Student Skills that Result from Arts-Based Data Literacy Teaching

Research Report: Full Study Findings

Final Report: 06/26/2025

J. Sickler CONSULTING

PREPARED BY

Jessica Sickler

Michelle Lentzner

BUILDING INSIGHTS THROUGH OBSERVATION

Table of Contents & Key Findings

04 **Background & Methods**

12 **Results: Interpretation & Evidence**

- BIO students showed significant growth in ability to make more robust interpretations about data. The number of inferential interpretations (claims) doubled from the pre-test.
- BIO students' growth in interpretation skills was significantly stronger than Control group peers.
- Post-BIO, over half of students were able to make claims and support them with robust evidence, an increase of 40 percentage points from pre-test.
- BIO students' growth in making claims and supporting with evidence was significantly stronger than peers in the Control group.

25 **Results: Observing & Decoding**

- BIO students showed much greater growth in their geography knowledge, through the ability to identify unlabeled locations, than their Control group peers.
- BIO students showed significantly greater increases in observing data patterns – including similarities, contrasts, and aggregate patterns – compared to Control group peers.
- BIO significantly increased the total number of observations made by a student compared to the Control condition, where there was almost no change.

39 Results: Flexible & Scientific Thinking

- After experiencing BIO lessons, 70% of students demonstrated flexible thinking; of these, 26% were able to generate 3 or more possibilities of what the visualization could be trying to show.
- Both groups saw increases in flexible thinking skills, but BIO students showed slightly higher rates of making 3+ distinct syntheses from the data, compared to the control group

47 Results: Self-Reported Learning

- Teachers felt BIO was most effective at helping students build visual literacy and engaging a broad range of students, including those that don't typically enjoy science topics.
- Almost 60% of interviewed students recalled something specific they had learned from BIO, including science information, thinking analytically, and awareness about data and visualizations.

52 Results: Differences Across Classrooms

- Each classroom increased the number of claims made (inferential interpretations) after BIO.
- Each classroom saw an increase in inferential interpretations supported by adequate/complex evidence after experiencing BIO lessons.
- In all classrooms, students grew the most in making observations about geography and data patterns.

56 Conclusions

BACKGROUND

Research Overview & Methods

Building Insights through Observation (BIO) was a research and development project to develop a new approach to teaching data literacy by integrating approaches from visual arts education for use with geospatial data visualizations in the middle school science classroom. The resulting BIO framework emerged from a multi-year collaboration between a multi-disciplinary team of curriculum developers, researchers, visualization specialists, and 10 middle school science teachers from around the U.S. The process involved teacher professional development, iterative curriculum development, and research reflection.

This Research Brief presents findings centered on Research Questions 5 and 6, shown on the following page.

BIO Research Overview

The BIO project aimed to address six research questions within three lines of inquiry – instruction, teacher growth, and student skills.

This brief addresses the bolded questions. Findings from other research questions can be found in additional Research Briefs.

Instructional Approaches

RQ1: Which applications or combinations of arts-based instructional approaches do teachers use and find successful at addressing their curricular needs and goals?

RQ2: In what ways do teachers' views and uses of techniques vary, and what does that suggest about the broader potential of this approach?

Teacher Professional Growth & Practices

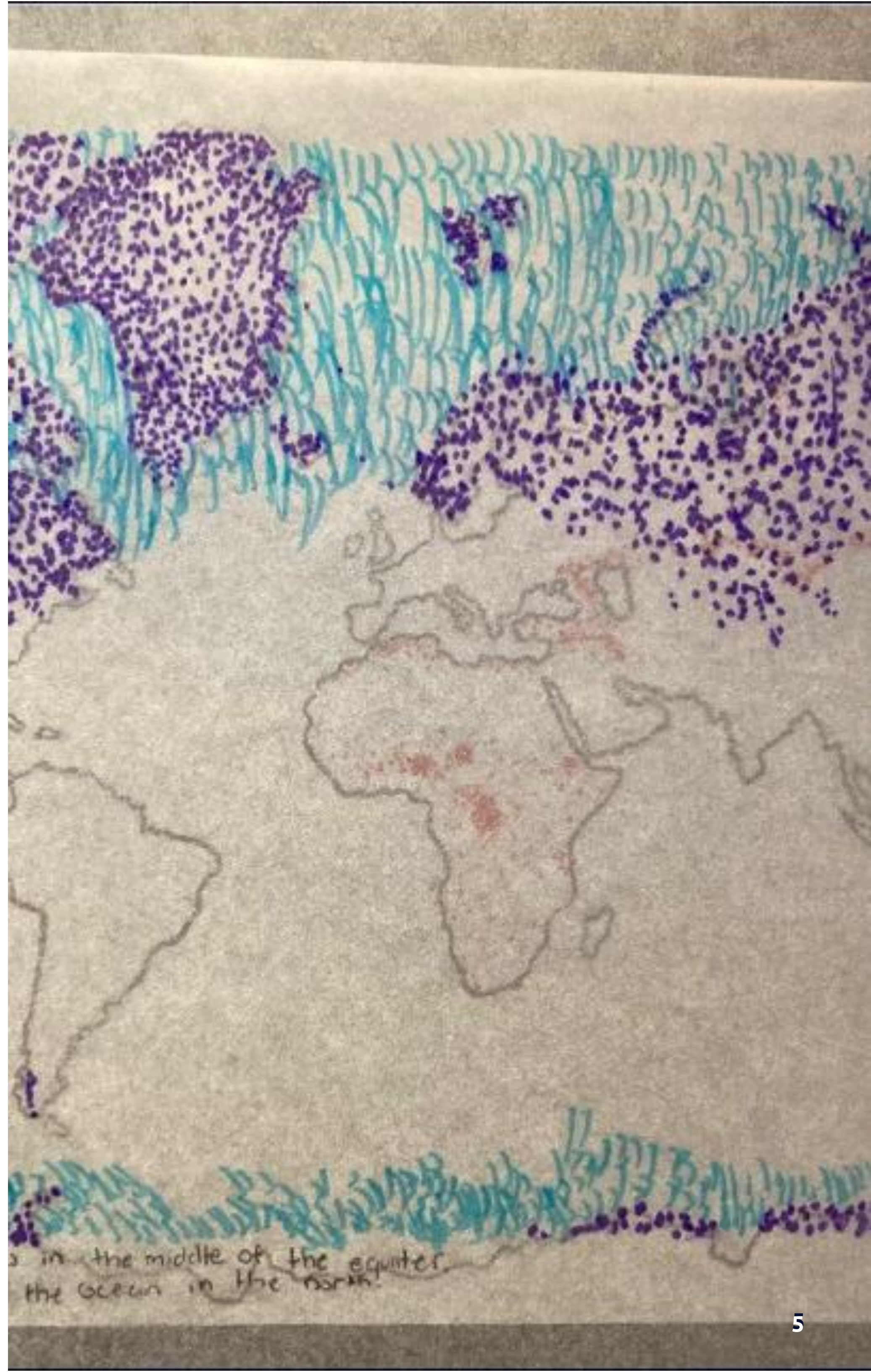
RQ3: In what ways do science teachers develop their professional knowledge and practice, particularly around geospatial data visualization content and pedagogy?

RQ4: What elements of the process of learning, application, reflection, and community-building do they find most influential to their professional growth?

Promise for Developing Student Skills

RQ5: When arts-based approaches are applied in science classes, in which of the data/visual literacy skills do students show the greatest growth?

RQ6: Compared to standard instruction, do any skill areas show particular promise of being influenced by the arts-based approach?



BIO Research Overview

This Research Brief presents the development of data literacy skills among students in the classrooms of BIO's Cohort 1 teachers. Students were given a data literacy performance task before and after experiencing two rounds of BIO lessons to examine areas of growth and promise.

The Building Insights through Observation (BIO) project was implemented in partnership with 10 middle school science teachers from around the U.S. Teachers were in 2 cohorts, participating from 2022 to 2024. The timeline of PD experiences within BIO were as follows:

- **2021-22 School Year:** A class from each Cohort 1 teacher does a data literacy performance task at the beginning of the school year and again at the end to serve as a control group for typical year-long learning among each teacher
- **Summer PD 2022**: Cohort 1 initial 3-day Workshop (Easton, PA)
- **2022-23 School Year Implementation**: Cohort 1 members create and implement 2 lesson plans per teacher; No student data collected
- **Summer PD 2023**: Cohort 2 & Cohort 1 at 3-day Workshop (Easton)
- **2023-24 School Year Implementation**: Both Cohorts' members create and implement 2 lesson plans per teacher; Students from selected Cohort 1 classrooms do a data literacy performance task at the beginning of the school year and again after completing two round of BIO lessons
- **Summer PD 2024**: Culminating workshop

BUILDING INSIGHTS
THROUGH
OBSERVATION

Research Questions 5 and 6:

When arts-based approaches are applied in science classes, in which of the data/visual literacy skills do students show the greatest growth? Compared to standard instruction, do any skills show particular promise of being influenced by the arts-based approach?

BIO Lesson Overview

BIO implementation centers around three key activities, plus a culminating discussion, designed to integrate visual thinking skills, design thinking skills, and content understanding to develop data literacy using arts-based practices and professionally collected geospatial data.

1.) Discussions inspired by Visual Thinking Strategies (VTS)*

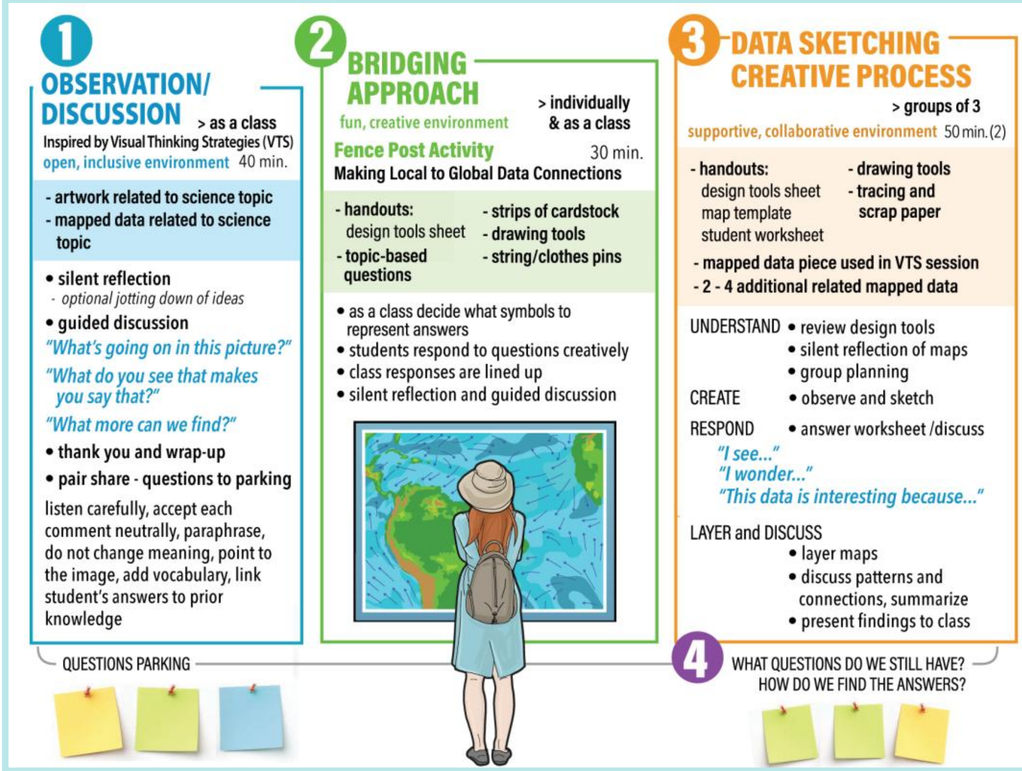
As a class, teachers guide students through a carefully facilitated discussion, first about a piece of artwork and then about a data visualization, using the VTS technique. Discussions are open-ended, generating interpretations and observed evidence to support them.

2.) Fenceposts

Students respond to a series of five teacher-designed poll questions, about their personal experiences and opinions related to the unit's topic. Students visually represent their individual answers on slips of paper using agreed-upon symbology (based in Design Tools). Class responses are lined up like fenceposts to generate discussion about the visual patterns that come from symbolic representation of data.

3.) Data Sketches

Each small group receives ~3 visualized geospatial data sets. Students are tasked with identifying ways to re-visualize each of the data sets using different symbology than the original. Each student resketches the data from one map on tracing paper. After sketching, students layer and compare their visuals, looking for correlations, patterns, and other notable features. Finally, discoveries are discussed as a whole class.



*VTS (https://vtshome.org/) is a well-established and highly researched arts-based discussion approach, traditionally used to generate ideas about artworks, notice details, and make new and surprising connections.

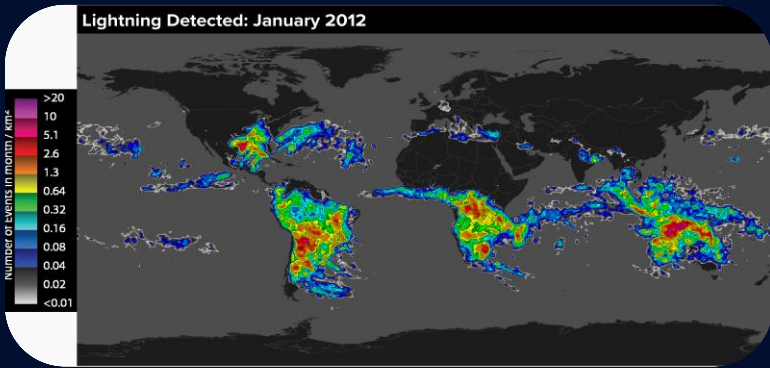
BACKGROUND

Research Study Design: Change in Student Skills

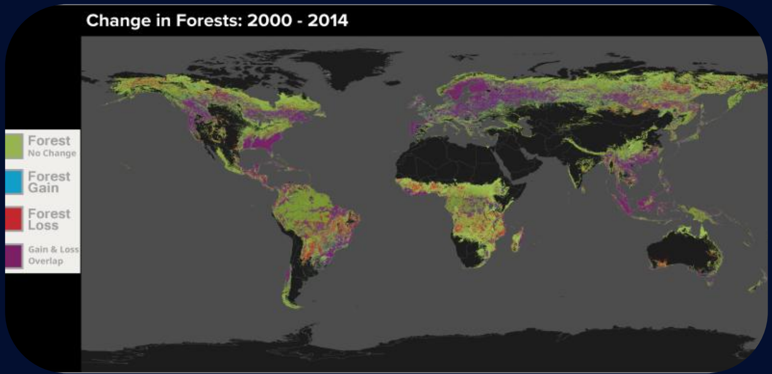
To control for variation in teachers' pre-existing skills, the control group was composed of students whose teachers would be trained in the BIO approach, but had not yet received any project training.

	Pre-Test	School Year Experience	Post-Test
Treatment Group	**Fall 2023** 1:1 data skill interviews	**2023-24** Teachers apply their adaptation of 2 arts-based, data units (after a full year of PD & practice)	**Spring 2024** 1:1 data skill interviews (same students from Fall 2023)
Control Group	**Fall 2021** 1:1 data skill interviews	**2021-22** Teachers apply their pre-PD curriculum and practices (same teachers)	**Spring 2022** 1:1 data skill interviews (same students from Fall 2021)

Pre-Test Image



Post-Test Image



BACKGROUND

Performance Task: Eliciting Student Skills

The Sample

The study samples were drawn from two years of classes of 5 teachers in Cohort 1 of BIO PD. Identifying classes in which they would ultimately use BIO teaching, each Cohort 1 teacher identified a sample of 10-20 students to participate in the data literacy performance task.*

The control sample was collected during the 2021-22 school year (n=76), when BIO teachers were identified but had not yet begun the PD. This allowed us to establish typical skill changes for students learning from these teachers. The treatment sample (n=69) was collected during the 2023-24 school year. These students experienced two complete BIO units with the same teachers, who had also had a full school year in between to learn, practice, and refine their adaptation of the BIO lesson framework. All pre-tests occurred in the fall and post-tests in the spring.

*Parental information and consent procedures, reviewed by an IRB, were followed to recruit and collect data with the samples.

The Data Performance Task

The task involved a 1:1 interview, as students looked at a novel geospatial data image (above). Images were selected using criteria to ensure they directly referred to common phenomena or concepts, understandable regardless of students' grade (6-8) or prior curricula – lightning and forests.

The interview protocol was highly structured to ensure consistency, so that all students had the same opportunity to express their thinking. The interview began with a very open-ended question and follow-up to elicit any idea(s) the student had about what was going on in the data (e.g., "What's going on in this data? What makes you say that?"). This open-ended series was repeated three times with each student. Unscripted probes or follow-ups were not used, to avoid inconsistent scaffolding support. This open-ended portion was followed by a series of four targeted questions to prompt specific types of skills, such as flexible interpretation and asking questions about data.

BACKGROUND

Analysis: Student Interpretation Skills

Interpretation Skills: Coding Structure

Interpreting Meaning: Each student had 3 opportunities to interpret data with minimal scaffolding. Each opportunity received an interpretation category (none, descriptive, or inferential). If a student made multiple types in one segment (a rare occurrence), the higher level was applied.

Evidence Support: When a student received an inferential interpretation code (i.e., made a claim), that segment was also coded for the quality of evidence provided to support the claim using a rubric.

Flexible Thinking: Scored with a rubric of degree of flexibility, focused on a segment of the interview where students were asked to state the visual's meaning and an alternative (as well as prior interpretations made).

Questioning Data: While several questions in the interview prompted students to pose questions, the presence of different types of questioning could occur and be coded at any point in the interview.

Final Coding & Reliability

When a final coding system was developed, 2 researchers independently coded a sample of 30 transcripts (10% of data), randomly selected but ensuring even representation across classes, conditions, and time. Coders achieved .77 or higher reliability. All discrepancies and requests for consensus coding (used any time a coder felt unsure about the correct code application) that arose were discussed and resolved.

Coding of the final 90% of data was completed independently by the two coders. All data from a single classroom was coded by a single coder; this minimized the possible effect of inter-coder variation on change scores in a student or class. While coding, any segment where coding seemed unclear was flagged for review by the second coder, to arrive at a consensus code. A final consistency review was completed on all segments within each code, looking for inconsistencies in application. A few segments were found and reviewed again by the original coder(s) to resolve.

Analysis: Observations

An additional layer of analysis was added to record the number and types of unique observations students made over the course of the interviews. This process focused on a counting and categorization of distinct observations about the data visual in front of the students, as used in prior research.*

This coding was done as an additional layer to the interpretation analysis (discussed on the prior page), as it examined a smaller grain size within students' interpretive statements. In this approach, each transcript was reviewed and individual segments of observation identified. Each segment was applied to a single type of observation being made, which included: (1) Using Title / Legend; (2) Applied Geography Knowledge; (3) Describing Presence of Data; and (4) Describing Data Patterns. Within each of these groups, sub-codes for types of observation were developed (see the Results section for details).

