compare locations, identify patterns, and monitor trends	Exploration of static data visualizations (see Figure 2); evaluation of hypothesis support
Soil Data Investigation and Future Directions	Interpretation of results; iterative science; modeling	Value of inconclusive or null findings; hypothesis generation; introduction to predictive modeling and future research directions	Padlet-based research question generation; reporting of findings; summative assessment

Note. For more detail about the design of the module, see (Thacker et al., 2025). For full access to the asynchronous module, see https://www.softchalkcloud.com/lesson/serve/IEmc5RhkOGrA9n/html.

Table A2.
Summary of Research and Findings

Research Questions	Data Sources	Analyses	Findings
RQ1. To what extent do student's perceptions of data science relevance and microbiology knowledge change after exploring place-based soil data and data visualizations?	Pretest, Posttest comparison of N=298 undergraduate students' microbiology knowledge and data science relevance perceptions.	Wilcoxon signed-rank tests.	Significant gains in microbiology knowledge (p < .001, d=.39) and data science relevance perceptions (p < .001, d=.33).
RQ2. (a) How do students describe the relevance of data science tools for exploring the microbial content of soil? **RQ2 (b)** What levels of analytical thinking are evident in those descriptions? And how do these dimensions change after students explore place-based soil data and data visualizations?	Qualitative responses to the question: "Explain why you think that data science is or is not relevant for the field of soil microbiology? Provide examples if possible."	**(a)** Inductive qualitative analysis (Appendix B) **(b)** Automated analysis using Linguistic Inquiry Word Count, Analytical Thinking dictionary to identify use of Analytic words at pre and posttest and Wilcoxon rank tests to compare pre and post.	**(a)** Students perceive data tools to offer: *understanding and insight* (comparing, contrasting, describing), and *utility* (to address real-world problems in the field, save time, communicate findings). A small number of students did not see data-science as relevant, and some provided vague responses. **(b)** Lower levels of discrepant (p = .006, d =-.22) and tentative (p=.006, d = -.19) language at posttest. (I.e., more confident/certain in their responses.)
RQ3. What soil-related variables do students identify as being relevant for data exploration? And how do these perceptions change from pre- to post-intervention?	Qualitative responses to the question: "Please make a list of any variables that might be relevant for data scientists to investigate when exploring information about soil."	Automated analysis using Linguistic Inquiry Word Count (LIWC), individual word counts presented and compared at pretest and posttest.	Relative frequencies of words similar at pretest and posttest, more specific language used when describing and identifying specific variables.
Synthesis: Students engaged with an asynchronous learning module focused on introducing students to the goals of the TinyEarth initiative, their data repository, and providing them an opportunity to posit and address their own research questions using that data. It was intentionally designed to provide coherent, compelling, and comprehensible narratives and data (see Thacker et al., 2025). Qualitative and quantitative evidence illustrate that students improved their microbiology knowledge and perceptions that data-science is relevant for the field. *How* and *why* students perceive data science to be relevant is clarified in the qualitative evidence across multiple dimensions, and in their improved confidence and identification of more relevant, specialized variables over time.			

Appendix B

Coding Guide for Interview Data

Vague

Evidence that the individual used broad responses without explaining how to specifically use data science within and outside the context of microbiology (e.g., “I think that data can be used as a beginning point for the field of soil microbiology,” “I believe it is relevant because, looking for and getting results for these is not all purely qualitative data.”)

Not Relevant

Evidence that the individual did not find data science to be relevant for soil microbiology. (e.g., “*N/A*”, “*IDK*”, “*Not relevant*”)

Utility: Perception that data science can be useful for specific real-world tasks and professions

Field-specific examples	Evidence that the individual sees the use of data science by providing examples of how the data can be used in specific fields/industries. (e.g., “*...data science is highly relevant in soil microbiology as it empowers researchers to make sense of large and complex datasets, model soil ecosystem processes, and apply findings to various fields like agriculture, environmental management, and climate change research*”)
Techniques	Evidence that the individual sees the use of data science methods and how they can be applied to study microbial processes in soil (e.g., “*Techniques like machine learning, predictive modeling, and data visualization enable researchers to identify patterns in microbial communities and predict factors influencing soil health or antibiotic production*,” also examples were noted involving “*statistical analysis*,” “*machine learning*,” “*record keeping*”)
Efficiency	Evidence that the individual sees the use of data science as a way to reduce time and increase efficiency and quality in microbiology research (e.g. “*It is important for the organization of data that can lead to discovery, increase efficiency, and ultimately spread information through things like data mining, visualization, and machine learning.”* or *“Makes organizing vast amounts of data more efficient.*)
Data presentation	Evidence that the individual understands that data presentations (written, numerical, or graphic) is useful for interpretation and communication in microbiology. (e.g., “*Data science is very relevant to microbiology, data science allows us to test the bacteria and show people the percentage of antibiotic resistance, etc.*,” as another example, “*data visualization tools can create interactive maps and graphs to illustrate the distribution of microbial communities in different soil types.*”)
Data Organization Uses	Evidence that the individual understands the importance and utility of managing and saving collected data (such as descriptive data and/or other associated data, for proper identification, organization, and/or future reuse, while potentially recognizing the need for large and diverse datasets; perhaps, to ensure comprehensive analysis and informed decision-making; e.g., “*Data science can organize and store important data such as the NCBI GenBank*,” or “*Data science is relevant in the field of soil microbiology due to its ability to organize and sort through data allowing for accurate findings and data collection.*" or “*Data science is relevant for the field of soil microbiology since it organizes complex data and allows us to identify patterns and correlations just like how cloud computing is useful for the Tiny Earth*

