Special Series: Scientific Literacy
JMBE Journal of Microbiology & Biology Education

Using an Observation Protocol To Evaluate Student Argumentation Skills in Introductory Biology Laboratories

Lindsey Clevenger,[a] Jennifer Teshera-Levye,[a] Joi P. Walker,[b] and Heather D. Vance-Chalcraft[a]
[a]Department of Biology, East Carolina University, Greenville, North Carolina, USA
[b]Department of Chemistry, East Carolina University, Greenville, North Carolina, USA

Argumentation is vital in the development of scientific knowledge, and students who can argue from evidence and support their claims develop a deeper understanding of science. In this study, the Argument-Driven Inquiry instruction model was implemented in a two-semester sequence of introductory biology laboratories. Student's scientific argumentation sessions were video recorded and analyzed using the Assessment of Scientific Argumentation in the Classroom observation protocol. This protocol separates argumentation into three subcategories: cognitive (how the group develops understanding), epistemic (how consistent the group's process is with the culture of science), and social (how the group members interact with each other). We asked whether students are equally skilled in all subcategories of argumentation and how students' argumentation skills differ based on lab exercise and course. Students scored significantly higher on the social than the cognitive and epistemic subcategories of argumentation. Total argumentation scores were significantly different between the two focal investigations in Biology Laboratory I but not between the two focal investigations in Biology Laboratory II. Therefore, student argumentation skills were not consistent across content; the design of the lab exercises and their implementation impacted the level of argumentation that occurred. These results will ultimately aid in the development and expansion of Argument-Driven Inquiry instructional models, with the goal of further enhancing students' scientific argumentation skills and understanding of science.

KEYWORDS argumentation, inquiry, laboratory, introductory biology, social, cognitive, epistemic, argument driven inquiry, assessment of scientific argumentation in the classroom, observation protocol

INTRODUCTION

Argumentation plays a central role in the development, evaluation, and validation of scientific knowledge and is an important practice that makes science different from other ways of knowing (1–3). Learning how to argue scientifically helps students construct knowledge by taking their personal sense-making and transferring it to their peers beyond a "correct answer" (2). When scientific arguments are developed in groups, the scientific understanding of individuals, as well as group interactions, influence the process and outcome of argumentation.

The theoretical framework behind argumentation in groups includes cognitive, epistemic, and social subcategories (Table 1) (3). The cognitive dimension of argumentation can be described as how the group interacts with ideas, specifically how they attempt to develop meaning or understanding. For example, this includes discussing multiple claims and lines of evidence, challenging ideas based on evidence, and modifying claims or explanations when faced with conflicting evidence. The epistemic dimension of argumentation includes how the group determines what counts as valid and how consistent the group's process is with the culture of science. Groups demonstrating high levels of epistemic argumentation would consider the value of a piece of evidence, examine relationships between pieces of evidence, evaluate how data were interpreted, base their arguments and claims on scientific theory, and use scientific terminology correctly. Finally, the social aspects include how the group members interact with each other. Social argumentation is shown by groups who reflect on and monitor their progress, allow everyone to express ideas, ask group members to clarify or elaborate on their ideas, discuss ideas that are proposed instead of dismissing them, and show high levels of engagement from all group members. These three subcategories of argumentation describe a collaborative process of proposing, supporting, evaluating, and refining ideas to make sense of a complex or ill-defined problem or to advance knowledge in a manner that is consistent with conceptual structures, cognitive processes, epistemological commitments, and the social norms of science (2–4).

Editor Justin Shaffer, Colorado School of Mines
Address correspondence to Department of Biology, East Carolina University, Greenville, North Carolina, USA. E-mail: vancechalcrafth@ecu.edu.
The authors declare no conflict of interest.
Received: 2 November 2022, Accepted: 8 May 2023,
Published: 22 May 2023

TABLE 1
Subcategories (cognitive, epistemic, and social) and their 6 items that were rated in the ASAC model