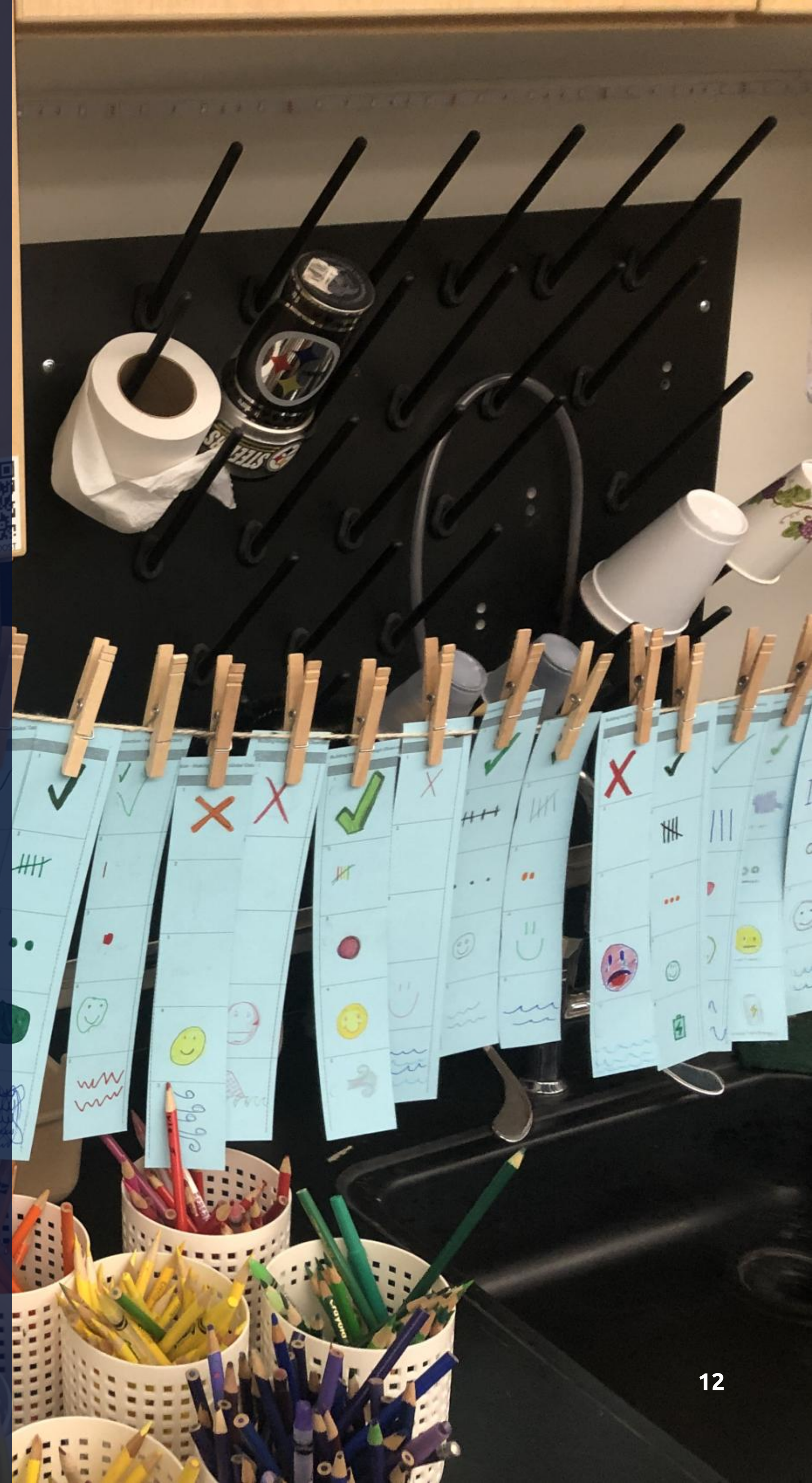
If a student repeated an observation they had previously made (e.g., repeated that there were areas without data/color), it was not double-counted. An emergent code (using population to explain data) was added, to explore if it was, in fact, an area of change.

From this coding, students received two scores: one score for the number of sub-codes (types of observation) used; and one score for the total number of distinct observations (even if they were of the same type). A single scorer was used for this element to minimize intercoder errors. All coded segments were reviewed 2+ times to ensure consistency in final code application.

*Naghshineh, S., Hafler, J. P., Miller, A. R., Blanco, M. A., Lipsitz, S. R., Dubroff, R. P., ... & Katz, J. T. (2008). Formal art observation training improves medical students' visual diagnostic skills. *Journal of general internal medicine, 23*, 991-997.

RESULTS

Interpretation & Evidence



BIO Students' Ability to Make Interpretations

BIO students showed significant growth in their ability to make more robust interpretations about data. The number of inferential interpretations (claims) made by BIO students roughly doubled from the pre-test.

Each of students' responses to the first series of open questions in the interview were scored to one of three levels, based on whether they opted out, described what they could see in the data, or if they made a claim that went beyond the data and legend.

After BIO lessons, making claims was the most common strategy, as compared to pre-BIO, when students more often described the data. Inferential interpretations most often stated causes of why the data might look the way they do. (Evidence used to support claims will be discussed on p. 15)

Descriptive interpretations were more common in the pre-test, and ranged in complexity, with some being as simplistic as noticing and reading the title, while others were comparing regions or noticing patterns. These differences are characterized in the next results section, *Observing & Decoding*.

Change in frequency of interpretation skills used by BIO students over the course of a year with arts-based data instruction.

Each student was given three open-ended opportunities to make interpretations about the geospatial data visualization during their data performance task. The n in the table below represents the total number of opportunities for interpretation (n=68 BIO students with 204 opportunities to make an interpretation).

Interpretive Skills*** Levels of interpretive strategies used; when more than one level was used in a response to a cycle of questioning, the highest level used in that segment of thinking was coded.	PRE	POST
Claim: Inferential Interpretation (Level 3) Student offers an idea that infers something about the data/image that goes beyond what is visible; it can be about what caused the data to look this way, what it could like in different situations, or what effects or implications it could cause.	26%	51%
Description: Descriptive Interpretation (Level 2) Student offers one or more ideas to describe what the data representation includes or shows; including descriptions about the entire map, segments, features, or an individual data point. Interpretation relies entirely on information contained within the visual itself.	69%	48%
No Interpretation (Level 1) Student says, "I don't know," requests to skip the question, or otherwise opted out of answering the question.	5%	1%

***BIO students showed significant growth in their interpretive skills from pre- to post-test (p<.001).

What it Sounds Like: Students' Interpreting Skills

Inferential

"It seems like some places have more forests than others, like at the top of Africa. I'm pretty sure that section has deserts now, so it's probably not as fertile for trees."

-Student 455, Post-BIO

"That it looks a lot like in South America in the center right here that there was a lot of loss that was going on. … Maybe there was a disease that the trees caught, or the vegetation caught, or maybe they had to cut down the trees to make paper or something."

-Student 463, Post-BIO

"I noticed that there's a lot of forests in the upper Northern Hemisphere and that there is less forests in the end of the Southern Hemisphere. In the bottom of South America, there's only a little strip of forest and Australia doesn't have much forest either. … I'd assume that it's a little bit colder because there's also not much forest in Greenland and Iceland, so I assume that since it's further from the equator, it's a little bit colder and the trees don't grow there as well, so there's less forest there."

-Student 661, Post-BIO

Descriptive

"It shows how forests, I guess, [how] the amounts of trees have changed and the key on the side shows by how much. … So the title says, "Change in Forests" and it gives me a time period, I guess, 2000 to 2014. And then the key, it has words like, "no change" or "gain" or "loss" which can show me the amount, I guess."

-Student 462, Post-BIO

"Student: [I think there's] not a lot of forest change. … There's a lot of green and in the key, it says forest no change for the green."

-Student 653, Post-BIO

"I think the data is showing how the forests have been changing from the year 2000 to 2014 with these different colors. … Well, from the legend, I can tell that the greens mean that the forests haven't really gained or lost much, and the blue shows that the forest gained more forest, I guess. And from the reds, you can tell that the forests have lost some of their forest, and the purple shows that gain and loss overlap, which may mean that it both lost and gained, I think."

-Student 351, Post-BIO

Growth in Interpretation Skills: BIO and Control

Students who participated in BIO had a significantly greater growth in their interpretation skills than their Control group peers. BIO students made nearly twice as many inferential interpretations at post-test than at pre-test. Control students showed much less growth.

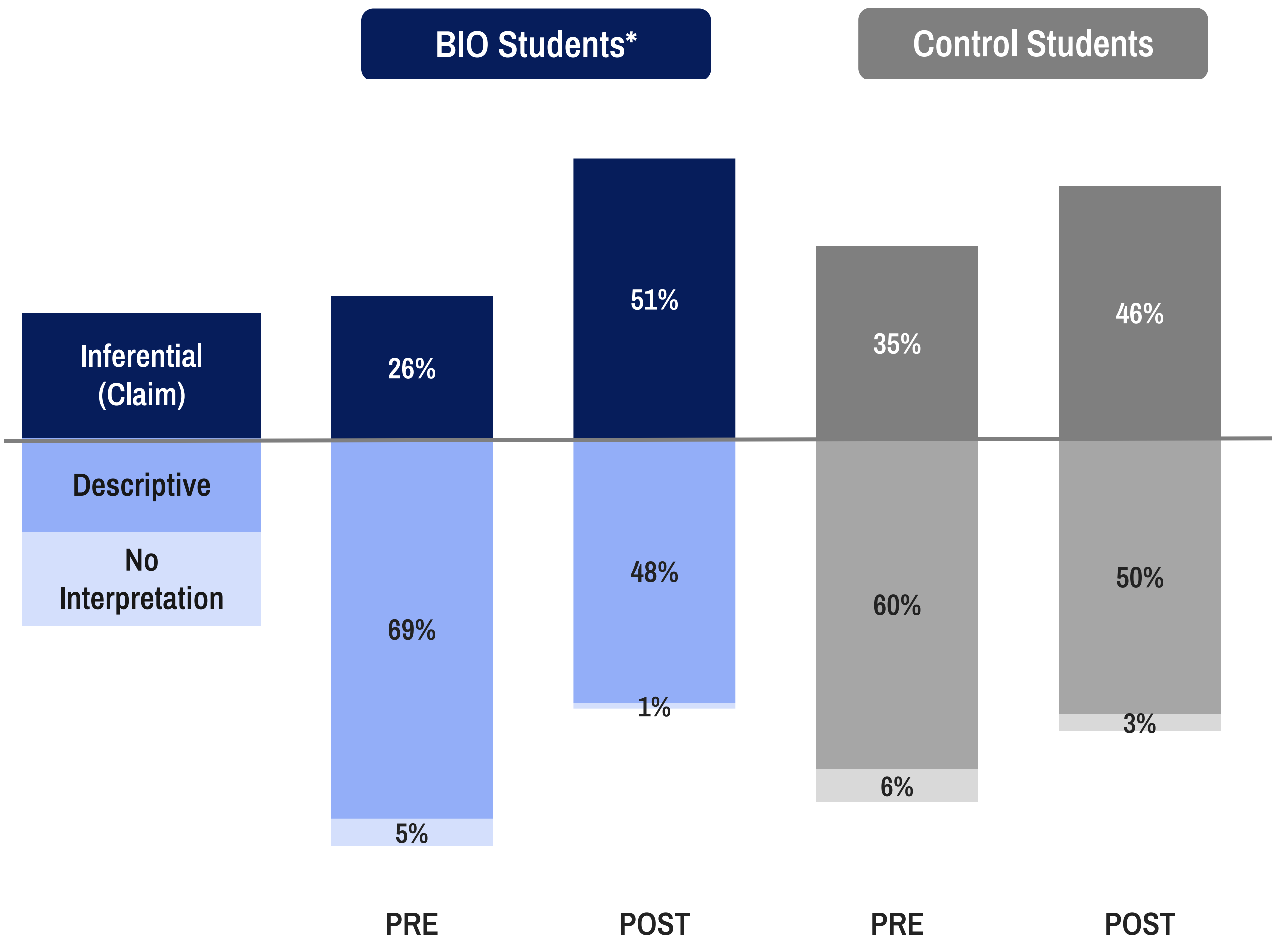
Prior to starting BIO lessons, nearly 7 out of 10 interpretations made by BIO students were descriptive. Students in the Control group started with a higher rate of inferential interpretations.

However, by the end of the school year, the BIO intervention group had shifted towards inferential interpretations and surpassed the control group, with 51% of interpretations being inferential versus only 46% among the control group. This difference was significant.

In addition, the post-BIO students had the lowest rate of opting out of making interpretations. Only 1% of interpretation opportunities (2 individual opportunities) were met with a student opting out of an attempt to interpret. The Control group remained at 3% opting out of opportunities (or 7 individual opportunities).

Differences in students' interpretive strategies before and after participating in two units of BIO lessons.

The graph below shows the percentage of interpretation strategies across each opportunity for interpretation, equaling three opportunities per student (n=204 BIO Pre/Post; n=228 Control Pre/Post).



*Students receiving BIO teaching showed significantly greater growth in interpretation skills than the Control students (p<.05)

BIO Students' Ability to Use Evidence with Claims

After two rounds of BIO lessons, the majority of inferential interpretations were supported with robust evidence, with nearly 1 in 5 supported with complex evidence. Use of complex evidence was rare in the pre-interviews.

When a student made a claim about the data (an inferential interpretation), the quality of their evidence was scored using a three-level rubric, from weak to complex support.

After BIO, 56% of all claims were supported with adequate or complex evidence. These two levels required students to incorporate *both* data and one relevant concept or phenomenon that could be logically related to their claim (usually about why the data patterns looked the way they did).

The frequency of weak support used decreased over time. Weak support tended to include students relying on personal anecdotes, imaginative explanations for cause, or scientific concepts that could not logically be tied back to the data. There were only a handful of instances in which a student failed to offer any evidence at all, which were included in the weak support category.

Change in frequency of evidence used to support claims (when those were made) by BIO students over the course of a year with arts-based data instruction.

When a student made an inferential interpretation (or claim), that segment was scored on the following rubric to reflect the quality of evidence that was used to support their claim.

Quality of Evidence Used to Support Claims Strength of evidence used to support an inferential interpretation, when such an interpretation was made.	PRE (n=52)^	POST (n=104)^
Complex Support (Level 3) Student provides support for their interpretation that involves: At least <u>two data points or patterns</u> plus <u>one relevant scientific concept or human factor</u> AND The data and concept used as evidence are <u>relevant and/or plausible</u>	6%	18%
Adequate Support (Level 2) Student provides support for their interpretation that involves: At least <u>one data point or pattern</u> plus <u>one relevant scientific concept or human factor</u> AND The data and concept used as evidence are <u>relevant and/or plausible</u>	23%	38%
Weak or Absent Support (Level 1) Student provides <u>nothing to support their interpretation</u> OR Student provides support for their interpretation, BUT: The evidence is <u>solely rooted in personal lived experience</u>, is <u>fanciful/imaginative</u>, or hinges on <u>data snafus</u> (broken sensors, glitches) OR Evidence is used that is only <u>weakly related to the interpretation</u> OR Evidence is <u>based only on information from the title or key</u>	71%	44%

^Note: Evidence was only coded when a student had made an inferential interpretation, where they were making a claim that went beyond what was evident from the data, title, and/or legend. Because frequency of making claims varied (see prior pages), the available opportunities to support with evidence (n) varied.

What it Sounds Like: Supporting with Evidence

Complex Support

"So, I noticed there's a lot of purple down towards southern North America, towards Florida and Alabama, and all the southern states, for the gain and loss overlap. And also some in Europe. [This is] probably due to cities and more houses getting built and taking down from the forest. And in South America, mostly in the northern part, there's no change in the forest. And same in Canada and Alaska. ...There's less population up there and they might not need any more buildings and construction that needs to go on." -Student 362, Post-BIO

"I think near the big cities I see lots of purple and green, and according to the key when it says gain and loss overlap, I think it's when trees are either cut down or something and then organizations try and grow them back. I guess I see in cities and in the Amazon forest, I know people cut trees down and then there are different organizations that try and plant trees..." -Student 462, Post-BIO

Adequate Support

"I noticed down here in Antarctica, there's zero forest at all, probably because that's where all of our glaciers are and stuff, so it makes sense that there's no forest.

-Student 617, Post-BIO

"Well, I know India is very highly densely populated, and it shows not much forest at all. I mean, there are a few here and there, but not that much of it has forests, so I'm thinking maybe because of deforestation and maybe developing natural areas with buildings or infrastructure, they lost a lot of forest. Well, it would've been... Well, I guess it might... I'm not sure. Because if it lost forest, then it should have been red, but maybe in this timeframe it was already gone."

-Student 351, Post-BIO

Weak Support

"Student: Maybe it's about the wildlife in it, if it's gaining or losing anything [the wildlife].
Interviewer: What makes you say it's about wildlife?
Student: Because forests have animals and stuff."

-Student 366, Post-BIO

"Student: In a lot of areas, it's just completely... there's nothing at all.
Interviewer: What do you think that might mean, the areas with nothing in them?
Student: [Pointing to Sahara Desert] Either, their population, people have taken over, because it's a pretty straight line from where my mouse is down where the green is, it's pretty straight across. So, it could be where a lot of people live and there's not any forest there, or nothing grows there from the dirt."

-Student 151, Post-BIO

Evidence Spotlight: Thinking about Human Population

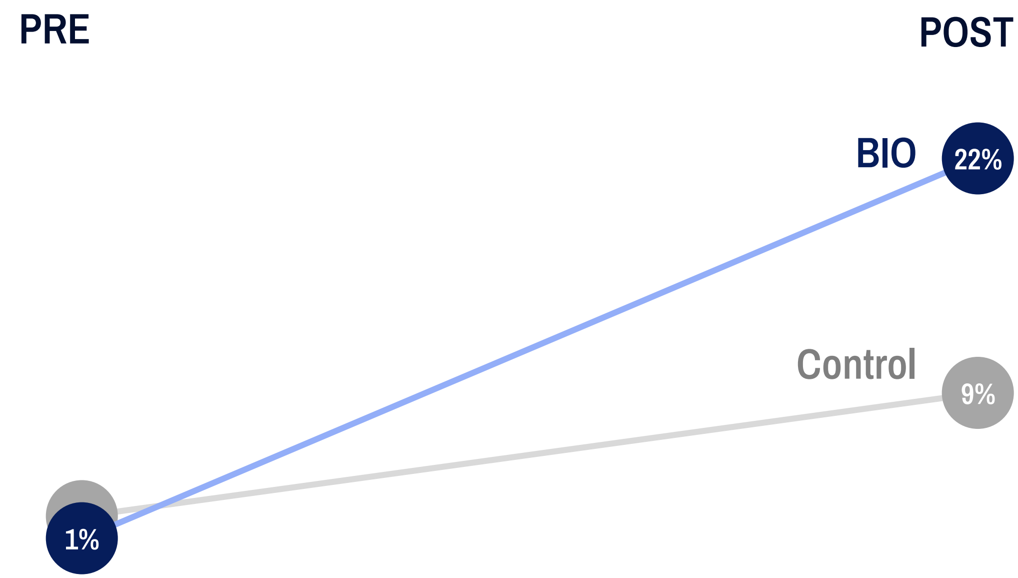
More than 1 in 5 BIO students ultimately brought up a connection between human population and the data they were looking at (forest changes over time). This was far more often than the Control group, and mirrored datasets that all BIO teachers had used in their lessons.

Teachers were encouraged to include at least one dataset that used human factors, and many had students explore a human population map during the Data Sketches activity. This exposure to global population data seemed to influence how students thought about and interpreted a new dataset. They far more frequently applied their gained knowledge about population when interpreting the data and when requesting additional information that might help them better understand the data.

References to population were near zero in pre-interviews, which may reflect lack of relevance to the dataset (lightning events). Post-interview data (change in forests) had more potential to tie to population factors, and slight growth was seen in the Control group. However, the growth among BIO students was much stronger. BIO students sometimes directly stated wanting to layer the population maps they now knew well.

Change in use of human population as evidence or a potential contributing factor students were considering during the data performance interview.

Each student was coded for whether they directly referenced human population as a potential cause or relating factor to the data at any point during their performance interview. (n=69 BIO group; n=76 Control group)



"I noticed there's lots of color where there's big cities, like in the US there's the west coast here and a lot of the east coast. And then I also notice lots of color where there's big forest or natural, untouched land, I guess, like up in Canada and Alaska and in the Amazon Rainforest, I think."

-Student 462, Post-BIO

"If you lined up the population density map, then it would probably line up closely to this [change in forests map]. … So, the population would help because we would know why. It would kind of give us the hint of why there were trees going down or the forest was changing."

-Student 358, Post-BIO

Changes in Use of Evidence to Support Claims

Students who participated in BIO nearly doubled their rate of using adequate or complex evidence, although the control group saw similar but slightly lower gains.