	project.").
Understanding: Perception that data helps to illuminate, explain, or provide insight into properties of soil.	
Compare and contrast (across locations, over time, etc.)	Evidence that the individual understands the relevance of data science by comparing and contrasting datasets from various locations and/or over periods of time to provide researchers insight (e.g., "It is relevant because it allows for the comparison between different soil samples and locations," *"...communication, image processing, data sampling, and knowledge of programming languages such as python are crucial for understanding the vast type of information interrelated with soil sampling and comparing Its data to other geographical locations to determine isolation of antibiotic-producing bacteria*," "*Data science is relevant to soil microbiology because it allows us to compare soil samples from various locations and soil over time*")
Describing data, trends, and patterns	Evidence that the individual understands the relevance of data science by its ability to describe, analyze, and/or identify trends and patterns in data to give researchers new insight (e.g., "*Data science techniques, including statistical analysis, machine learning, and data visualization, can be used to analyze and interpret these datasets to extract meaningful insights about microbial communities, their interactions, and their impact on soil health*," also "...analyze and interpret data so as to identify trends in the data," also "*Now I feel that previous answer has been proven. Data science is essential to the soil microbiology field to analyze and interpret data so as to identify trends in the data, getting us closer to the potential discovery of new antibiotics that work against the ESKAPE pathogens*")

Appendix C

Microbiology Knowledge Measure (Researcher Created)

ESKAPE definition: ESKAPEs are 6 bacteria that are considered to be major threats as they comprise the majority of antibiotic-resistant infections. (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter species*)

1. What is the source of most of the antibiotics in current use?

a) Chemical labs
b) Plants
c) Soil bacteria
d) Water bacteria

2. The pharmaceutical industry is investing most of its money in identifying novel compounds from soil bacteria.

a) True
b) False

3. Which of the following is NOT a goal of the Tiny Earth Initiative

a) Give students some research experience
b) Discover new antibiotics
c) Gain an understanding of the antibiotic crisis
d) Determine the amount of antibiotic resistance in the community

4. Antibiotic resistance means that

a) A bacterium produces antibiotics
b) A bacterium is susceptible to antibiotics
c) A bacterium is not susceptible to antibiotics
d) A person is immune to antibiotic treatment

5. Compare the various definitions and justifications below and pick the one that is most accurate.

a) The bacterium has changed physically or chemically in some way to be able to destroy the drug or avoid its action, allowing it to grow unimpeded by the drug.
b) The bacterium becomes immune to the drug; the drug no longer kills or inhibits the bacterium.
c) The person becomes resistant to the drug; the body adjusts to the dosage of the chemical and no longer responds to its action.
d) The drug is changed in the body and is inactivated physically and chemically so it no longer works properly against the bacterium.

6. What is an ESKAPE pathogen?

a) A group of 6 bacteria that are considered major threats
b) A group of 6 bacteria that comprise the majority of antibiotic-resistant infections
c) A group of 6 bacteria that have "escaped" successful antibiotic treatment
d) All of the above

7. Environmental factors, like temperature and pH, have an impact on the growth of antibiotic-producing bacteria in the soil. How do these factors influence the production of antibiotics by antibiotic-producing bacteria in the soil?

a) **Temperature and pH can influence growth and the regulation of secondary metabolism in antibiotic-producing bacteria, which ultimately determines the intensity of antibiotic production**
b) Environmental factors stimulate antibiotic production during primary metabolism in antibiotic-producing bacteria
c) All antibiotic-producing bacteria can survive drastic environmental changes, so temperature and pH are not relevant
d) Humidity is necessary for antibiotic production

8. The antibiotic resistance crisis converges with the soil crisis

a) **because antibiotics are sourced from soil, soil erosion threatens the discovery of novel antibiotics to combat the antibiotic resistance crisis**
b) because antibiotics are sourced from water, soil erosion contaminates water streams, threatening the discovery of novel antibiotics to combat the antibiotic resistance crisis
c) because antibiotic resistant microbes are in the soil and soil erosion causes increased interaction with humans and these pathogens
d) because antibiotics are sourced from plants which need the nutrients from the soil

9. Assign numbers, from start to finish, of the Tiny Earth Antibiotic Discovery Process: (Ranking item)

1. Soil Sampling
2. Bacterial isolation
3. Antibiotic screening
4. Isolate Characterization and Genomic Analysis

Appendix D

Data Visualization Literacy (Researcher Created)

Assessment principles were adapted from page 1862 of: Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. Proceedings of the National Academy of Sciences, 116(6), 1857-1864.

1. Data Science Processes
In your opinion, are data science techniques **relevant** for addressing questions in soil microbiology?

- Not at all relevant, Slightly relevant, Relevant, Fairly relevant, Very relevant.

2. Explain how/why you think that data science is or is not relevant for the field of soil microbiology. Provide examples when relevant.

3. **[Display logic: if response to (1) is NOT "not at all relevant", display:]** Please make a list of any variables that might be relevant for data scientists to investigate when exploring information about soil. (Separate entries with commas)