Cognitive	Epistemic	Social
The talk of the group was focused on solving a problem or advancing understanding.	The participants used evidence to support and challenge ideas or to make sense of the phenomenon under investigation.	The participants were reflective about what they know and how they know.
The participants sought out and discussed alternative claims or explanations.	The participants examined the relevance, coherence, and sufficiency of the evidence.	The participants respected what each other had to say.
The participants modified their explanation or claim when they noticed an inconsistency or discovered anomalous information.	The participants evaluated how the data were gathered, analyzed, or interpreted.	The participants discussed an idea when it was introduced into the conversation.
The participants were skeptical of ideas and information	The participants used scientific theories, laws, or models to support and challenge ideas or to help make sense of the phenomenon under investigation.	The participants encouraged or invited others to share or critique ideas.
The participants provided reasons when supporting or challenging an idea.	The participants made distinctions and connections between inferences and observations explicit to others.	The participants restated or summarized comments and asked each other to clarify or elaborate on their comments.
The participants attempted to evaluate the merits of each alternative claim or explanation in a systematic manner.	The participants used the language of science to communicate ideas.	There was equal participation from all members of the group.

Argument-Driven Inquiry (ADI) is an instructional model designed to elicit scientific practices, such as argumentation (1). Inquiry-type exercises have the potential to serve as an effective platform for formulating arguments, owing to the special features of this learning environment (5). ADI instructional models for lab courses at the undergraduate level have been described in the literature previously with slight variations (1–4, 6) but often consist of prelaboratory activities, proposals, inquiry labs, argumentation sessions using whiteboards, lab reports, and peer review (Fig. 1). Prelaboratory activities (completed in the lab) are where the students are introduced to the techniques and content that they need for the inquiry lab, followed by their preparation of a group proposal which they use as their guide for the inquiry lab. Once they complete the inquiry investigation using their proposal, the students analyze their data and prepare a tentative argument consisting of a claim, evidence, and justification. The claim answers the guiding question, the evidence is the data that helps support their claim, and the justification is where the students make sense of phenomena using their data and scientific knowledge. The claim, evidence, and justification are displayed on a whiteboard for other groups to see during the argumentation session (Fig. 2). During this session, one member of the group stays behind during the whiteboard rotation and presents their whiteboard findings to visiting groups. The other group members rotate to visit other groups' whiteboards and ask questions regarding their techniques and data collection. This dialogue at the whiteboards is called an argumentation session. After the argumentation session, each group has a chance to change their claim and argument if they desire, based on information gained through the session. Then, each student writes an individual lab report that ultimately elaborates on the claim, evidence, and justification from their whiteboard. Once the lab report is peer-reviewed, it is individually submitted for grading.

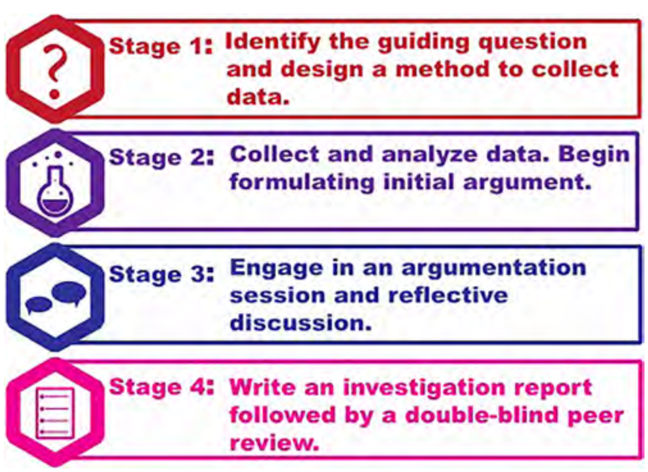

FIG 1. The four stages of Argument-Driven Inquiry instructional models.

The ADI lab format has been used in a variety of disciplines at multiple institutions. For example, laboratory courses in general chemistry (1, 6), human anatomy (7), and introductory physics (8, 9) have been converted to the ADI instructional model. Studies have reported greater student engagement and improvements in reasoning and argumentation on a scientific level after implementation of ADI labs (6, 7).

Using this framework of argumentation in ADI instructional models, we evaluated three research questions.