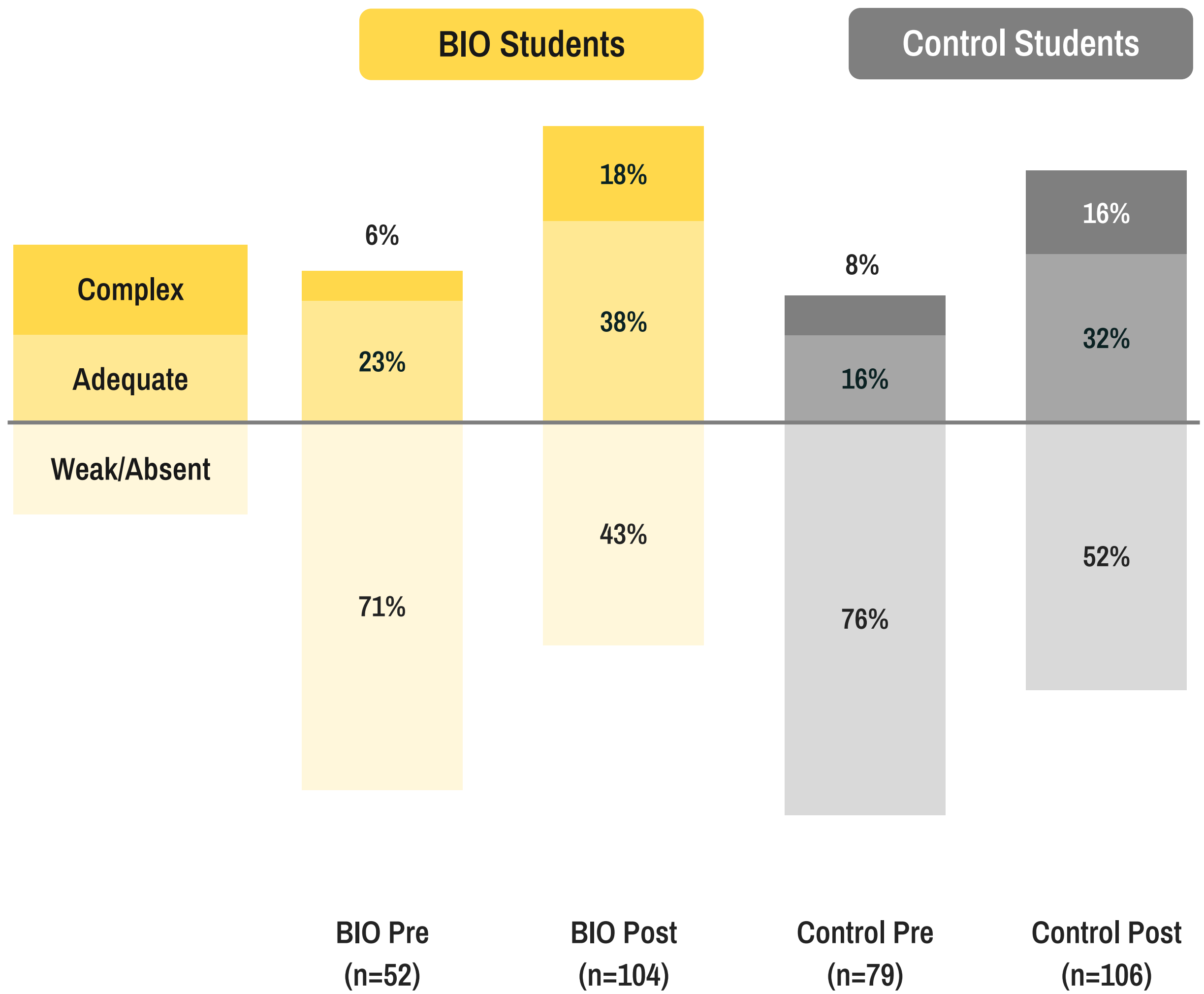
Prior to starting BIO lessons, nearly 7 out of 10 instances of provided evidence in the BIO intervention group's data performance tasks were weak support. Students in the Control group started out with a slightly lower rate of providing adequate or complex evidence, compared to students who would go on to experience the BIO lessons (24% versus 28%, respectively).

By the end of the school year, adequate or complex evidence was used to support 57% of all inferential claims among students that had experienced BIO lessons, versus only 48% among the Control group.

Overall, complex and adequate support tended to increase after the school year while weak support decreased in both the BIO and Control groups.

Differences in students' use of evidence before and after participating in two units of BIO lessons.

The graph below shows the percentage of evidence qualities across each inferential claim made, with a maximum of three opportunities per student interview to make a claim and support it with evidence. The n for each pre/post group changed based on the quantity of inferential claims made.



Growth in Ability to Make & Support Claims

After experiencing BIO lessons, over half of students were able to make inferential interpretations with robust evidence, an increase of 40 percentage points from pre-test. This growth was significantly stronger than their peers in the Control.

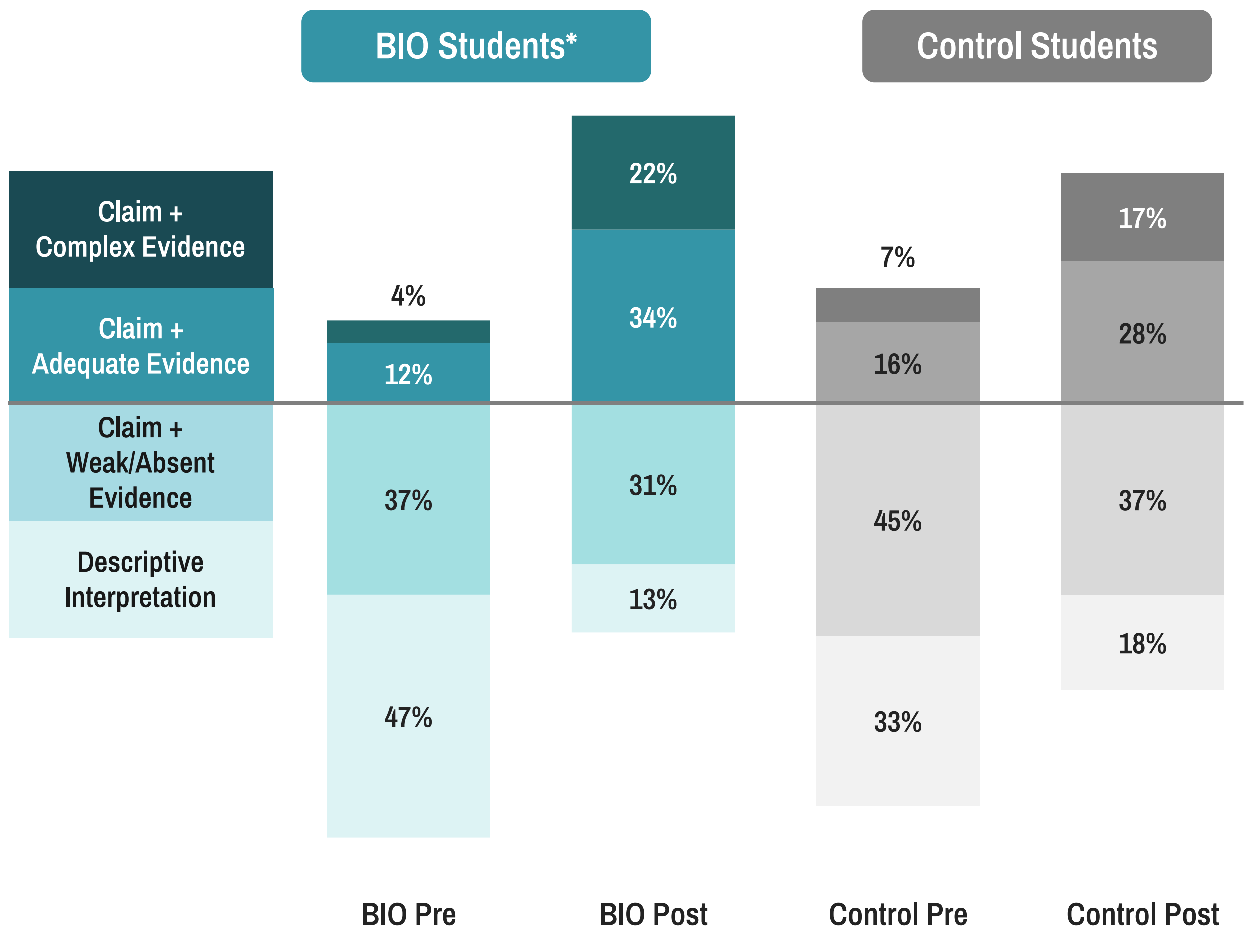
In the data performance task, students had three chances to make an interpretation and, if they made a claim, support it with evidence from within and beyond the data. Here, we analyze each student's "personal best" on that scale – the highest they scored on this rubric across three opportunities. (All students made at least one interpretation attempt.)

Prior to BIO, nearly half of all students stopped at describing the data, and just 16% made inferential interpretations with adequate or complex evidence. But this spiked to 56% of students using robust evidence to support inferential claims, and only 13% who stuck to descriptive interpretations.

While the Control group saw similar overall patterns in the shifts of their scores, the gains made following BIO instruction were significantly more dramatic and ended up substantially higher than the Control group.

Changes in highest level achieved (out of 3 attempts) of making interpretations and supporting with evidence from pre- to post-test, comparing both groups.

In the graph below, each student is reported based on their most advanced interpretation and use of evidence during the data performance task, out of three tries. The percentages represent how many students fell into each level as their "personal best" during the task. (n=68 BIO group; n=76 Control group)



*Students receiving BIO teaching showed significantly greater growth in ability to make claims and support them with strong evidence than the Control students (p<.05)

Changes in Individual Students' Skills

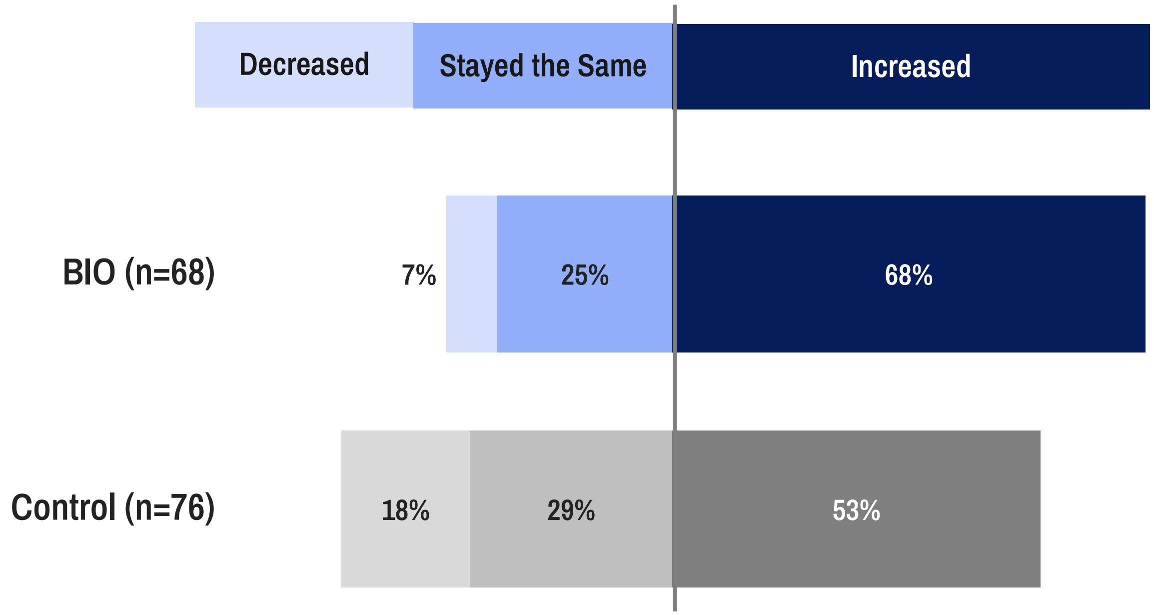
After participating in BIO lessons, 68% of students improved their ability to make claims with data or better supported claims with evidence, compared to just 53% of students in the control group.

Among those students who showed increased skill in their post-task, the average increase was 1.9 levels – which was the same for the Control and BIO groups. However, BIO students were more likely to make this upward shift.

Further, only 7% of students in the BIO group decreased their personal best interpretation score, compared to 18% of students in the Control group. It was expected that some students would have lower performance in the post-task, either due to day-to-day shifts in attention or energy, varied comfort with the data presented in each task (lightning vs. forests), or other factors. However, the BIO students seemed more consistently able to withstand those factors, and maintained their ability level in both times of testing. The control group students were far more likely to decrease. This control group also included four students that decreased scores dramatically, by 3 or more levels; no students in the BIO group showed such substantial decreases from pre to post.

Frequency with which students changed their highest level of interpretation/evidence score between their pre and post data performance task.

Each student was assigned to their most successful interpretation and/or use of evidence during the data performance task, out of three tries – their "personal best." The graph below shows the percentage of students whose "personal best" went up, went down, or stayed the same, comparing the control and BIO intervention groups.



*Students receiving BIO teaching showed significantly greater growth in ability to make claims and support them with strong evidence than the Control students (p<.05)

Pathways of Skill Change: BIO versus Control

Individual BIO students showed more consistent changes and growth patterns, while the changes among Control group students were more chaotic, with more students dropping levels and dropping further.

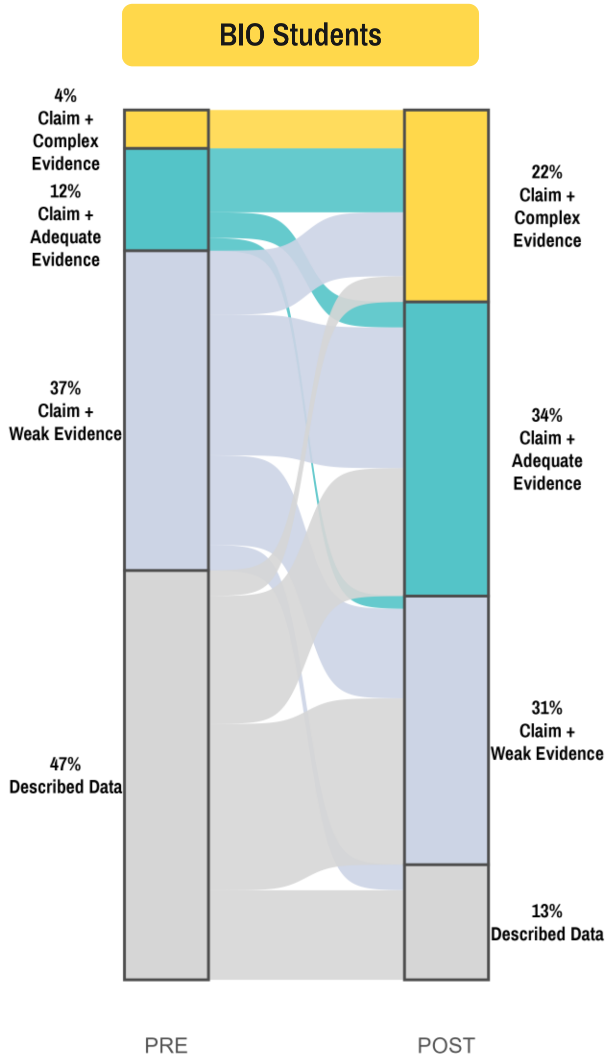
The diagrams to the right trace each student's pathway from pre-test to post-test, focusing on their "personal best" performance of interpreting from the data and supporting their interpretation with evidence.

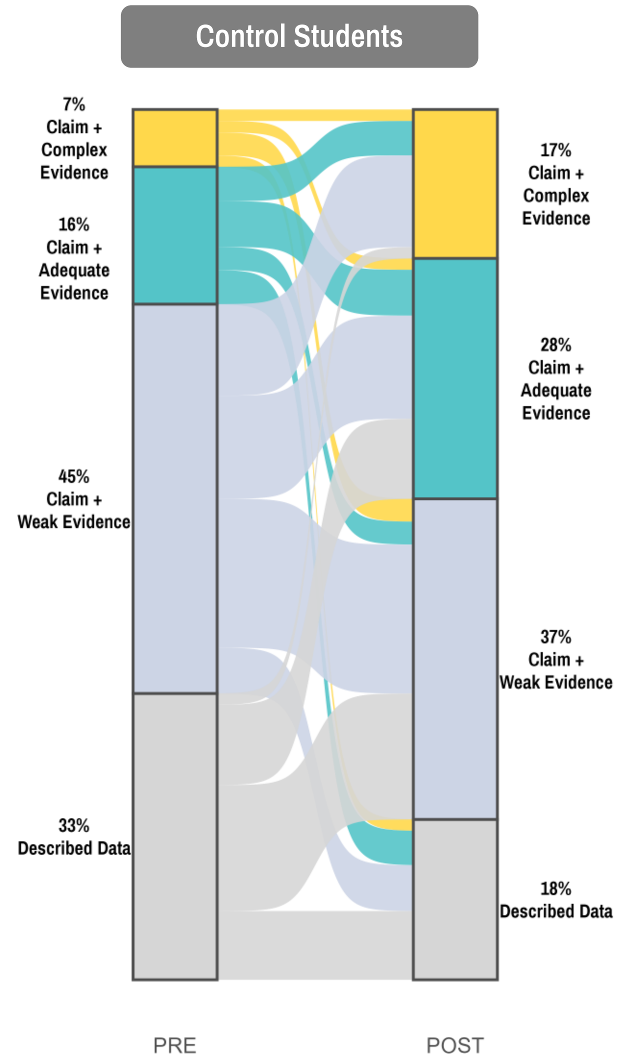
While the aggregate improvement in growth is strong on its own, this analysis highlights that the BIO treatment brought greater stability to increases and abilities to do this task. For example, all BIO students who started out using complex evidence continued to do so; while the majority of Control students who started out with complex evidence dropped to lower levels at post.

This reflects the skill gains with BIO. It may also reflect a broader, increased comfort in seeing and discussing geospatial data, that led to less variability in patterns in post-interviews.

Alluvial diagrams tracing the movement of each individual student from pre- to post-test in their "personal best" skill of interpreting data and supporting with evidence.

The data below represent each student's highest demonstrated level of ability in interpreting data and supporting it with evidence, out of three opportunities. (n=68 BIO; n=76 Control)





Variation in Observations Used for Interpretations

BIO students used different types of observations as support depending on whether they were making a descriptive or inferential interpretation. Students relied on data similarities, case values, and contrasts to support inferential claims. When simply describing data, they focused on aggregate patterns and decoding the meaning of colors.