1. Are students equally skilled at all subcategories (cognitive, epistemic, and social) of argumentation?

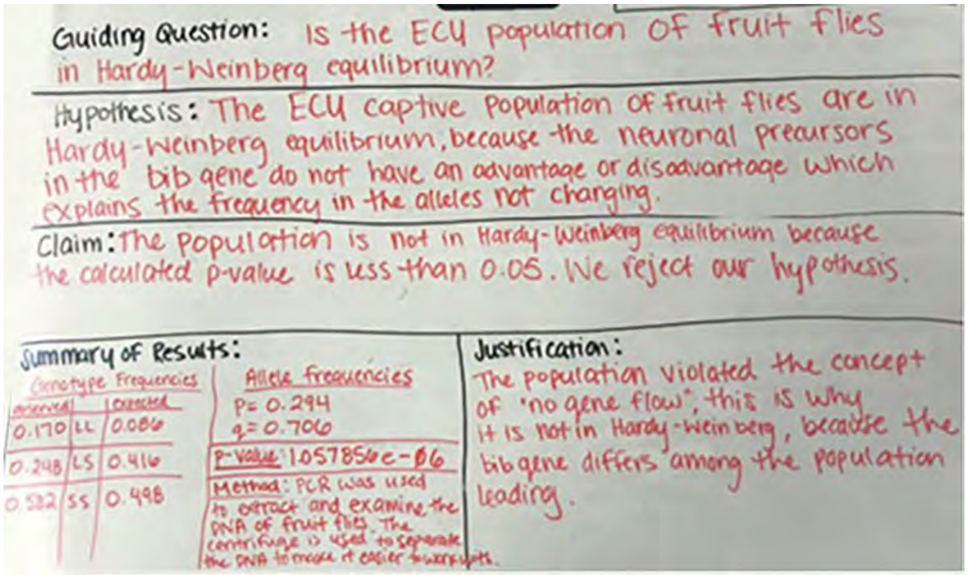

FIG 2. An example of a whiteboard used in an argumentation session within an ADI lab. Most whiteboards had the guiding question, a hypothesis, a claim, results, and a justification.

2. How do students' argumentation skills differ based on lab exercise (within a course)?
3. How do students' argumentation skills differ based on lab course (between two different lab courses)?

These results can identify areas in which students need additional support to increase mastery of argumentation and provide valuable information to curriculum developers interested in creating ADI-style lab exercises.

METHODS

Study population

This study took place at East Carolina University (ECU), an R2 institution in the southeast United States. At ECU, each introductory biology lecture course is accompanied with a 1-credit laboratory that follows the ADI instructional lab model. Introductory Biology Laboratory I, Biol 1101, covers topics related to cellular and molecular biology, and Introductory Biology Laboratory II, Biol 1201, covers topics related to ecology and evolution. Each lab course is split into sections, and all sections of both introductory biology laboratory courses follow the same ADI format so students in Introductory Biology Laboratory II have generally already participated in ADI labs in the first laboratory course. Students in the lab course are mostly biology majors, but they may be majoring in another science, engineering, or health-related field. The students work in groups of four in each course, with groups generally staying the same within a semester.

Each lab section is taught by a graduate student teaching assistant (TA), with different TAs teaching different sections. All new TAs attend a presemester workshop that introduces evidence-based teaching methods and the ADI instructional approach. This interactive workshop models student-centered approaches while the teaching assistants act as students and complete a short experiment, write their claim, evidence, and justification on a whiteboard for other groups to see during the argumentation session (i.e., "develop a whiteboard"), and have an argumentation session. Throughout the process, the facilitator points out key aspects of the instructional approach. For example, the workshop emphasizes the need for instructors to allow students to take the lead in the argumentation, as instructor participation in argumentation has been shown to negatively impact the argumentation process. At the conclusion of the argumentation session, the facilitator concludes by asking workshop participants to reflect on what they learned from the session and any changes they need to make to their whiteboards or for their arguments that would be featured in their lab reports. Beyond the presemester workshop, TAs attended weekly planning meetings with course personnel to reflect on instructional successes and challenges, as well as to discuss evidence-based practices for the upcoming week.

This study includes data from student participants in two sections from the 17 sections of Principles of Biology Laboratory I in Fall 2019 and seven sections from the 16 sections in Principles of Biology Laboratory II in Spring 2020. Biology Laboratory I had a maximum of 48 students per section, while Biology Laboratory II had a maximum of 24 students per section. Data on a total of 10 groups of students (38 students total) from Biology Laboratory I and 21 groups of students (84 students total) from Biology Laboratory II are included in this research study. Focal sections were determined by logistical considerations and the number of students consenting to participate in the research study.

Ethics statement

ECU's Institutional Review Board approved the human subject research (UMCIRB 17-001117).

Lab exercises

We focused on two of the laboratory investigations (each taking place over 3 weeks) from each course. Each course had four ADI investigations in a given semester that comprised the entire content of the lab, but we focused on only two from each course, based on logistical constraints, as the Spring 2020 data for Biology Laboratory II was collected in person but the rest of the semester was moved online due to the 2019 coronavirus disease (COVID-19) pandemic. For Biology Laboratory I, the investigations that were studied were an enzyme lab ("What factors decelerate the apple browning process?"; weeks 4 to 6 of the semester), and a photosynthesis lab ("What factors led to the greatest increase in the rate of photosynthesis?"; weeks 7 to 9 of the semester). These investigations were chosen because they had the most video data collected.

During the enzyme lab, students were assigned either alcohol, salt, or pH and had to determine whether these factors accelerated or decelerated the apple browning process (10). Browning occurred when catechol in the apple reacted with oxygen in the presence of catecholase (an enzyme found in apples). This reaction resulted in a brown product called benzoquinone (11). For this investigation, the groups added water, then catechol, and then the extract (variable of pH, salt, or alcohol) to the apple solution containing catecholase in a tube and incubated it for 3 min before being read on a spectrophotometer to determine if, and how much, benzoquinone was produced in the apple solution.

During the photosynthesis lab, students submerged leaf discs into a water solution consisting of baking soda and soap. The students then chose an independent variable and determined its effect on photosynthesis. The options for the independent variable were distance of light from the leaf discs, different colored light on the leaf discs, different wattages of light on the leaf discs, or other ideas the students had using the available supplies. The rate of photosynthesis was measured by the number of leaf discs that became suspended in the solution when introduced to the independent variable (10).

For Biology Laboratory II, the investigations that were included in this study were a population genetics lab ("Is the ECU captive population of fruit flies in Hardy-Weinberg equilibrium?"; weeks 1 to 3 of the semester) and a phylogenetics lab ("Do the phylogenetic trees for morphological data and genetic data agree when classifying spider genitalia?"; weeks 4 to 6 of the semester). For the population genetics lab, each group extracted DNA from ECU captive fruit flies and used PCR to replicate the DNA (12). The students then used gel electrophoresis to separate the DNA according to molecular size, enabling the students to compare different DNA samples. These results were used to determine whether the fruit fly population was in Hardy-Weinberg equilibrium or not and then explain why (13) (Fig. 2).

For the phylogenetics lab, the groups explored the evolutionary relationships of spider species by analyzing morphological and molecular characteristics (12). The students compared the morphological traits of trapdoor spiders (genus *Myrmekiaphila*) using pictures of genitalia morphology, as those traits are critical to identifying trapdoor spiders to the species level (14). Molecular Evolutionary Genetics Analysis (MEGA) software (15) was then used to mine information on the molecular traits of these spider species. Students used this morphological and molecular information to infer evolutionary relationships between the species and make a phylogenetic tree that illustrated the evolutionary relationships between several species of trapdoor spiders (12).

Data collection and observation scoring

Students who consented to participate in this research study were video recorded during the whiteboard development and argumentation sessions for each specified lab section and investigation. These videos included at least three rounds of visitors from other groups coming to the focal whiteboard. We watched each video and coded each using the Assessment of Scientific Argumentation in the Classroom (ASAC) observational protocol (16).

The ASAC observation protocol (16) was designed to measure the cognitive, epistemic, and social subcategories of scientific argumentation. Each of these three subcategories are evaluated based on 6 items (Table 1) that are rated based on number of times an item occurs, regardless of the length of time it occurs. For example, if a group changed their whiteboard based on information gained during the whiteboard argumentation session, this would be one instance of "The participants modified their explanation or claim when they noticed an inconsistency or discovered anomalous information," in the cognitive subcategory. With the cognitive subcategory, a group explaining that they had based their claim on their pH results would be one instance of "The participants used evidence to support and challenge ideas or to make sense of the phenomenon under investigation." One instance from the social subcategory ("The participants respected what each other had to say") is when a student mentioned that they had learned a lot from a different student group during the whiteboard argumentation session.

Each of the six items for each subcategory were given a score of 0, 1, or 2 for each group. A score of 0 meant no instances occurred, a score of 1 meant that a single instance occurred, and a score of 2 meant that two or more instances occurred. The maximum score in each subcategory was 12, for a total maximum score of 36 for each group. Each group was evaluated for the time in which they were analyzing their data and developing their argument to display on their whiteboard, explaining their arguments to their peers during a whiteboard argumentation session, and discussing their thoughts as a group after the argumentation session (with the possibility of revising their whiteboard and argument). High scores imply that the group demonstrated more aspects of argumentation (when developing, discussing, and revising their arguments), while lower scores reflect a less sophisticated degree of argumentation.