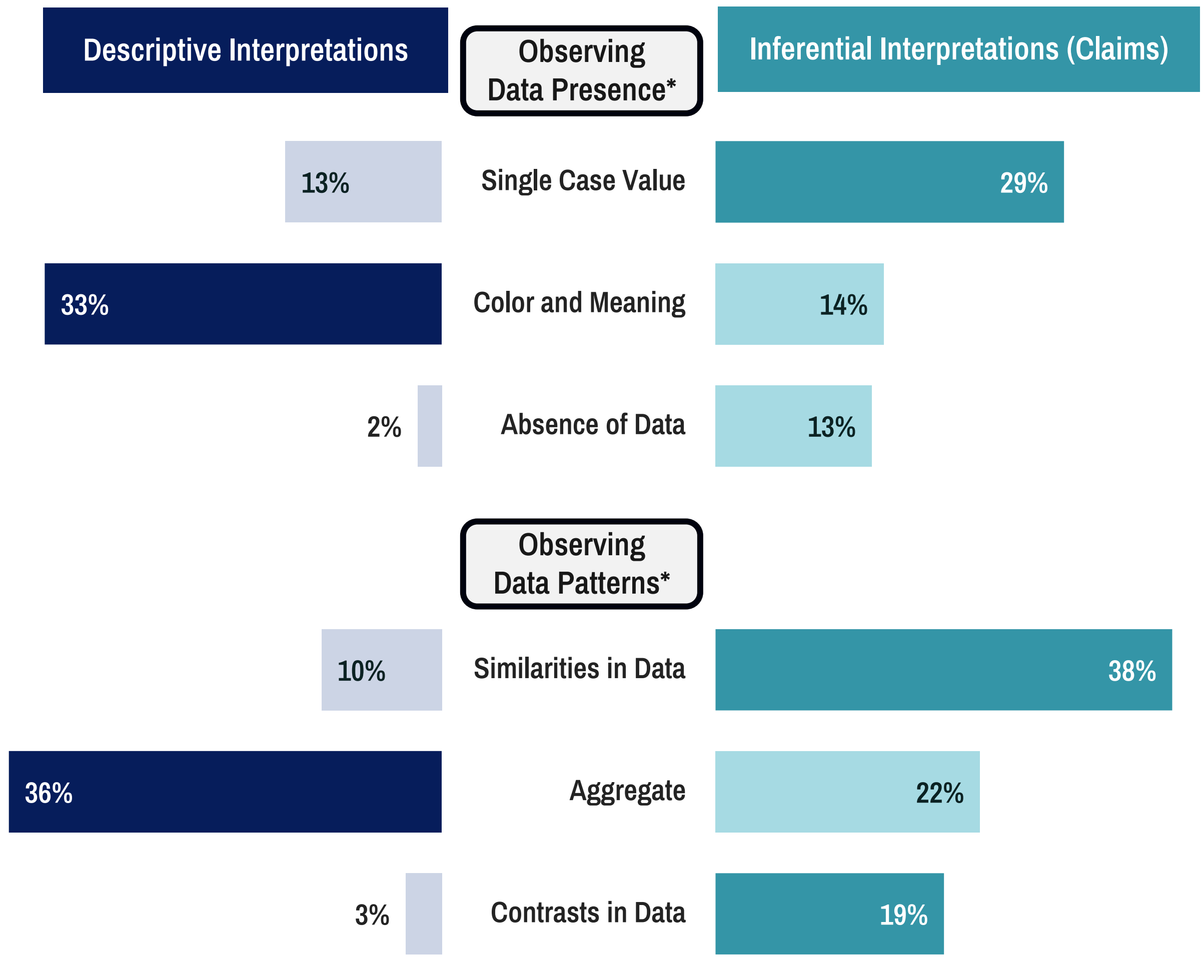
Making claims (inferential interpretations) tended to be supported with data patterns at a higher rate than descriptive interpretations. In descriptive interpretations (using only information contained in the data visualization), students more often noted aggregate patterns, such as which color was "the most" or "the least" present overall.

When students made claims about why data look the way they do, they used a greater variety of observations. Similarities, contrasts, and case values were most used as data support for these claims, with many students incorporating multiple observations to support their broader claim about the data.

Observation types that were less common are not shown here.

Comparing the types of observations students used to support their descriptive and inferential interpretations during the post-BIO data performance interview.

The percentages below represent how often a given type of observation was used among all of the descriptive and inferential interpretations made. (n=98 Descriptive interpretations; n=106 Inferential interpretations)



*Detailed descriptions of Observing Data Presence & Patterns codes are discussed in the Observing & Decoding section of this report.

Variation in Observations and Support Quality

Following BIO, students who provided complex evidence to support their claims overwhelmingly used observations of similarities in data points, along with case values, data contrasts, and design choices.

BIO students who provided adequate or complex support for claims seemed more likely to use data patterns as part of their support compared with BIO students who provided weak evidence.

Students who offered weak evidence were more likely to include an observation of the absence of color. These were often cases where students proposed a cause for the absence of color, but reached for fanciful explanations of why that was happening, not grounded in data or phenomena. These students also occasionally used similarities and case values in their thinking.

Notably, observations about design choice only occurred as part of a student constructing complex evidence to support their claim. It should be noted that observations about design choices were *also* occasionally used as part of descriptive interpretations, occurring among BIO students and the Control group alike.

Types of observations students used as evidence for claims, based on the ultimate quality of their evidence, during the post-BIO data performance interview.

The percentages below represent how often a given type of observation was made in relation to instances of each level of coded evidence to support an inferential interpretation. (n=19 instances of Complex Support; n=41 instances of Adequate Support; n=46 instances of Weak or Absent Support)

Complex Support	Adequate Support	Weak or Absent	Observation Type*
			Data Presence
42%	32%	22%	Single Case Value
5%	22%	11%	Presence of Color and Meaning
5%	7%	22%	Absence of Color
			Patterns and Comparisons
89%	27%	26%	Similarities in Data Points
42%	20%	9%	Contrasts in Data Points
21%	32%	13%	Aggregate Data Pattern
11%	0%	0%	Design Choices

*Detailed descriptions of each Observation Type code are discussed in the Observing & Decoding section of this report.

RESULTS

Observing & Decoding

BIO Students' Use of the Title & Legend

Each graph feature – the title and legend – was directly referenced (pointing, naming, or directing the interviewer's attention) by around two-thirds of students in post-BIO tasks. This was only a minimal change (and slightly lower) than BIO students' pre-tests.

Far more than two-thirds of students clearly used the title and legend in their process of interpreting data. For example, they used words and phrases only available in the title or legend. But not every student directly cited the feature as evidence of why they were saying what they were saying. It was likely they, like many adults would, assumed it was obvious they had used these tools.

However, the rate at which students directly cited the title and legend as tools in their interpretation did not meaningfully change from the pre-interview to the post-interview.

There was also no substantial difference between the BIO and Control groups at any stage of testing. This may indicate that basic decoding (and explicit use of) the title and legend was an already-established skill among the students, at least in this study.

Frequency of referencing or using information from the title or legend as evidence during BIO post-interviews.

Each student was given a binary code to indicate whether they directly referenced the title, legend, or both at any point during their data performance interview. (n=69)

Basic Decoding: Title & Legend Percentage of students that referenced the title or legend at any point in their interview.	PRE	POST
Used the Title as Evidence Student directly states or indicates that they used the title to get information, including reading the title aloud, pointing to the title, or using it as evidence.	71%	68%
Used the Legend as Evidence Student directly states or indicates that they used the legend to get information, including reading the legend aloud, pointing to the legend, or using it as evidence.	74%	64%

"Well, from the thing up here, the [title], it shows change in forests 2000 to 2014."

-Student 361, Post-BIO

"The legend on the left says no change, gain, loss, and gain and loss overlap."

-Student 165, Post-BIO

BIO Students' Applied Geography Knowledge

BIO students grew substantially in their ability to apply geographic knowledge – especially names of places and natural locations – as they reasoned about a dataset after their school year.

Maps used during testing were unlabeled – no political, natural, or map landmarks were identified with text. When names were used by a student in their interpretation, they were drawing on their outside knowledge. The most common way this was displayed was by identifying political locations and boundaries (countries, continents, etc.). Identifying specific natural features and geospatial orienting occurred, but were less common.

For political locations, in particular, BIO students increased use of geographic knowledge by 14 percentage points. Moreover, after BIO teaching, they demonstrated this skill far more often than their peers in the Control group (68% to 54%, respectively).

In Data Sketches, students worked closely with maps where countries and continents are labeled, which may have helped support growth in knowing where countries are and/or looking for data in those places.

Frequency of pointing or referring to specific locations or landmarks on the map during BIO post-interviews.

Each student was coded when they directly referenced or named geographic elements during their data performance interview. (n=69)

Basic Decoding: Applied Geography Knowledge Percentage of students that referenced a geographic feature at least once at any point in their interview.	PRE	POST
Identified Political Location Student accurately refers to or names a specific location with a political boundary, such as countries, continents, states, or regions.	54%	**68%**
Identified Specific Natural Location Student accurately refers to or names a specific natural feature of location (such as the Amazon rainforest, the Himalayas, the Rockies, the Sahara desert, the Atlantic ocean, etc.)	9%	**17%**
Identified Map Landmarks Student names specific geospatial orienting features or landmarks, such as the equator, hemispheres, latitude, longitude, or the poles.	17%	**17%**

"I noticed that there's more forestry down here like the south of Africa than in the northern part of Africa."

-Student 451, Post-BIO

"I see the countries with red are mostly located near the Amazon rainforest…"

-Student 359, Post-BIO

"I noticed that there's a lot of forests in the upper Northern Hemisphere and that there is less forests in the end of the Southern Hemisphere."

-Student 155, Post-BIO

Changes in Applying Geography Knowledge

BIO students showed much greater growth in ability to identify political locations and map landmarks than the Control group.

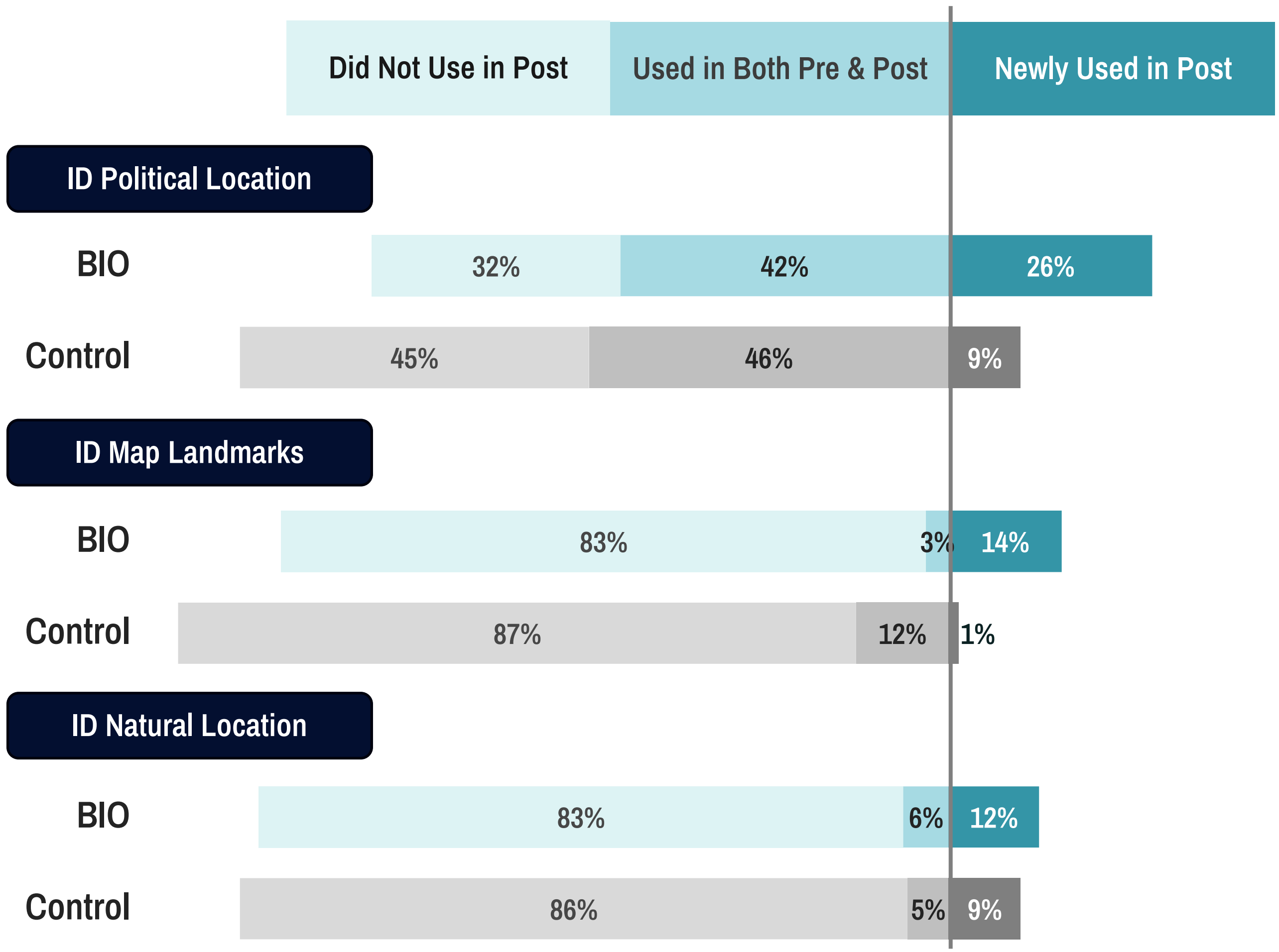
26% of BIO students correctly identified a political location at least once in their post-interview after failing to do so in their pre-interview. For the Control group, only 9% of students showed this growth.

And although naming map landmarks (e.g., equator, hemispheres, etc.) was a far less common skill, 14% of BIO students newly applied this skill at the end of the year, compared with only 1% of the Control group.

Identification of specific natural locations or biomes (e.g., the Sahara Desert) was a skill where BIO did not show far greater improvement. In both the BIO and Control groups, a similar percentage of students made gains in this area (12% and 9%, respectively).

Change in presence of applied geography knowledge types during the data performance interview.

The graph below shows the percentage of students that applied each category of geography knowledge at least once in their post interview after not doing so in their pre, used geography knowledge in both interviews, and students who did not use geography knowledge at any point in their post interview. (n=69 BIO; n=76 Control)



Changes in Frequency of Applying Geography

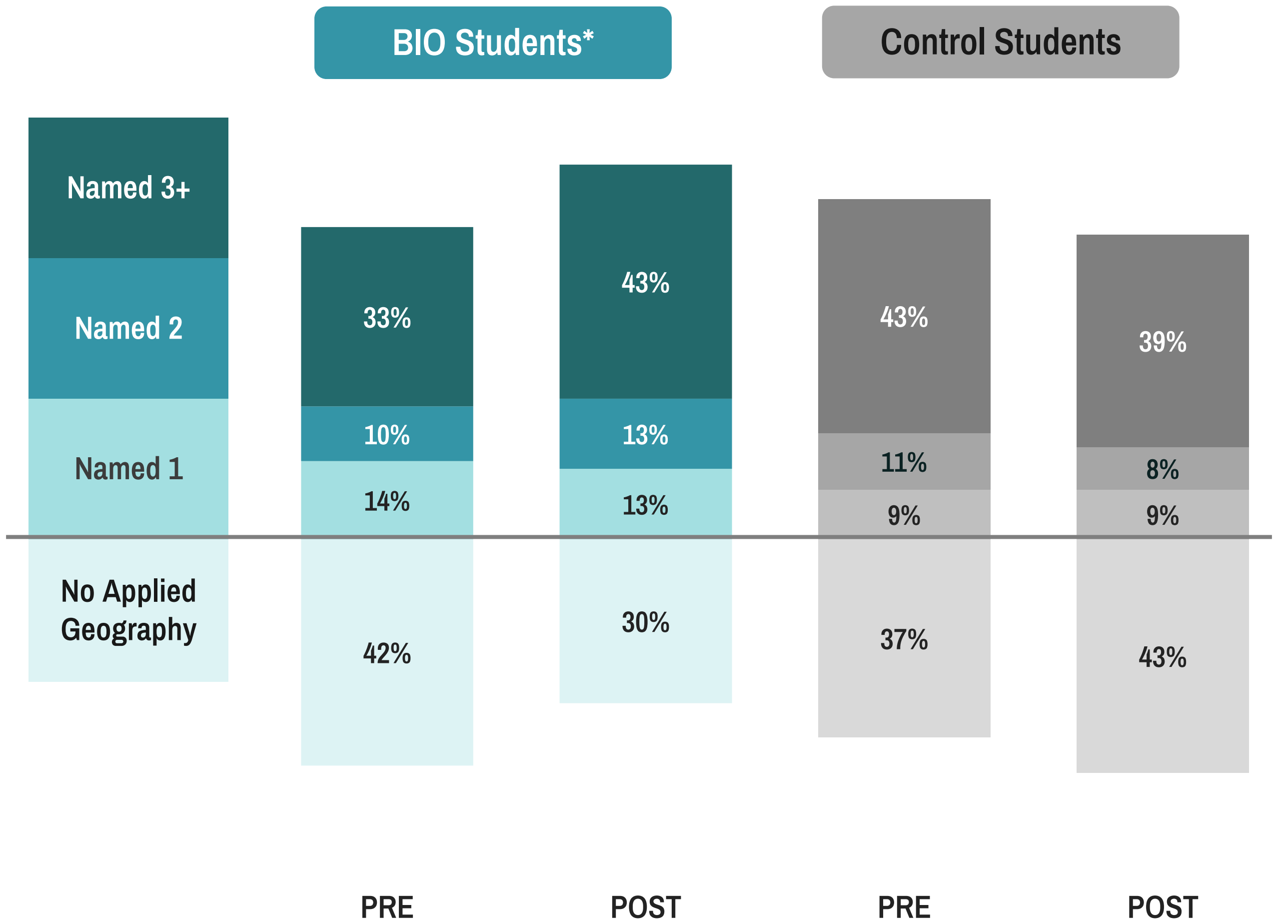
BIO students also applied geographic knowledge more frequently during their data performance interviews, compared to their pre-test. 10% more BIO students named 3 or more specific locations in post-interviews than in pre-interviews. Control students decreased slightly.

While Control students started their year with a higher percentage of students who made three or more geographic references (including identifying political regions, natural locations, or map landmarks), frequency of geographic mentions seemed to decrease by the end of the school year. Only 39% of Control students hit the 3+ level, and 43% failed to apply any geography knowledge.

BIO students, in contrast, entered with less prior skill use, with only one-third making geography references at the beginning of the school year. But by the end of the school year, after experiencing two BIO units, 43% of students named more than three geographic elements, and only 30% of students made no geographical references at all.

Change in frequency of applied geography knowledge during the data performance interview.

The graph below shows the percentage of students that applied any type of geography knowledge by frequency. Students that named more than 3 geographic elements are grouped here for clarity; the highest number of geographic mentions in a single student interview was 41. (n=69 BIO; n=76 Control)



**BIO students' change from pre to post is significantly greater than the Control group (p<.01).

BIO Students Observing the Presence of Data

When students observed the presence of data after BIO, they most often either noted the colors and meanings from the key or called out the data at a specific point (case value) on the map. These observations were limited to data presence, and did not extend to broader patterns of data.

Noting presence of data and associating it with its meaning was the most common skill exhibited post-BIO. This was also an area of great growth – with 21% more students doing this post-BIO.

This skill was also an area of growth for Control students, which may have to do with affordances of the legends used in the pre- and post-test maps. However, BIO students' gains were greater, compared to the Control group's gains. (This is discussed further on p.34.)

Alongside this, there was a substantial decrease in BIO students who made observations about the simple presence of data – the least specific observation in our coding system, where they only stated only that there was color on the map. This suggests students got better at moving more quickly to meaning alongside color presence.

Change in frequency of different ways of describing data presence or absence terms used by BIO students in their pre and post tests. Highlighted cells show increases.

Each student was coded when they noted the presence or absence of data without noting a specific location, larger pattern, assigning meaning, or anything more specific. (n=69)

Describing Data Presence: General Features Percentage of students who made one or more observation that was limited to describing general data presence or absence without noting any larger pattern (in that segment).	PRE	POST
Presence of Data Student observes that the data representation uses the presence of colors to visualize something, without assigning meaning or noting anything more specific.	35%	19%
Absence of Data Student notices that non-specific areas of the map are black without noting any specific location or larger pattern.	16%	36%
Absence-Presence Comparison Student states a contrast between the presence of color/data and the absence of color/data on the map, without being specific about where it occurred, which patterns they noticed, or any other details.	17%	4%
Presence of Data + Meaning Student reads or interprets the legend; identifies the color(s) and what they mean without noting any larger pattern.	41%	62%
Single Case Value Student notes or describes color/data in a single location, with or without interpreting what that might mean or relate to.	33%	51%

What it Sounded Like: Describing Presence of Data

Data Presence + Meaning

"[The legend] says "forest loss," and then there's a bunch of red on here."

-Student 154, Post-BIO

Single Case Value

"That it looks a lot like in South America in the center right here that there was a lot of loss that was going on."

-Student 160, Post-BIO

Absence of Data

"I do see that there's a lot of areas that are blacked out."

-Student 368, Post-BIO

Presence of Data

"I see a bunch of green."

-Student 657, Post-BIO

Presence/Absence Comp.

"There's a lot of places that have forest, but there are some that don't have any at all, not even a small amount."

-Student 363, Post-BIO

Changes in Observing the Presence of Data

In this group of observation types, BIO students showed greater growth in newly decoding color meanings from the legend (without further interpreting) and describing single case values, as compared with students in the Control.