The ASAC was originally validated using a 7-step process for use in students at a similar academic stage to ours (16). To validate the ASAC scoring between individuals in our study, two individuals (L.C. and H.D.V.-C.) scored three videos from ADI lab sections that were not included in this study. We chose three because we were limited in the number of videos we had that were not included in this study. We then compared our scores using a measure of interrater reliability (IRR), the percent agreement. IRR is often necessary for research designs where data are collected through ratings provided by trained or untrained coders (17). After one round of scoring, the IRR did not meet an acceptable threshold (<0.75), so the researchers met to discuss each scoring criterion and come to consensus and then each scored an additional three videos. After the second round of scoring, a high percent agreement on total scores per video (0.98) was established, and the remaining videos were coded by L.C.

Statistical analysis

The total argumentation scores and argumentation subcategory scores (cognitive, epistemic, and social) for each lab investigation and course were calculated. To determine if students were equally skilled at all subcategories (cognitive, epistemic, and social) of argumentation, we used data from both introductory lab courses in a Kruskal-Wallis rank sum test to compare the cognitive, epistemic, and social subcategory scores, followed by pairwise Wilcoxon rank sum tests with a Bonferroni correction for multiple comparisons. To address whether students' argumentation skills differed based on lab exercise in the same course, we used separate Wilcoxon rank sum tests for the total score and each subcategory score. Finally, Wilcoxon rank sum tests on total and subcategory scores for argumentation were used to compare Biology Laboratory I and II. All statistical analyses were conducted in R (R Core Team).

RESULTS

Within the cognitive subcategory, the ASAC items with the highest number of instances were "The participants provided reasons when supporting or challenging an idea," "The talk of the group was focused on solving a problem or advancing understanding," and "The participants sought out and discussed alternative claims or explanations." The items with the highest number of instances within the epistemic subcategory were "The participants used evidence to support and challenge ideas or to make sense of the phenomenon under investigation," "The participants used the language of science to communicate ideas," and "The participants evaluated how the data were gathered, analyzed, or interpreted." Finally, within the social subcategory, the items with the highest number of instances were "The participants respected what each other had to say," "The participants encouraged or invited others to share or critique ideas," and "There was equal participation from all members of the group."

The results from the Kruskal-Wallis rank sum test and the pairwise Wilcoxon rank sum tests with the Bonferroni correction showed a significant difference between subscores (chi-squared = 54.548, $P < 0.0001$) (Fig. 3). The social subcategory

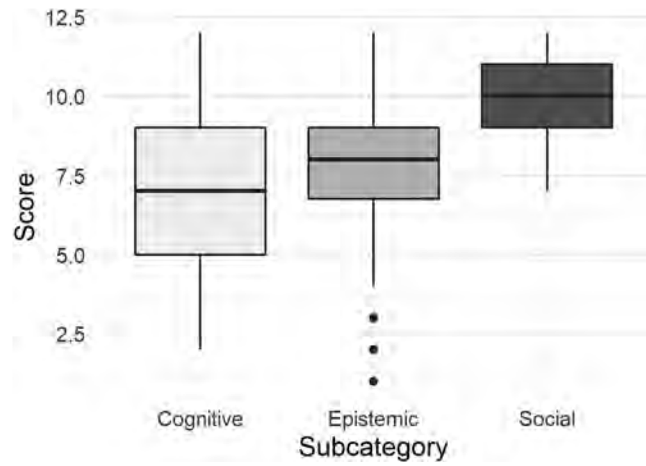

FIG 3. On average, social subcategory scores were significantly higher than the cognitive and epistemic argumentation scores, while the cognitive and epistemic scores were not significantly different from each other.

scores were significantly higher than those for the epistemic subcategory of argumentation ($W = 526.5$, $P < 0.0001$) (Fig. 3) and significantly higher than those for the cognitive subcategory of argumentation ($W = 461.5$, $P < 0.0001$) (Fig. 3). There was not a significant difference between the cognitive and epistemic subcategories of argumentation ($W = 1851.5$, $P = 0.10$).