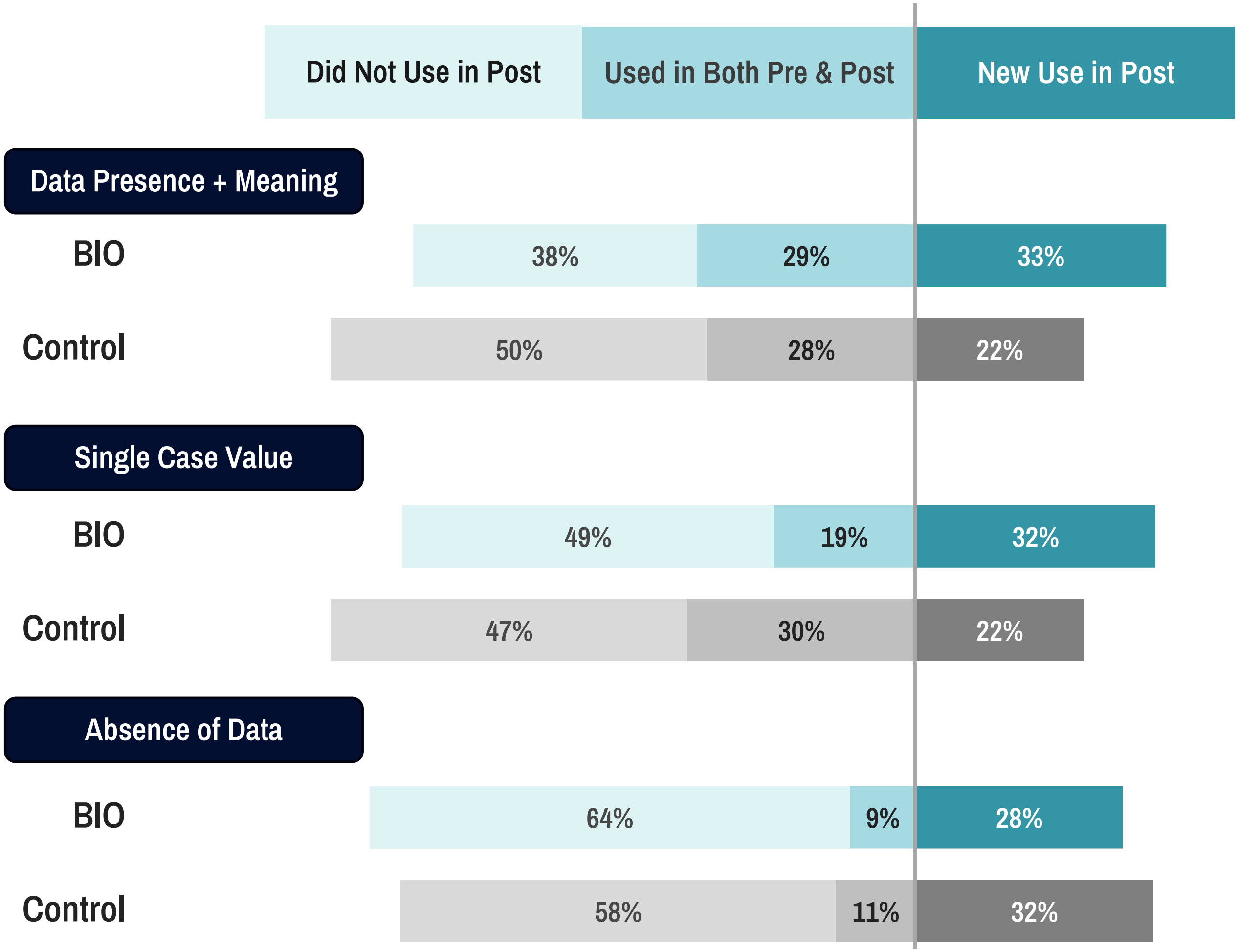
33% of BIO students newly made an observation that described the presence of data with its meaning at some point in their post-interview. Control students also increased in this skill, but only 22% of students in the Control group did the same.

The pattern of growth in skills was very similar, with nearly a third of BIO students noticing case values for the first time in their post-interview, and only 22% of Control students showing this growth. However, more Control students used this skill in both pre- and post-interviews.

An observation that simply noted the absence of data was an area of gain, but the growth did not seem to be influenced by BIO any differently than regular instruction. Changes in other types of observation of data presence are not shown; there was little difference between groups and little growth from pre- to post-interviews in any of these.

Change in noting the presence of data during the data performance interview.

The graph below shows the percentage of students that noted presence of data at least once in their post interview after not doing so in their pre, noticed presence of data in both interviews, and students who did not note simple data presence at any point in their post interview. (n=69 BIO; n=76 Control)



BIO Students Describing Patterns in Data

Post-BIO, the most common types of patterns students observed were aggregate patterns about the whole map and similarities in multiple specific points. Each skill was used by the majority of BIO students, with aggregate thinking showing the greatest growth.

Descriptions of aggregate data patterns saw the most change post-BIO, jumping from only 16% of students in the pre-interview to a whopping 71% in the post-interview. However, this jump was mirrored in the Control group, suggesting that this skill may have been acquired outside of BIO.

About a quarter of BIO students made observations that referred to a design element in the data representation. These observations referenced the stylistic choices behind the visual. This was a major growth for BIO students.

A few students described highly nuanced patterns in the data, such as co-occurrence or distribution of colors/data. This was, interestingly, a skill that decreased between pre and post for both groups. The use of aggregate more often and within data patterns less often seem likely attributable to affordances of the two data visualizations used.

Frequency of describing observed patterns in the data during BIO post-interviews.

Each student was coded when they noted a type of observable pattern in the data, either as an observation on its own or in support of a claim.

Describing Data Patterns: Similarities, Contrasts, Aggregate Percentage of students that described a pattern they noticed in the data at least once during their data performance interview.	PRE	POST
Aggregate Data Pattern Student describes an aggregated pattern about the data that applied to the whole map, often using words like most, least, majority, more, and less.	16%	71%
Similarities in Data Student describes a similarity in data pattern in two or more locations on the map, either through naming or pointing.	59%	61%
Contrasts in Data Student describes a difference or contrasting relationship between what is visible in two or more locations on the map, either through naming or pointing.	39%	36%
Design Element Noticing Student makes note of symbology, stylistic choices, or representation types to interpret meaning or describe imagery.	6%	23%
Relationship within Data Student describes a pattern, nuance, or relationship about how the colors/data appear on the map, including distribution and proximity of colors/data.	16%	10%

What it Sounded Like: Patterns and Comparisons

Aggregate Pattern

"See, there's mostly 'forest no change' where everything is just staying there, and that's what I see most in this map."

-Student 153, Post-BIO

Similarities in Data

"The top half of Africa has nothing. Same with a lot of Australia. It's just really blank right here and here. Same with this area. Really blank. … Probably [means there are] no forests, because there's nothing in Greenland also."

-Student 165, Post-BIO

Contrasts in Data

"I notice that in Africa, the northern parts of Africa are completely black. If you go more towards the south, there's a lot more color."

-Student 460, Post-BIO

Design Elements

"I don't know how they chose the colors like this, but the colors could represent where it's purple and red as bad things that happen when people are there, and green shows that environment is ... It's much better places where there are no people."

-Student 359, Post-BIO

Relationships w/in Data

"There's forest loss in a larger section of green, which is just normal forest. So [forest loss is] a bit more spread out than just normal forest, I'd say, or overlap."

-Student 353, Post-BIO

Changes in Describing Patterns in Data

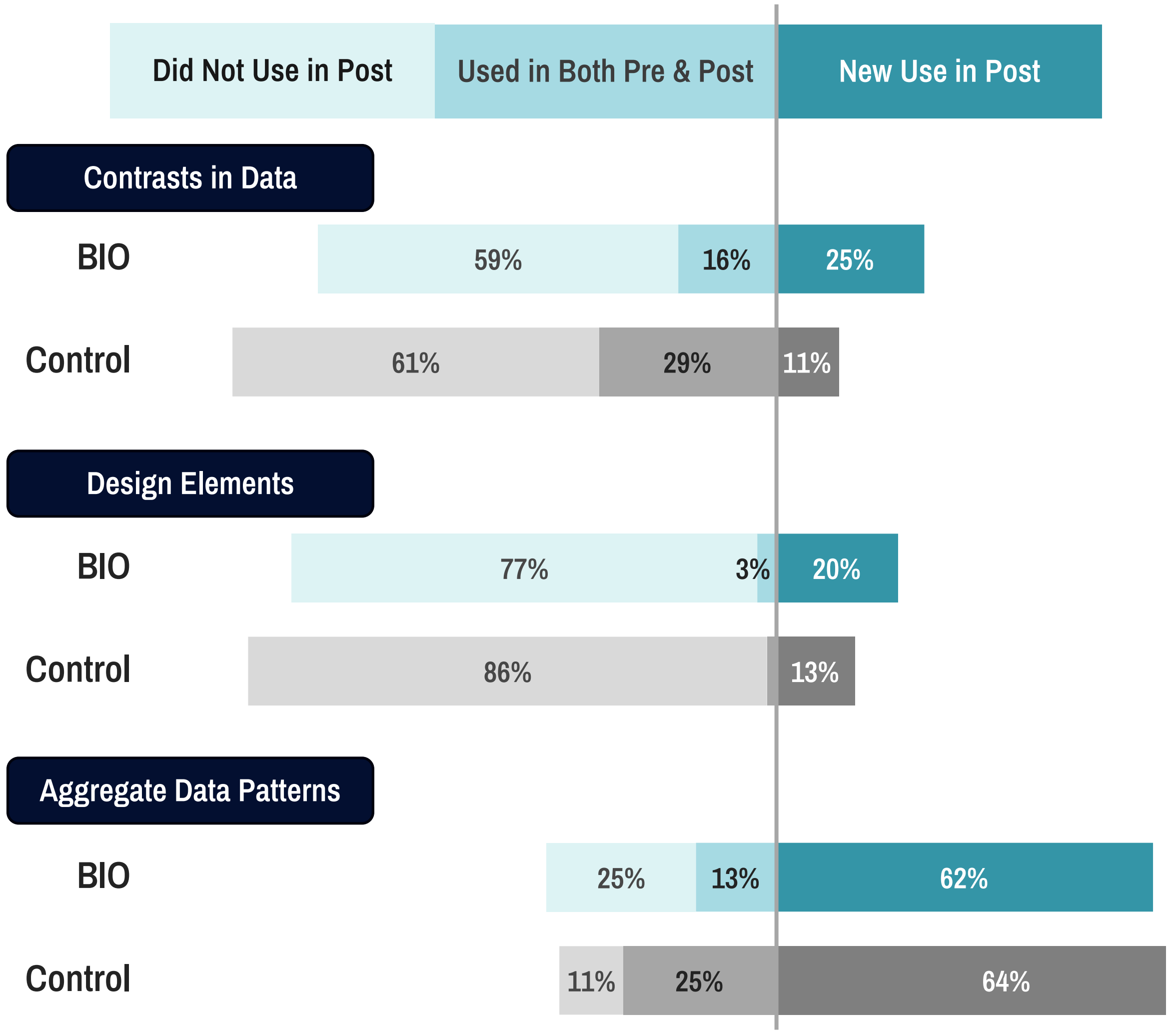
BIO students showed much greater growth in observing contrasts in data points, compared to their Control group peers. While they ended at similar points, more BIO students only began using this skill at the post-interview.

There was a slightly higher growth in the skill of describing design elements within data representation, with 20% of BIO students newly doing this in post-interviews, versus 13% of the Control group. However, this tended to be an area of growth during the year for all students – or was an affordance offered by the post-test data visualization. Almost no students in either condition referred to design elements in the pre-interview.

In the other coded patterns, BIO students did not show markedly different growth patterns than the Control students. When it came to observing aggregate patterns, similarities in data, and relationships within data, the degree of skill growth from pre- to post-test was fairly similar for both groups.

Change in describing data patterns during the data performance interview.

The graph below shows the percentage of students that noticed patterns and made comparisons at least once in their post interview after not doing so in their pre, noticed patterns in both interviews, and students who did not note patterns or make comparisons at any point in their post interview. (n=69 BIO; n=76 Control)



Changes in Using Different Observation Types

There was a greater increase in the number of BIO students who applied each of the categories of observation – geography, data presence, and data patterns – compared to the changes seen in the Control group.

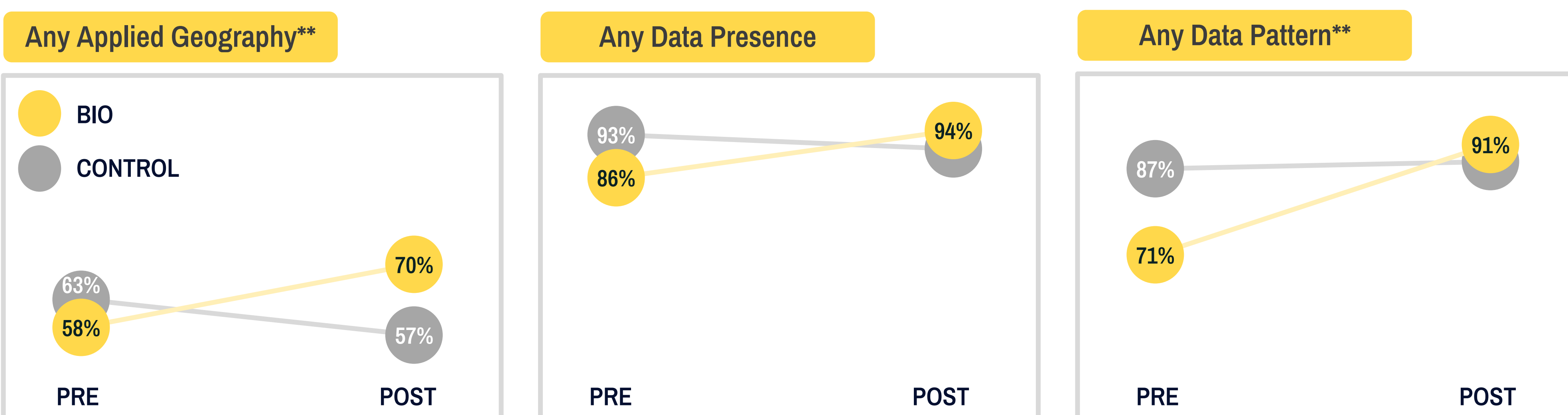
In all three categories, fewer BIO students used the skill at the beginning of the school year; but they still out-performed students in the Control group in the post-interview. A striking example is applied geography knowledge, where there was a 12-point increase in the proportion of students using the skill, compared with a 6-point decrease among Control students.

While a similar number of BIO and Control students made observations of data patterns at the end of the year (91% versus 88%), BIO students reached this point from a much lower proportion of students describing patterns in their pre-interviews. The Control stayed static, while students had a 20-point increase in noticing data patterns after completing BIO lessons.

The changes were less dramatic in observations made about the presence of data (without patterns). However, BIO students again saw an 8-point increase in the number who used this skill, while the Control students decreased slightly from pre to post.

Change in presence of observational types among students.

The graph below shows the average percentage of students that were coded at least once in each of the four observational parent codes during their data performance task. (n=69 BIO; n=76 Control)



**There is a significant relationship between treatment (BIO) and likelihood of newly making these types of observation at post-test (p<.01).

Changes in Total Number of Observations Made

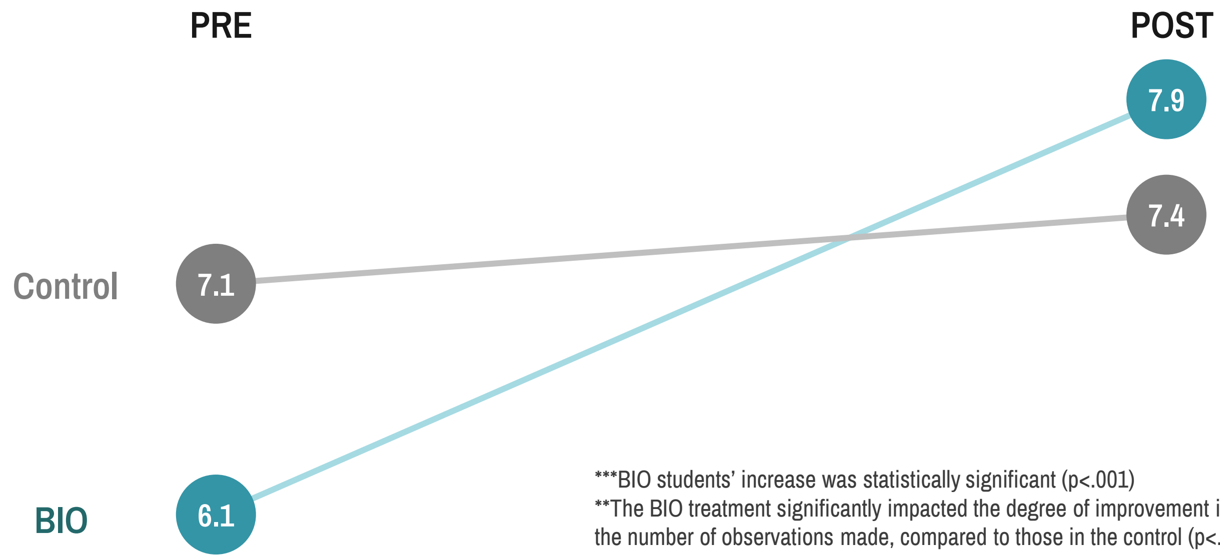
The BIO learning experience showed a significant impact on increasing the total number of distinct observations made by a student – particularly compared to the Control condition, where there was almost no increase in observations.

The graph to the right shows the change in average number of observations made by students in each group. Statistically, the difference in gains was significant – with the BIO treatment showing a significant effect on outcomes, far more than the Control group (which effectively did not change).

The finding was similar when we examined the diversity of *types* of observation made – in other words, when we only counted 1 point for each type of observation on the list (right). BIO students also grew much more in their use of a greater variety of *types* of observations. This means that BIO students grew to notice many different types of patterns – similarities, contrasts, aggregate, presence, etc. – as opposed to using just a few strategies (similarities and presence) and repeating them more often.

Comparing the change in average number of unique observations made during interviews between the two conditions.

Each data performance interview was given a score for the number of distinct observations made. The graph below shows the average number of observations made at each time and condition. Detail on the components of the score is outlined below. (n=69 BIO; n=76 Control)



***BIO students' increase was statistically significant ($p<.001$)
**The BIO treatment significantly impacted the degree of improvement in the number of observations made, compared to those in the control ($p<.01$).

Possible Components Comprising Observation Score

- Used Title as Evidence (1 pt)
- Used Legend as Evidence (1 pt)
- Identified Political Location(s) (1 pt)^
- Identified Natural Location(s) (1 pt)^
- Identified Map Landmark(s) (1 pt)^
- Observed Presence of Data (1 pt)
- Observed Presence of Data + Meaning (1 pt)
- Observed Absence of Data (1 pt)
- Observed Presence / Absence Contrast (1 pt)
- Case Value Observation (1 x distinct case value)
- Similarity Observation (1 x distinct similarity)
- Contrast Observation (1 x distinct contrast)
- Aggregate Observation (1 x distinct aggregate)
- Within Data Relationship Observation (1 x distinct pattern)
- Design Element Observation (1 x distinct feature)

^Geographical knowledge varied so widely between individual students (e.g., range of places named was 0-40+), including each named place as a unique observation skewed scores to primarily reflect geographical knowledge. To minimize this effect in scoring, we counted each type of geographical knowledge as one unique observation, regardless of how many individual places were named.