Separate Wilcoxon rank sum tests were used for the total scores and each subscore from the two investigations (lab exercises) from both Biology Laboratory I and Biology Laboratory II. There was a significant difference in total score ($W = 46.5$, $P = 0.04$) (Fig. 4) and the cognitive subcategory ($W = 47$, $P = 0.03$) (Fig. 4) between the two investigations in Biology Laboratory I. There was no significant difference in the epistemic subcategory ($W = 32$, $P = 0.68$) (Fig. 4), but there was a significant difference in the social subcategory between the two investigations in Biology Laboratory I ($W = 47$, $P = 0.03$) (Fig. 4). There were no significant differences in total scores or any subscores between the two lab exercises in Biology Laboratory II (all $P > 0.05$).

The Wilcoxon rank sum tests were not significantly different in total score or any subscore (cognitive, epistemic, or social) between Biology Laboratory I (1101) and Biology Laboratory II (1201) (all $P > 0.05$) (Fig. 5).

DISCUSSION

Students in this study were not equally skilled at all subcategories of argumentation; they were significantly better at the social subcategory than the cognitive and epistemic subcategories. In addition, there were differences in argumentation scores between the two lab exercises in Biology Laboratory I. Finally, there was no significant difference in argumentation levels between two sequential biology laboratory courses (Biology Laboratory I and II).

Students were more proficient engaging in social aspects of argumentation rather than cognitive and epistemic aspects.

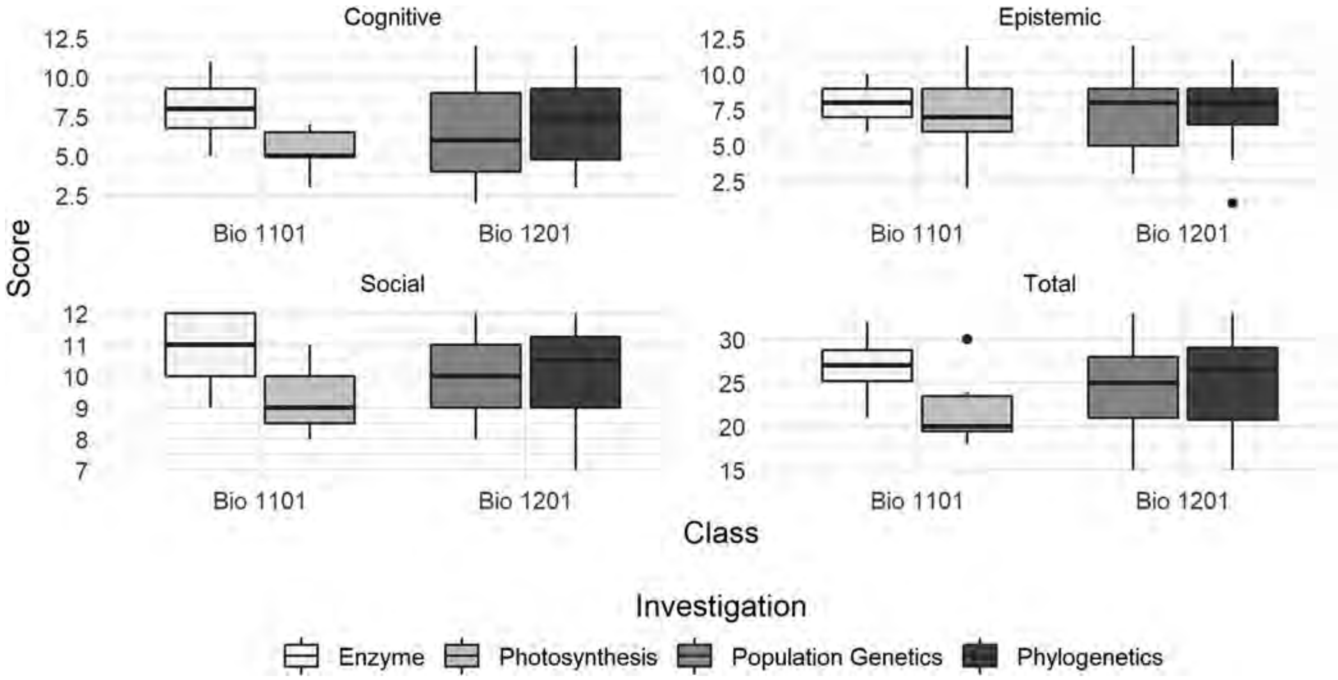

FIG 4. The total, cognitive, and social subcategory argumentation scores differed between the two investigations for Biology Laboratory I (1101). There were no differences between the two investigations in Biology II (1201) in any argumentation score.