Comparing Level of Growth in Total Observations

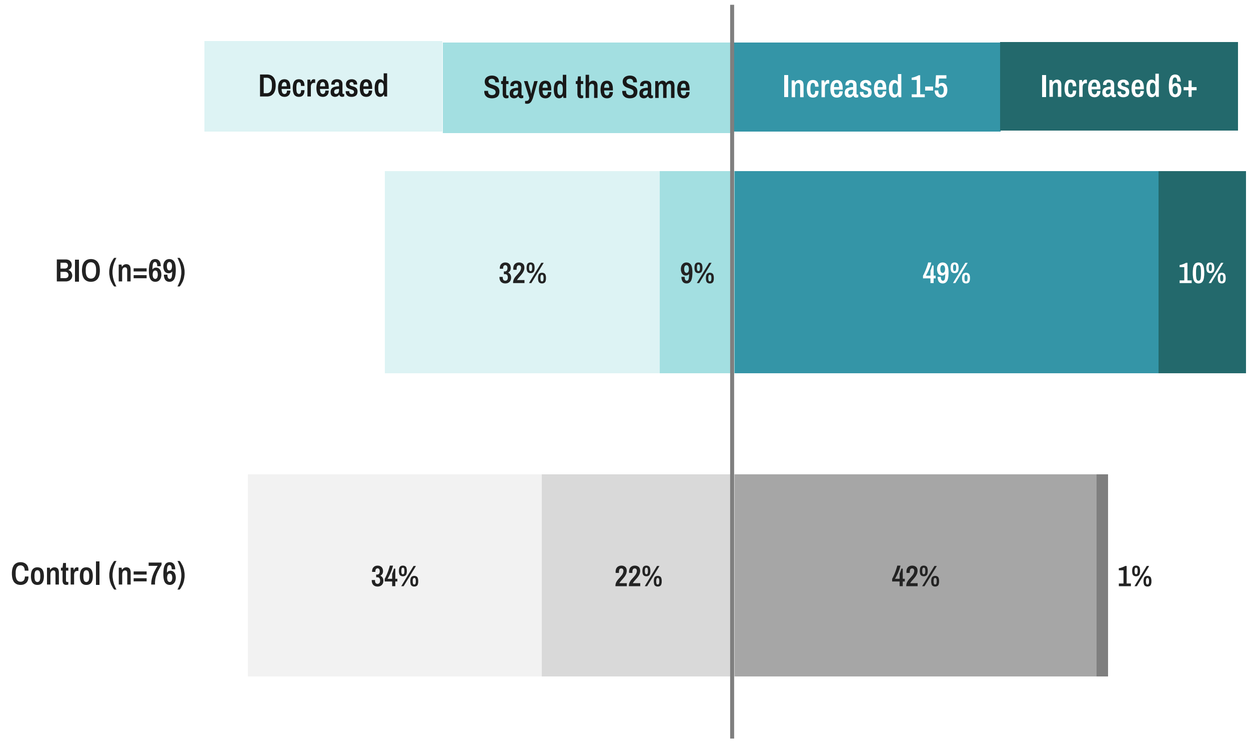
Over half of BIO students increased the number of distinct observations made in their interviews, with 1 in 10 making an additional 6 or more total observations than they did in their baseline interview. The increases seen in Control students were far more modest.

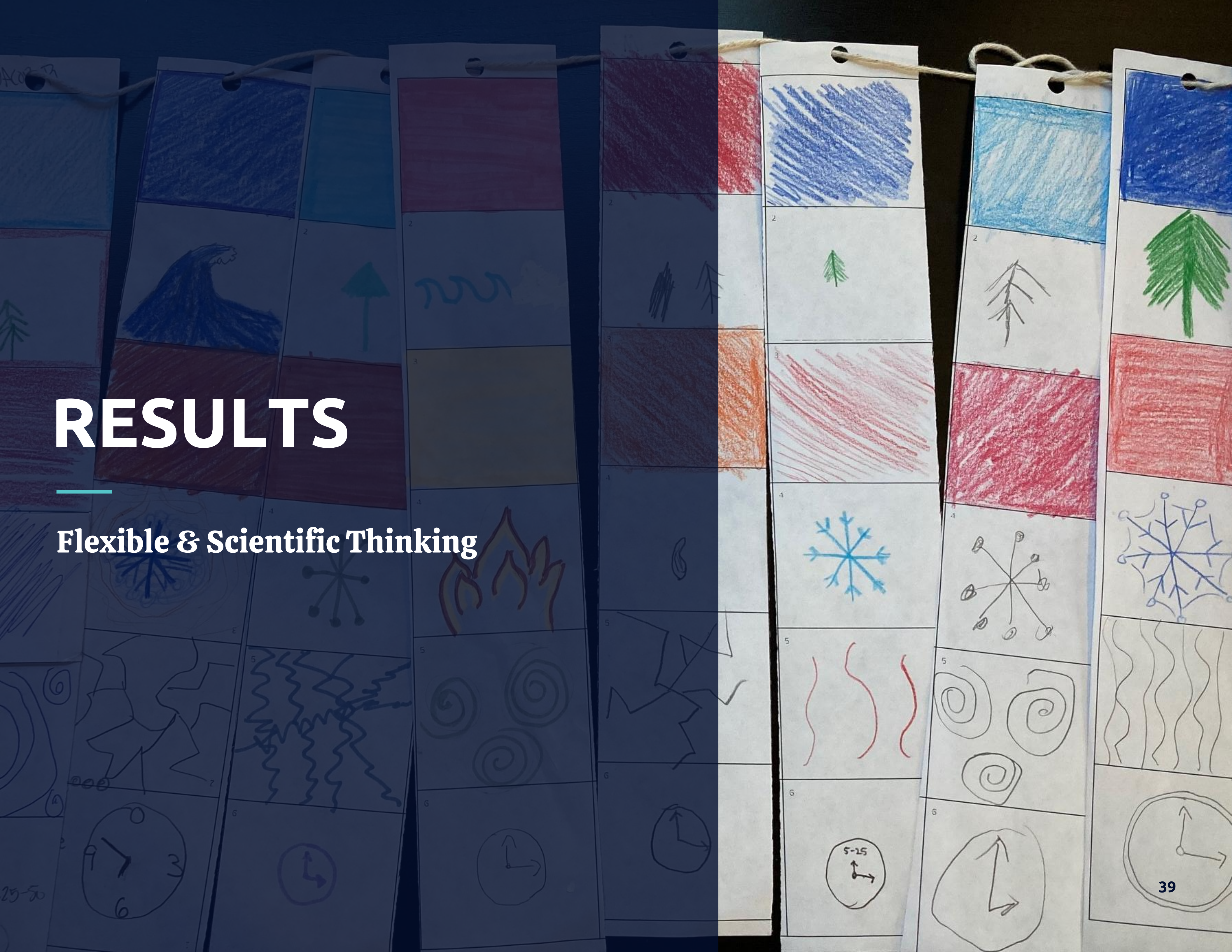
To unpack the change in total observations a bit more, the graph to the right shows the distribution of how much individual students' scores changed from their pre- to post-test. 59% of BIO students increased their total observation score at the end of the school year, compared to only 43% of the Control students.

Most notably, it was exceedingly rare for a student in the Control group to increase their score by 6 or more points between pre- and post-test. But 1 in 10 BIO students made this level of increase. **One BIO student made 17 more observations after their BIO learning experience**. The greatest increase any student in the Control group showed was making 8 more observations.

Degree of change in average scored observations during the data performance interview.

Each data performance interview was given an observational score based on the number of unique observations made by each student. The graph below shows the average number of scored observations per student in each condition. The scoring rubric is shown in the table below. (n=69 BIO; n=76 Control)





RESULTS

Flexible & Scientific Thinking

Increased Ability to Think Flexibly Post-BIO

After experiencing BIO lessons, a majority of students demonstrated flexible thinking, with 26% able to generate 3 or more possibilities of what the visualization could be trying to show.

Flexible thinking was considered as being able to identify 2 or more distinct possibilities of what data could show, which was achieved by 70% of students post-BIO. Most often, students had one core interpretation about the data's meaning, which they maintained until the interview gave a specific prompt about "anything else that it could be trying to show," when they were able to generate a second possibility.

Students who only provided a single synthesis stayed fixed on a single idea about the data's meaning, even after being asked for alternatives. Others struggled to interpret any meaning from the visual on their own, but could come up with just one when prompted for a synthesis.

It should be noted that this coding only measures how readily and flexibly students could generate multiple possible meanings for a single data visualization; it did not assess the plausibility or support for those interpretations.

Students' ability to generate multiple possibilities of what the data visualization could be trying to show people, from pre- to post-BIO.

The first half of the interview along with two targeted questions were reviewed and coded to identify how many distinct ideas a student produced about what the data might be showing. These syntheses were not scored for accuracy, plausibility, or quality – this set of codes is only meant to measure willingness to generate multiple possibilities. (n=68)

Degree of Flexible Thinking during Synthesis	PRE	POST
Multiple Syntheses (Level 4) Student provides three or more distinct possibilities of what the data could be showing. These syntheses could be stated alongside interpretations, or as a response to the two questions designed to prompt flexible thinking.	9%	26%
Added to or Revised Thinking (Level 3) Student describes two different possibilities of what the data could be showing: after making an initial interpretation, the student either offers another possible meaning or substantially revises their initial synthesis.	32%	44%
Single Synthesis (Level 2) Student provides a single consistent interpretation of what the data is showing throughout the interview and cannot come up with an alternative idea when prompted.	57%	26%
Does Not Synthesize / Opt Out (Level 1) Student cannot or chooses not to describe what they think the data is trying to show.	1%	3%

***BIO students showed significant growth in their flexible thinking skills from pre- to post-test ($p<.001$).

Students' Words: Flexible Thinking

Multiple Syntheses

"**Interviewer**: After all your looking at this image, what do you think it's trying to show people?
Student: I mean, I guess in a large sense **it's trying to show how the world has changed**, but you can also compare it with other sets of data to see how you might find **correlations between changes in trees and maybe the global surface temperature** or something like that. It could teach us more about **climate change**.
Interviewer: Can you think of anything else this data could be trying to show?
Student: ...If you look at places with lots of forest loss, **it could show more open space for new cities and buildings**.... Because I feel like nowadays people are trying to find more space because there are more people on the planet, and I feel like we're losing space. And **maybe it's for agricultural purposes or just residential purposes**."

-Student 462, Post-BIO

Added/Revised

"**Interviewer:** After all of your looking at this image, what do you think it's trying to show people?
Student: I think it's trying to show people the cause and effect of **what the ecosystem of trees looked like back then 10 years ago versus now**, because I feel like there would be a lot less green in a data map now than a data map from this data, like 10, 20 years ago.
Interviewer: Can you think of anything else that this data could be trying to show?
Student: I think this data could be trying to show, [it could] kind of give people confidence that **forests aren't going anywhere, and we don't really need to be as worry-able about forests**, even though forests are pretty good for the environment. ... I think that's a possibility because it's a lot of forests with no change and a lot of forest with an overlap rather than a forest with loss."

-Student 659, Post-BIO

Single Synthesis

"**Interviewer**: After all of your looking at this image, what do you think it's trying to show people?
Student: How the forest is. How they collected data over the past 14 years, and they're putting it all together to show us **where it's grown and where it's decreased**.
Interviewer: Can you think of anything else that this data could be trying to show?
Student: **I can't really think of anything else**."

-Student 354, Post-BIO

Comparing Student Gains in Flexible Thinking

Both groups saw increases in flexible thinking skills by the end of the school year, but BIO students showed slightly higher rates of making multiple distinct syntheses from the data when prompted, compared to the control group.

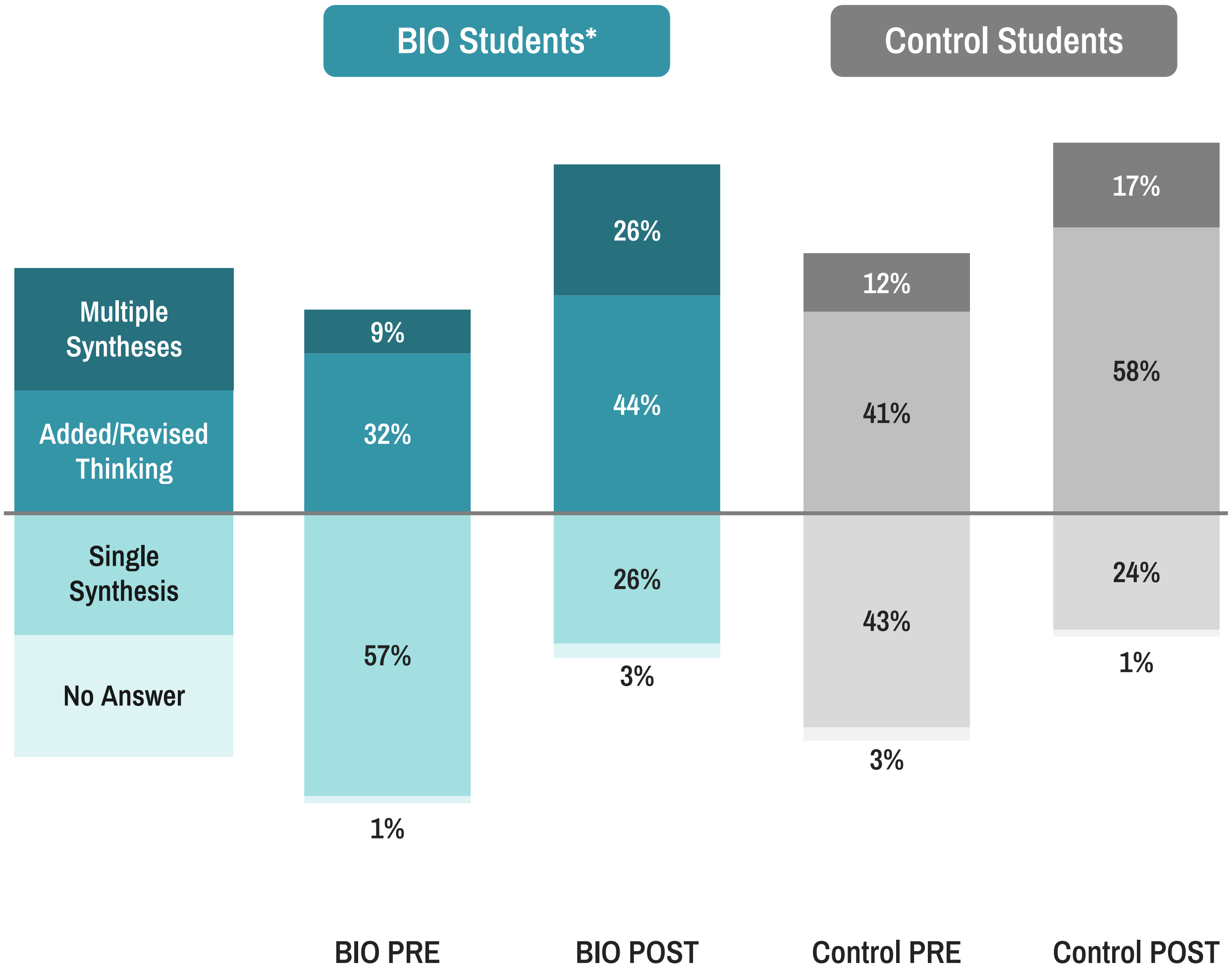
Both groups of students saw increases in flexible thinking from pre- to post-test, particularly shifts from sticking to one synthesis to generating at least one additional interpretation. However, students in the BIO group appeared to more regularly come up with three or more possibilities of meaning during the data performance task, as compared with the control group, who more typically stopped at two possibilities.

Students who opted out of making any synthesis about what the data was trying to tell people were rare and did not show any change.

Statistically, the relationship between treatment and the level of change from pre to post was not enough to be significant. However, this skill does show promise, and could be worth more targeted investigation in a future study of student skill development from arts-based teaching methods.

Degree of flexible thinking exhibited throughout the data performance task.

The percentages below represent how many students fell into each level of flexible thinking. These include the number of unique ideas about the data's meaning that students brought up as they made observations and interpretations, as well as their responses to two questions later in the interview: asking them what the data was trying to show people, and if they could think of anything else the data could be showing. (n=68 BIO group; n=76 Control)



Students' Scientific Questioning Pre- and Post-BIO

After experiencing BIO lessons, students slightly increased their requests for additional data or information and/or for alternate ways of representing or presenting the data. But overall, the types of questions remained unchanged.

In post-BIO interviews, students who requested new or additional data or context most often suggested additional data sets that might be correlated with the data they were looking at. Requested data was often related to the dataset they were familiar with after doing Data Sketches in class.

Another common question was wondering about the scientists' purpose or intentions in gathering and/or visualizing the data and/or how the data had been collected. This did include wonderings about whether the data were reliable or collected consistently across the world.

Other students wanted to see the data visualized differently, for example in a table, or with different information highlighted (such as key place names). Some students wondered what the same data would look like at a different time, or from a different visual perspective.

Comparing frequency of the types of scientific questioning and thinking students posed before and after experiencing BIO lessons.

Scientific questioning was identified throughout the student interviews as it occurred. Late in the interview, students were asked specifically if they had any questions for the scientists that made the visualization, and if there was any other information that might help them figure something out about the data. (n=68)

Scientific Questioning of Data	PRE	POST
New/Additional Data or Information Student suggests or asks for additional data, information, or context beyond what the representation contains, including that could be layered on the current data	21%	28%
Scientists' Purpose Student wonders or asks why the scientists collected the data or how it might be useful to people.	21%	24%
Data Collection or Source Student wonders or asks about where the data came from, how the data were collected, how one would measure a given phenomena, potential for error, etc.	21%	22%
Alternate Representations Student suggests or asks for a different way of representing the data; makes comments on graphic design choices in the context of data.	6%	16%
Same Data, Different Conditions Student wonders what the same data would look like under different conditions (e.g., year, season, area, zoom in/out, etc.).	9%	10%

There were no significant differences or differing patterns between BIO and Control students in the types of questions posed.

Students' Words: Scientific Thinking

New/Addl. Data or Info

"I would probably ask them why this map is showing forests and only forests, but not also the population in different places. Like a map overlap, which would make a better understanding of where forests and people relate."

-Student 359, Post-BIO

"[I would want to know] what [the scientists] meant by the gain-and-loss overlap, and if it just meant they lost a lot and gained a little bit or gained a lot and only lost a little bit or if it just evens out pretty well. … Because if the trees or if the forest loses a lot of trees every year and only gains a few, then they're still losing forest, but it's just less significant because they're growing trees back. But if they gain a lot of forest and only lose a little bit, then it's going to be really crowded, and it could possibly be a hazard because there's not enough nutrients and soil to be used for all the trees."

-Student 155, Post-BIO

Scientists' Purpose

"Maybe [I want to know] what was [the scientists'] intention with the data? … Because I think data can be interpreted in a lot of ways, especially if there's not much context. So, it'd be interesting to see what they were doing."

-Student 455, Post-BIO

Different Conditions

"[I want to know] what did it look like in 2000? How did it change over time? … Because I feel like that'd be a very interesting comparison to see how much it changed and just to see if there was forest loss in the same areas or if that happened more in the later years."

-Student 160, Post-BIO

Data Collection or Source

"[I want to know] how they were able to get the data, because I know it's over 14 years, which is a pretty long time, how they were able to measure how much change there was. … Because I feel like how old I am, the world seems like a really big place, so being able to have all this data from all around the world is sort of really cool."

-Student 462, Post-BIO

Alternate Representations

"I mean, it's kind of weird because it doesn't show [black] right here [in the key], so I kind of just want to like, "Oh, what does that mean?" I want to know more."

-Student 257, Post-BIO

Students' Requests for Explanation Pre- and Post-BIO

When prompted about their questions at the end of the interview, students regularly asked for aspects of the data visualization to be explained to them, consistently in pre- and post-interviews.

Students coded into these two categories wanted something explained to them about what was shown in the visualization, what caused the data to look the way it did, and/or what an aspect of the representation meant, (e.g., color or word choice).

Requests for an explanation of meaning were very common, and most often occurred when students were asked if they had questions for the scientists who made the visualization. Some of these requests were from students that struggled to understand the visual, but others were from students who wanted to verify their own claims about the data made earlier in the interview. There was a small decrease in these questions post-BIO, but it was not substantial. Fewer students asked questions to have aspects of the representation or visual to be explained to them.

Many students requested explanation alongside scientific questioning strategies (see prior page); categories were not mutually exclusive.

Comparing frequency of fact-seeking questions and requests for an explanation about the data or representation before and after experiencing BIO lessons.

Late in the interview, students were asked if they had any questions for the scientists that made the visualization, and if there was any other information that might help them figure something out about the data. Requests for explanation were coded throughout the student interviews, although they rarely occurred before those two targeted questions. (n=68)

Requests for Explanation or Fact-Seeking Questions	PRE	POST
Requests Explanation of Meaning Student asks/wants the scientists who made the data representation to explain what is depicted in the data and/or what phenomenon is causing something they noticed.	53%	46%
Requests Explanation of Representation Student wonders or asks about a specific aspect of the visualization; tries to get clarity about what a particular measurement or word in the visual means (e.g., event, sq-ft, "loss").	19%	19%

"Maybe [I want to know] why there's no data in some places. Like if right here, there's all this data, and then, directly right then, there's none, what's the reason behind it? Why does it just stop?"

-Student 253, Post-BIO

"Make it more clear. Don't just make the title 'Changes in Forests'. Stop being vague. It's because of the vagueness that we have to go through all this. You can say 'Changes in Forests', but you're going to have to include, what do you mean by forest?"

-Student 253, Post-BIO

Comparing Treatments: Questioning Strategies

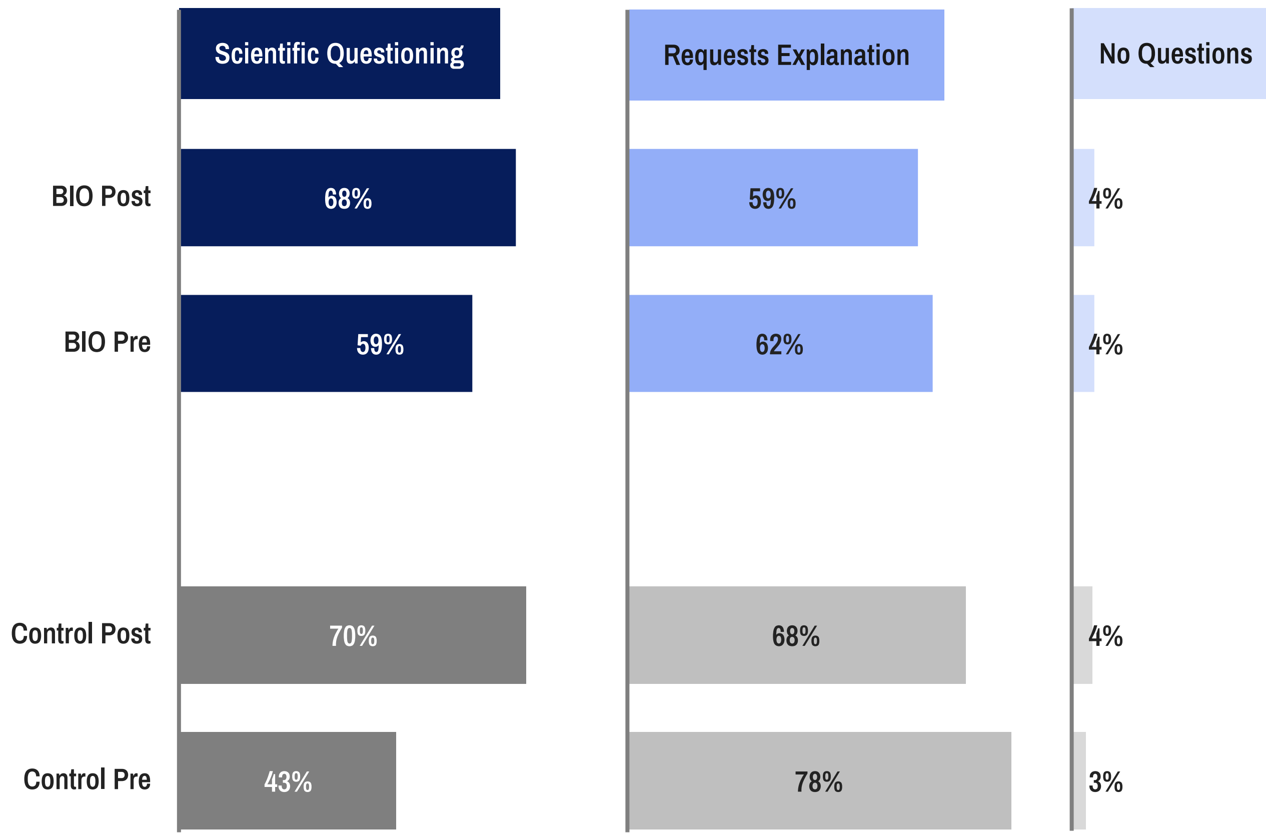
Rates of scientific questioning and requests for the data and/or visual to be explained did not change meaningfully after BIO. These results were similar to the patterns seen in students in the control group. This indicates that most students are already able to pose questions about data.

The types of questions students asked about data did not change significantly from pre to post, in the interview approach used here – including the use of scientific questions and simply requesting information. Similarly, rates of requesting information or explanation changed very little. There were also very minimal changes in the specific types of questions (sub-codes reviewed on the prior few pages).

Based on our data, when offered the opportunity to pose questions about data – particularly framed as "questions you'd ask the scientist" behind a visualization – is something most middle school students are able to do well. In future study, we would likely need to narrow the task to discern what, if any, questioning strategies are improved more by BIO instruction.

Rates at which students engaged in scientific questioning and/or requesting an explanation about the data at each time period and treatment condition.

The percentages below represent how many students engaged in some form of scientific questioning, requested information or explanation about the data and/or visualization, or chose not to ask any questions about the data. Students were able to engage in scientific questioning and request explanation; percentages do not add up to 100%. (n=68 BIO group; n=76 Control)



RESULTS

Self-Reported Learning

Teachers' Observations of BIO's Impact

Teachers felt BIO was most effective at helping students build visual literacy and engaging a broad range of students, including those that don't typically enjoy science topics.

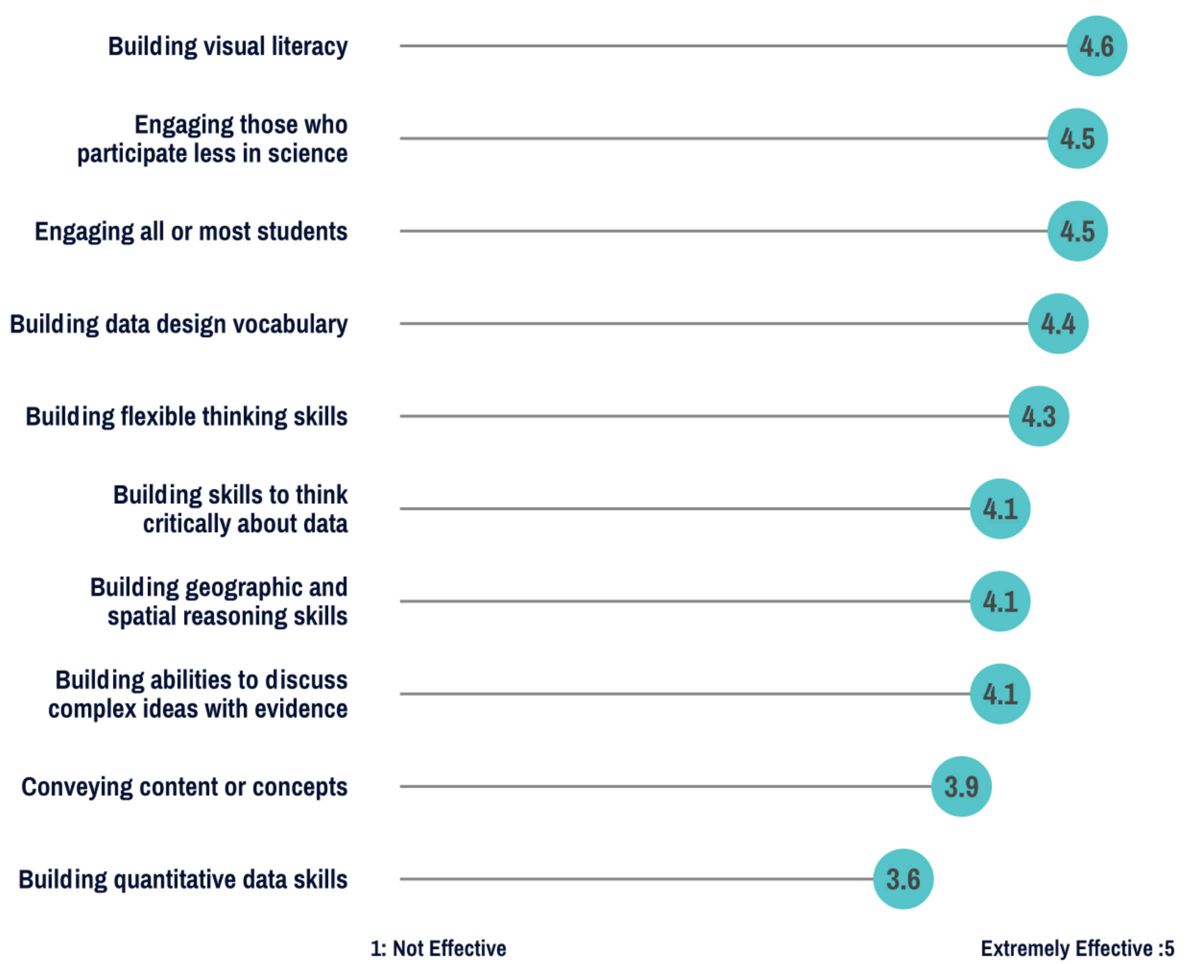
A majority of teachers felt the BIO approach had been either 'very' or 'extremely' effective in supporting student learning in each skill type provided in the survey question, with one exception. Teachers did not feel it was particularly good at building quantitative data skills – which it was not designed to do. It is notable that a few teachers even found it useful in that regard. It also may point to some degree of desirability bias, with "somewhat" effective ratings being as negative as teachers were willing to go.

BIO lessons seemed to excel at engaging students, encouraging flexible or divergent thinking, and developing visual literacy and design skills.

Teachers felt the BIO lessons were less effective at more traditional math/science skills, such as quantitative skills, conveying content, and building abilities to discuss complex ideas with evidence.

BIO's effectiveness in supporting areas of student learning.

Teachers rated each area of student learning on a 5-point scale, with 1 being Not at all Effective and 5 being Extremely Effective. No teacher rated any statement as a 1 out of 5.



What Students Liked about BIO Lessons

Students from Cohort 1 classes most often said they enjoyed Data Sketches. What students liked the most about all of the activities was making discoveries using data and being able to make creative design choices.

In terms of what components students liked, the VTS-style discussions were identified by a quarter of students interviewed. Fenceposts were rarely mentioned by students as the most enjoyable element. Data Sketches was the most recent activity students had done, however interviewers described all BIO activities to jog memories.

More importantly, what they liked wasn't just "doing art is fun." **Students commonly described that they enjoyed the process of discovering new ideas or connections through the process, as well as being in the role of *making* design choices for a quality visual product**.

Other students enjoyed collaborating. This included group work and hearing from other students that noticed different details or came to different conclusions than them. Some students simply enjoyed drawing, felt calm, or had other positive reactions.

How students described what they liked about the BIO lessons they experienced

Coded, open-ended responses to a student interview question asking what students liked about BIO lessons. The n reflects students that were asked and provided a response to each question and/or sub-question.

Which BIO Lesson(s) Students Recalled Liking	(n=67)
Data Sketches	60%
VTS-Style Discussions (Art or Data)	25%
Fenceposts	4%
Other (Unclear or non-BIO activity; not coded below)	15%

What Students Liked about the BIO Lessons	(n=60)
Making **observations**, **discovering** new **connections**, and **meaning-making**	40%
Making **design choices**, exercising **creativity**, creating a **quality** product	32%
Collaborating or **connecting** with others; hearing **others' viewpoints**	23%
Drawing or using their **artistic skill**	12%
Feeling **relaxed** or **calm**	8%
General **positive** (fun, cool, different)	7%

What Students Learned from BIO Activities

Almost 60% of interviewed students were able to recall something specific they had learned about data or maps in a BIO classroom. This included science information, thinking analytically, and awareness about data and visualization.

Over a quarter of students (28%) recalled learning new science information or a surprising fact through exploring geospatial data visualizations in the various BIO activities. 13% described how they learned that there isn't always a single right answer or otherwise described using flexible or analytical thinking about complex data or phenomena.

Other students commented on discovering how many different types of data were available, and how mapped data could be useful in a number of ways. Others learned about making purposeful design choices or how data representations are made. And finally, some reported learning the fundamental components of data visualizations, such as using keys and legends.

Around 40% of students weren't able to recall anything in particular that they had learned about maps or data that school year when directly asked.

What students recalled learning from BIO lessons.

Coded, open-ended responses to a student interview question asking what students recalled learning that semester about data and maps. The n reflects students that were asked the question and chose to respond.

(n=69)	What Students Learned from the BIO Lessons
59%	Students that **could recall something they learned** from BIO
28%	Discovered **new science information** through mapped images and/or data
13%	Comfort with **having no "right" answer,** thinking analytically about data
9%	Understood the **wide range of geospatial data** and how it **could be used**
9%	Insights about **design choices** and **representation of data**
9%	Improved understanding of **data representation elements** such as keys
41%	Students that **couldn't recall anything** they had learned

Students' Words: BIO Experiences and Learning

"One thing that I learned was one data set that [Teacher] assigned us was, I think it was the wildfires, the amount of wildfires around the world and increase. And so, then we connected that data set to maybe the amount of droughts that happened this year. If there was an increase or decrease in droughts, then maybe that affected the wildfires because if there was a lot of droughts, then maybe things like the air would get very dry and it would make a bit easier for wild forests to catch on fire and create more wildfires."

-Student Interview

"I liked the color matching part to where all the data was. ... Oh, I think it was about population, so I used a light pink for areas that were low in population and a dark purple for areas that were dense in population. I found that really interesting, because I liked seeing where a lot of people were, basically. ... It was kind of interesting to see that the bigger an area is doesn't necessarily mean that there's going to be more people there.

-Student Interview

"I think there was a lot of emphasis on being able to visualize this data in certain ways and communicate it to an audience. So, I think I practiced those skills this year a lot. I learned about the different methods for visualizing data... I learned, in respect to this sort of activity, I learned some about how some data was collected, specifically when it tied in with [topics from] the rest of the class."

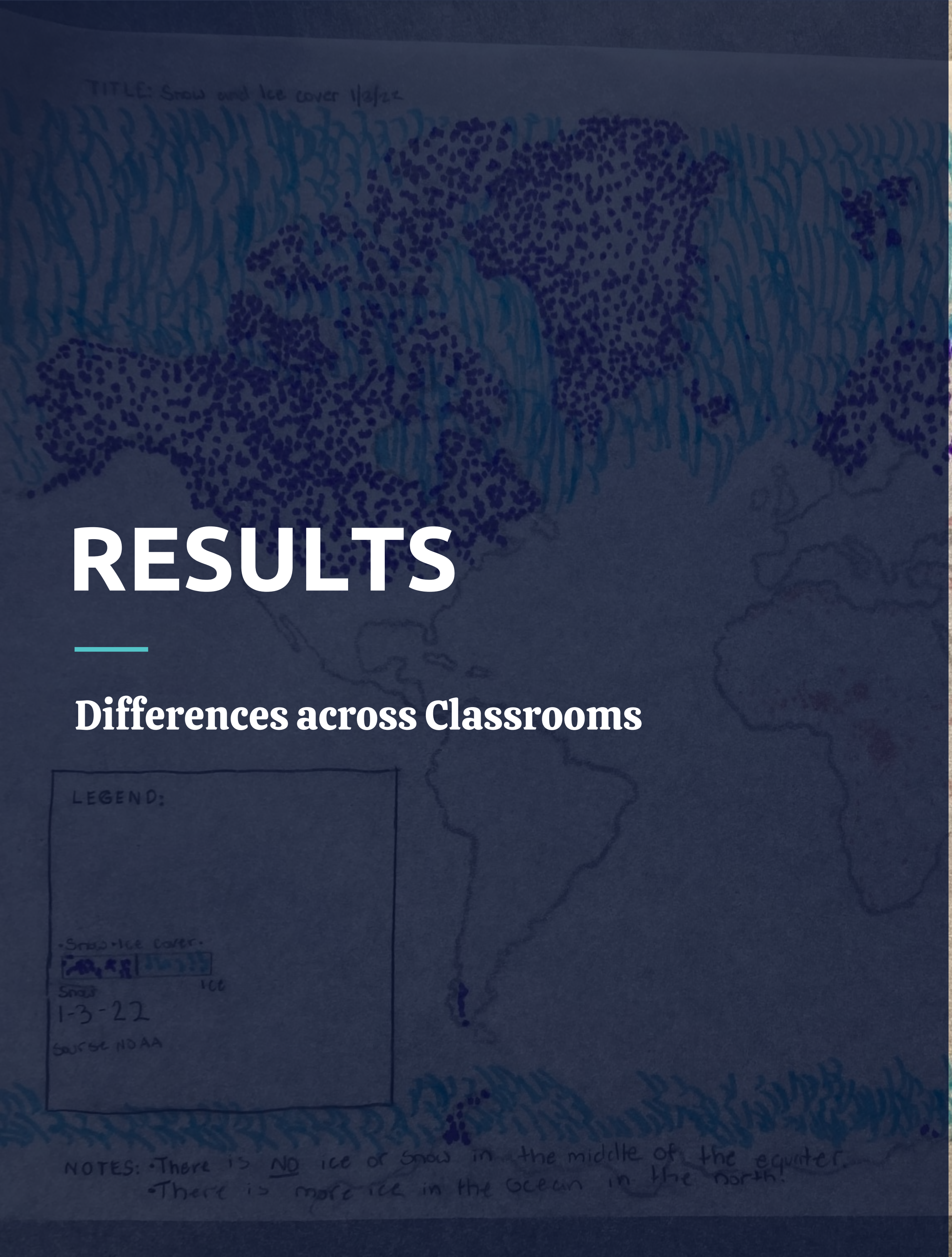
-Student Interview

"I really liked the teamwork aspect of it because all of your maps had to correlate, and especially with the keys, if everyone wrote their symbol at the very top of the key box, then all of them would overlap, so you just ... It really helps with communication because you have to know what everyone else is doing, so that way it all confines to a visible map. ... It makes it easier to understand when you have multiple opinions on it, and so when you're working together as a team, you can make sure that the symbols that you're using represent the math that you're describing, and so that way it's easier to understand..."