Cognitive processes are usually difficult for students to grasp, because cognitive learning requires students to "think about their thinking" in a way for them to open a deeper understanding of a concept. Epistemic practices are also challenging, because they require students to evaluate their knowledge, coordinate theory and evidence, and make sense of patterns in data (18, 19). The cognitive and epistemic scores reported here, however, were similar to those found in prior studies, while our social scores were higher (1–3). While there is no known "acceptable" level of argumentation, these scores provided insight on what parts of argumentation had the most room for improvement. Additional emphasis on determining the sufficiency of evidence and weighing alternative explanations may be beneficial to students.

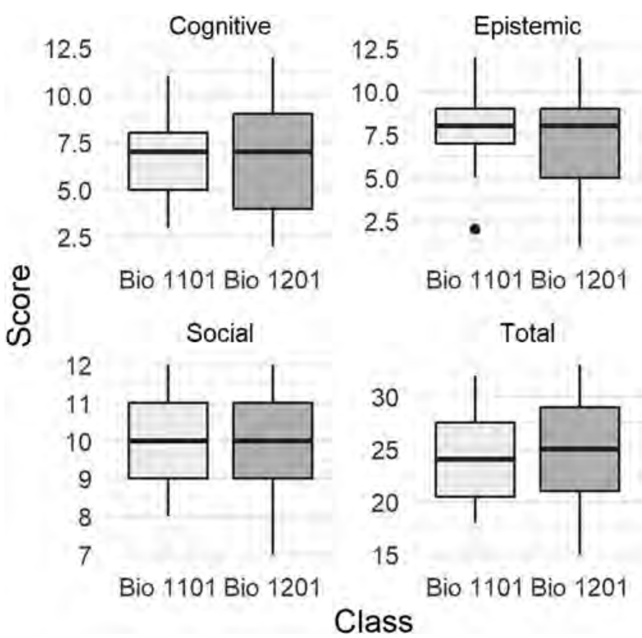

FIG 5. The argumentation scores (total and each subcategory score) of the two lab exercises together from Biology Laboratory I (1101) and Biology II (1201) were not significantly different.

Grooms et al. (20) found that ASAC scores (especially the epistemic and social subcategory scores) on familiar content were higher than those for unfamiliar content. They concluded that how a group of students interacts with each other during an episode of scientific argumentation is predominately influenced by how familiar the group members are with the fundamental concepts related to a given topic. This first laboratory in the Principles of Biology Laboratory I was on a food-related phenomenon, for which students would have been familiar. This could explain the higher-than-expected social score for this investigation.

There was a significant decrease in the cognitive and social subcategory from the first to the second investigation in Biology Laboratory I. Hosbein et al. (1) found an increase in subcategories among two sequential lab investigations as students learned how to scientifically argue with practice and time. Some intervening factors appeared to decrease argumentation. Investigation 2 (the photosynthesis lab) lacked robust discussion in one of the two focal sections, possibly due to the influence of the graduate TA's instruction. In the photosynthesis lab, there were many factors that the students could have chosen as the independent variable, but the TA of this section was heard on the videos telling the students which factor would give the best or easiest result. As a result, almost all groups chose the same factor to test and came up with the same results, which led to decreased discussion.

There was also no significant difference between the quality of argumentation in the two investigations from Biology Laboratory II. One possible factor that may have prevented an overall increase in subcategory scores from the first to the second investigation in this lab course was the content of the lab exercise on phylogenetic trees. For example, on one video the presenter stated that her group was very confused throughout the whole process of making their phylogenetic trees and the representatives from the other groups agreed that they were confused too. Moreover, not all students accessed all available resources. There was a reference map that was posted for students to use when preparing their tree, and only some groups from all sections used it. During the argumentation sessions, a member from another group visiting a whiteboard said their group got the same result, but their justification was different because they didn't use the reference map that the presenting group did. Hearing the student comments on these videos prompted us to change this lab exercise in future semesters to reduce confusion and, hopefully, allow increased argumentation.