-Student Interview

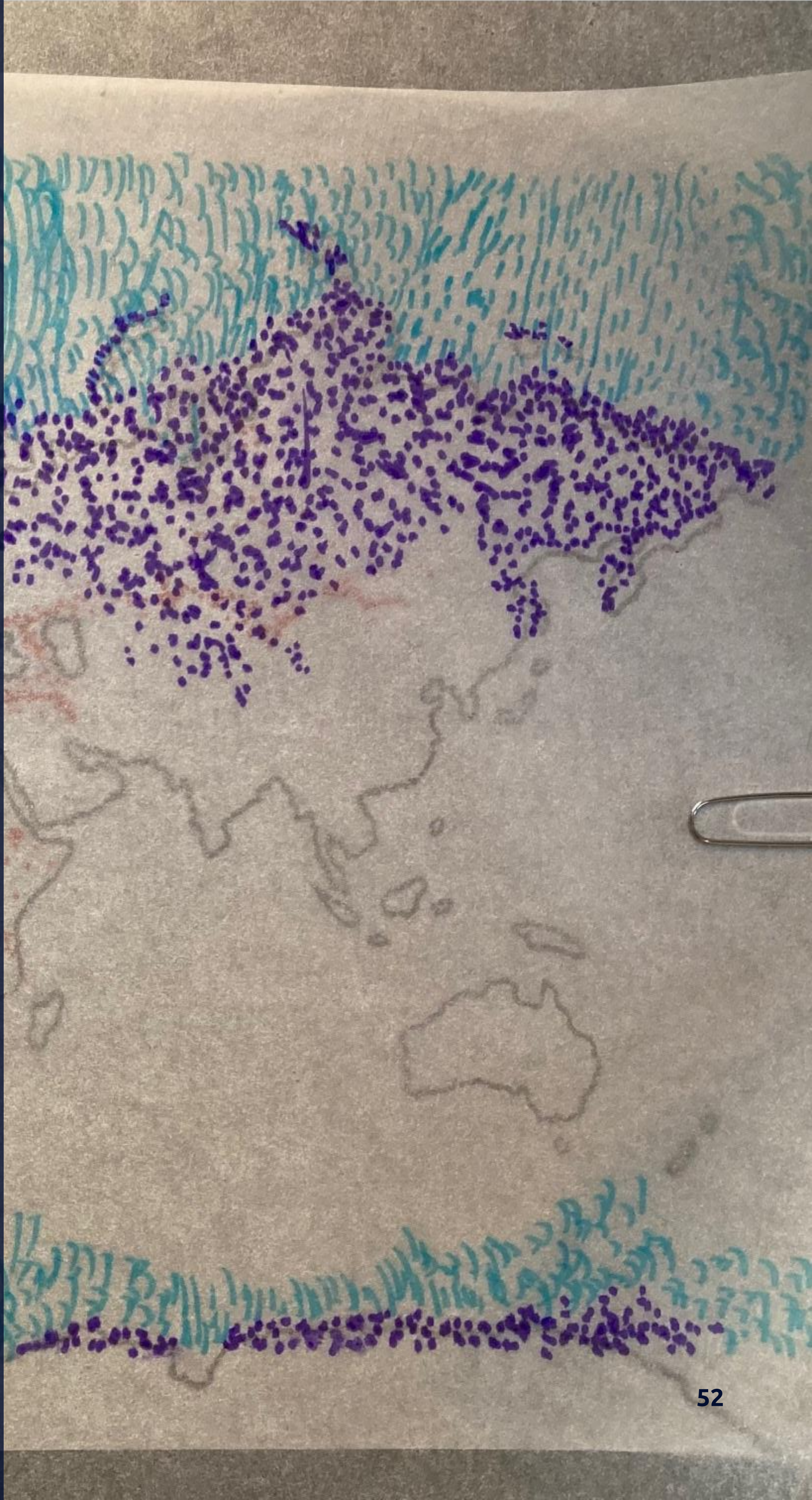
"[Teacher] said, "Oh, let's do a different way of teaching." [So, for] a few days, he shows us a diagram, and everyone gets up and they say like, "Oh, what do you think this is about? What do you think this is about?" And you have to use all the information around it and just figure it out, which I found that really fun and entertaining how we could just disprove each other's arguments because of more stuff in the picture. So, I found that fun."

-Student Interview



RESULTS

Differences across Classrooms

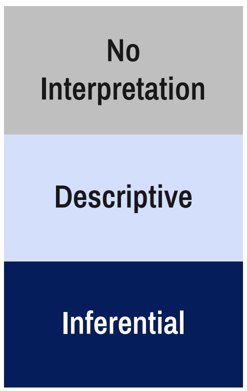


By Classroom: Change in Interpretation Types

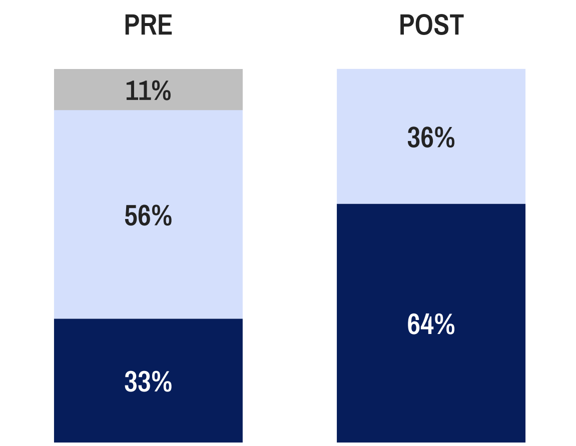
Frequency of descriptive and inferential interpretations before and after BIO instruction, by each of 5 participating classrooms.

The n for each class represents the three opportunities per student to make an interpretation during the data performance interview.

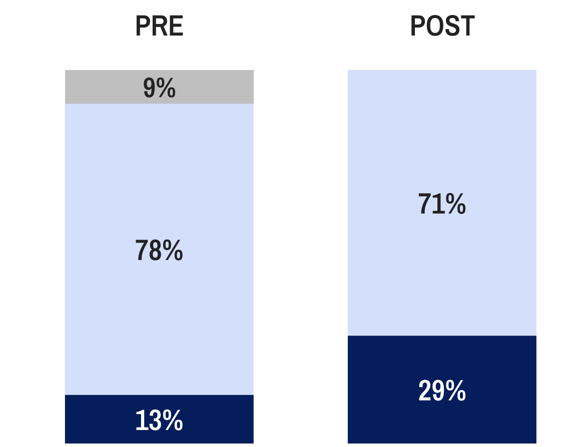
Each classroom increased the number of claims made (inferential interpretations) after BIO teaching. Classrooms 1, 4, and 5 saw the greatest growth.



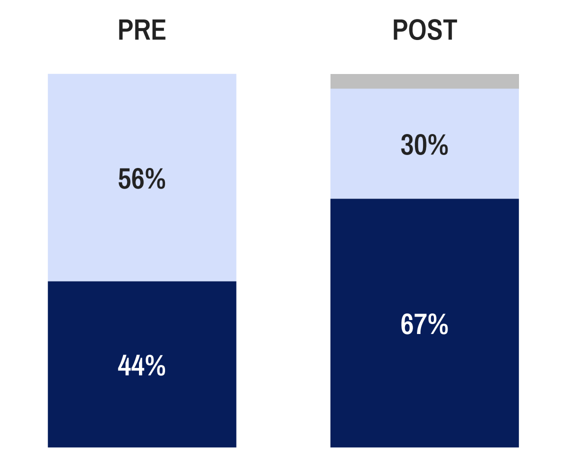
Classroom 1 (n=36)



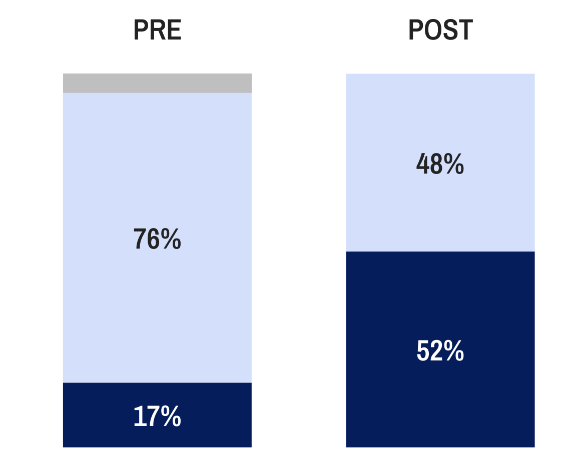
Classroom 2 (n=45)



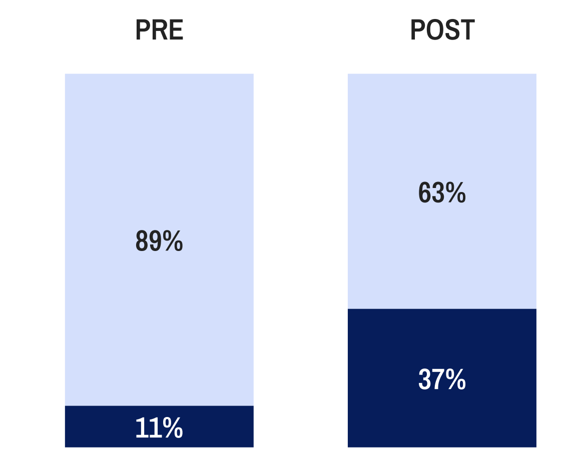
Classroom 3 (n=54)



Classroom 4 (n=42)



Classroom 5 (n=27)

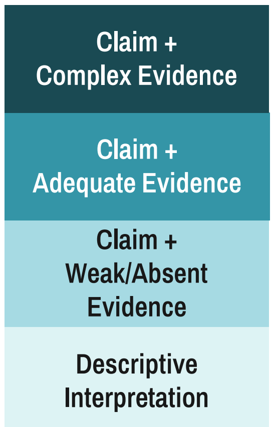


By Classroom: Ability to Make and Support Claims

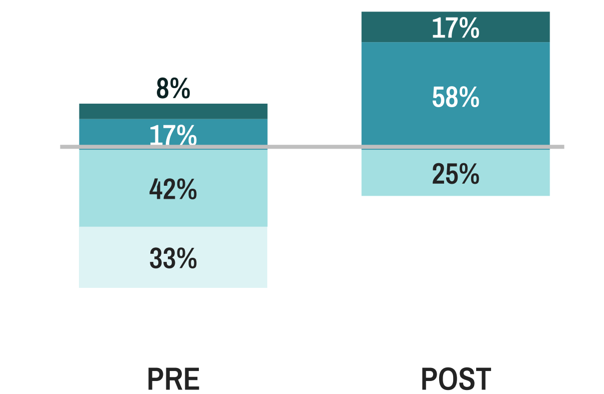
Changes in highest level achieved in making interpretations and supporting with evidence before and after BIO instruction.

Graphs below show the score for each students' most advanced interpretation/evidence segment during the data performance task, out of three tries. The percentages represent how many students fell into each level as their "personal best" during the task.

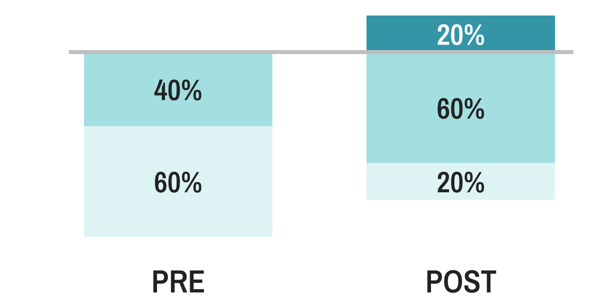
Each classroom saw an increase in inferential interpretations and adequate/complex evidence after experiencing BIO lessons.



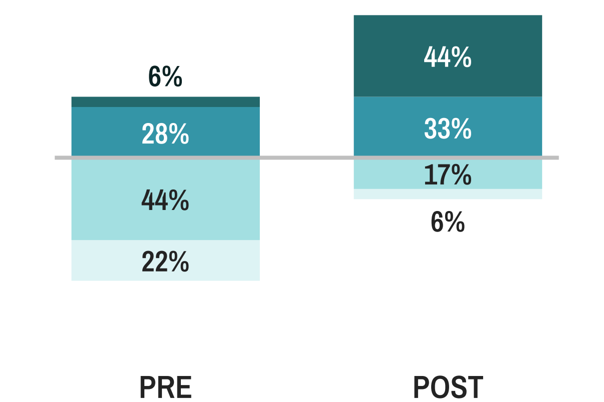
Classroom 1 (n=12)



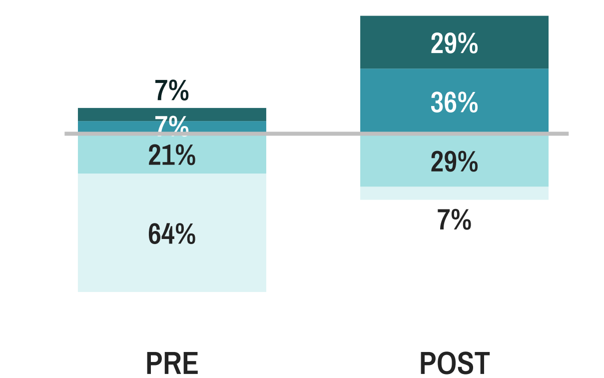
Classroom 2 (n=15)



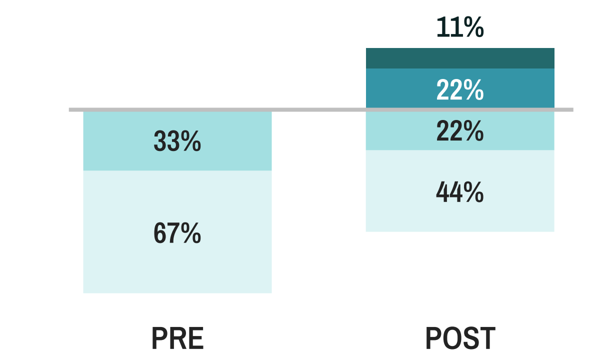
Classroom 3 (n=18)



Classroom 4 (n=14)



Classroom 5 (n=9)



Each Classroom's Frequent Observations Post-BIO

The three most common types of observations made by students after completing BIO, by classroom, as well as the observation type that showed the greatest increase after the BIO instruction.

Graphics below show the most common observational types within a classroom, as well as the observation category that saw the greatest increase from Pre to Post. All observational types and categories are described In greater detail on pages 26-34 of this report.

Each classroom differed in what visual elements students observed most. Classrooms 3 and 4 had high rates of noticing geography and data similarities; other classrooms more often noticed presence of data. In all classrooms, students grew the most in making observations about geography and data patterns.

Classroom 1 (n=12)

- Data Presence and Meaning
- Contrasts in Data Points
- Similarities in Data Points

Greatest Increase: Geography (+41%)

Classroom 2 (n=15)

- Presence of Data
- Identify Political Location
- Contrasts in Data Points

Greatest Increase: Data Patterns (+47%)

Classroom 3 (n=19)

- Identify Political Location
- Similarities in Data Points
- Data Presence and Meaning

Greatest Increase: Data Patterns (+16%)

Classroom 4 (n=14)

- Similarities in Data Points
- Identify Political Location
- Data Presence and Meaning

Greatest Increase: Geography (+14%)

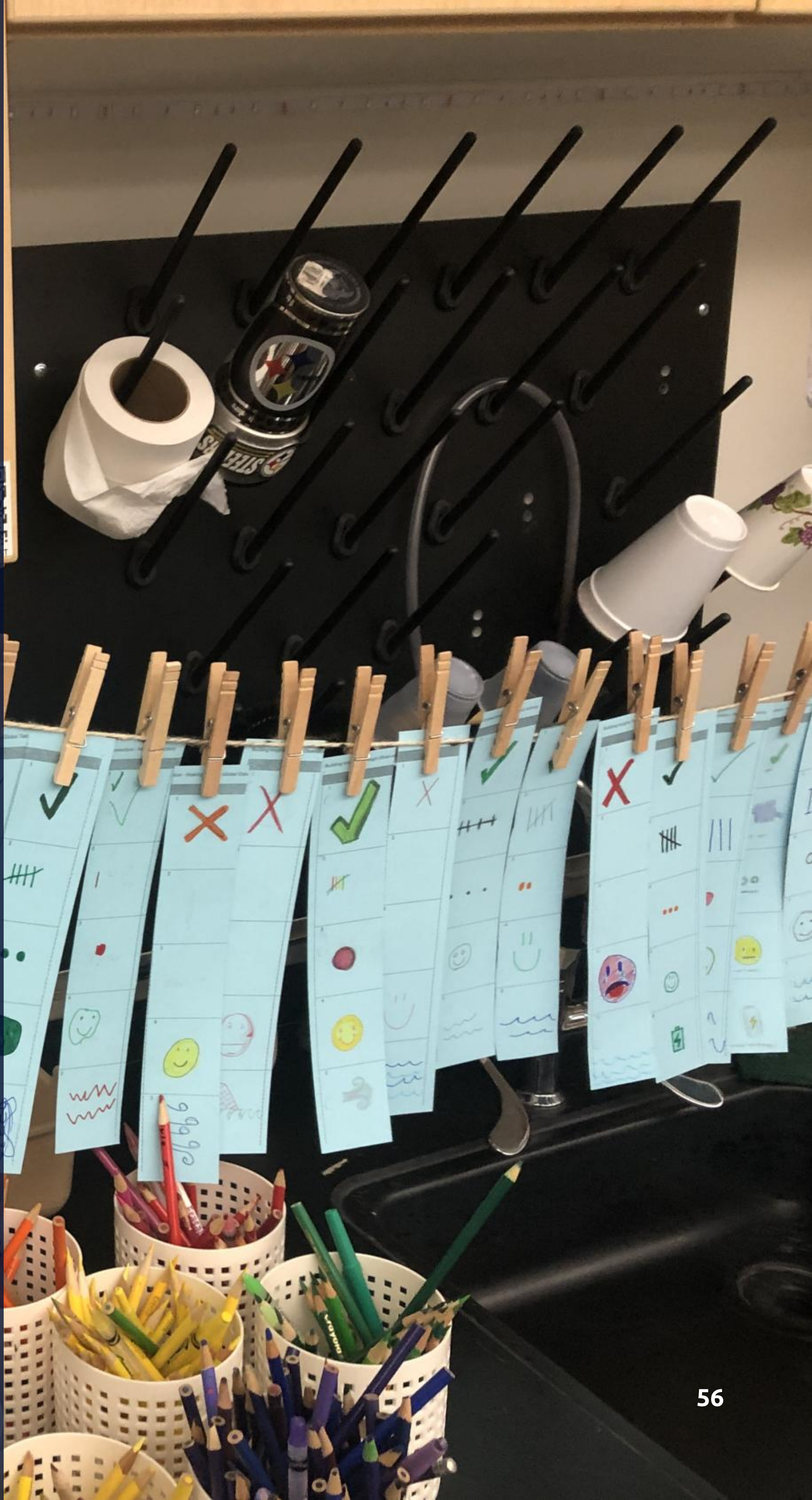
Classroom 5 (n=9)

- Similarities in Data Points
- Presence of Data
- Single Case Value

Greatest Increase: Data Patterns (+22%)

CONCLUSIONS

Development of Student Skills



Conclusions

>> Improved Data Literacy

Students who experienced the arts-based data literacy instruction, using the Building Insights through Observation approach, showed significant gains in core data literacy skills. They got better at making claims with data, supporting those claims with evidence, making more and different observations, seeing patterns in data, and broadening their geographical knowledge. Moreover, the gains achieved by BIO students were greater than their peers in control classrooms. Although this study was preliminary and exploratory, the evidence suggests that there is great promise in the approaches of teaching visual thinking and design thinking skills in the science classroom for improving data literacy skills and outcomes.

1 Better Data Interpretation + Evidence

In BIO classrooms, students were far better at making claims about data – making interpretations that went beyond what the data show. In addition, they got better at supporting those claims with robust evidence, reasoning about how data patterns relate to natural phenomenon, human interactions, and use of data. Further, exploratory analysis indicated that BIO instruction led to greater and more consistent improvements over the year than the control group.

2 More Observations from the Data

Students in BIO classrooms grew substantially in the number of unique observations they generated from geospatial data after their arts-based instruction. BIO students saw more unique features or patterns in the data – and wider variety of types of features and patterns. As with interpretation, experiencing the BIO arts-based instruction significantly impacted the degree of improvement in the number of observations made, with BIO students showing far greater growth, compared to their control group peers.

3 Seeing Patterns & Knowing Geography

In the types of observations students made during pre- and post-interviews, the data showed that BIO students especially grew in their ability to identify data patterns (similarities, contrasts, aggregate) and to apply geographical knowledge to their observations. Importantly, there was a significant relationship between having experienced arts-based BIO instruction and the increases in use of these skills during the performance task, compared to the control group.

Considerations for Future Research

Focus on Flexibility

BIO students showed an increase in their ability to think flexibly about data, but the differences between the study and control groups was smaller than other outcome measures. For a future study, this skill would benefit from a more focused measure to specifically examine level of flexibility at generating data interpretations, to determine the degree of impact of the method on this skill.

Narrowing Measurement of Questioning Skills

Middle school students were very capable of asking questions of data, with the types of questions asked barely affected by time or treatment. In the future, this could benefit from a more focused study of specific types of questions that we anticipate are most prompted through the BIO activities – adding new data or considering alternative representations – or if students continue to be able to generate these questions when prompted, regardless of the form of instruction received.

Consider Teacher-Observed Outcomes

Our study design required the study instrument and protocols to be developed prior to the BIO curriculum. As a result, the measures were very broad and not yet informed by the experience of teachers observing students in classrooms. Future study of this model will benefit from tighter protocols and measurements to reveal the skills observed by teachers as most targeted, including things like design vocabulary (which was not explicitly targeted, but showed hints of promise in the final data).

This material is based upon work supported by the National Science Foundation under Grant #2101310. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of NSF.

J. Sickler CONSULTING

For more information about this report, contact:

Jessica Sickler

jessica@jsickler.net

www.jsickler.net