Despite the fact that students in Biology Laboratory II had previously participated in ADI labs during Biology Laboratory I, there was no significant difference in argumentation scores between these two courses. Although we cannot say specifically why there was no improvement, we hypothesize that the content and instruction of the lab exercises played a role. Lab exercises need to have sufficient variability to give students the opportunity to practice argumentation (20). This variability can be created by allowing students to choose different variables for the same exercise (like the photosynthesis lab) or use different methods or resources for the same exercise. The lab exercises we were able to record in Biology Laboratory II before the COVID-19 pandemic moved instruction online had less variability for the students and focused on concepts that were known to be very unfamiliar for introductory biology students (i.e., Hardy-Weinberg equilibrium and phylogenetics). These lab exercises were more abstract than the exercises in Biology Laboratory I and were difficult for the students to understand deeply. Based on these argumentation scores and student feedback, we have now changed these lab exercises to be more explicit and hands-on with organisms that are more familiar to students while still covering the same difficult content in an ADI format. Future data collection will allow us to determine if making the labs less abstract and more familiar to students can improve the quality of argumentation that results.

The lack of a difference in argumentation scores between Biology Laboratory I and II may also have been influenced by instructional practices. The graduate TAs in Biology Laboratory II generally seemed more skeptical of their students' ability to fully understand these difficult and abstract lab exercises and tended to provide more guidance than is ideal. Laboratory instructors need to fully understand and "buy in" to the lab format so that they allow students to get sufficiently variable results and then fully benefit from the argumentation sessions, rather than encouraging students to get the "best" results. Professional development opportunities and planning meetings with graduate teaching assistants may need to explicitly discuss the benefits of having students engage in argumentation and the problems that can arise from too much intervention in the process. Fostering graduate teaching assistant trust in their students' abilities may also be helpful. Finally, having students examine primary literature during the prelaboratory exercises may help students be more confident when designing their experiments and generating their claims, which may reduce the graduate teaching assistant's desire to intervene too much in this process.

As we were conducting this study, some limitations became evident. In part due to the COVID-19 pandemic, we only had video recording for one semester of each course, and it would have been preferable to have videos for additional semesters of each course. We also only focused on two investigations for each course, because some attempted recordings for other investigations did not record important aspects of argumentation. Finally, we only had data from two sections of Biology Laboratory I, whereas we had data from seven sections of Biology Laboratory II. As the graduate TAs for the sections differed, having videos from more sections would have provided a more representative picture of the argumentation in these laboratory courses.

This study highlights factors that instructors and curriculum developers should keep in mind when emphasizing argumentation in the laboratory classroom. Studies such as this are necessary to improve lab exercises in the future, increase the level of argumentation, and ultimately increase students' understanding of scientific concepts. These results also suggest that additional research on professional development of people implementing these exercises may be warranted. With further research in other disciplines using ADI lab formats, the benefits of implementing additional ADI labs across multiple institutions can be evaluated. Finally, studies modifying this ADI format for course-based undergraduate experiences (CUREs) would be valuable to determine if making argumentation more explicit in CUREs facilitates more effective use of evidence by students to support their claims.

ACKNOWLEDGMENTS

We thank Katy Hosbein for assistance with interrater reliability calculations. T. Crenshaw, J. Marik, E. Tate, A. Smith Joyner, M. Reyes, V. Wekam, and many graduate student TAs facilitated the video recordings used in this research. Finally, we thank the National Science Foundation (DUE award 1725655 to J.P.W. and H.D.V.-C.) for supporting and funding this research.

REFERENCES

1. Hosbein KN, Lower MA, Walker JP. 2021. Tracking student argumentation skills across general chemistry through Argument-Driven Inquiry using the Assessment of Scientific Argumentation in the Classroom observation protocol. J Chem Educ 98:1875–1887. https://doi.org/10.1021/acs.jchemed.0c01225.

2. Walker JP, Van Duzor AG, Lower MA. 2019. Facilitating argumentation in the laboratory: the challenges of claim change and justification by theory. J Chem Educ 96:435–444. https://doi.org/10.1021/acs.jchemed.8b00745.
3. Walker JP, Sampson V. 2013. Learning to argue and arguing to learn: Argument-Driven Inquiry as a way to help undergraduate chemistry students learn how to construct arguments and engage in argumentation during a laboratory course. J Res Sci Teach 50:561–596. https://doi.org/10.1002/tea.21082.
4. Sampson V, Grooms J, Walker JP. 2011. Argument-Driven Inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: an exploratory study. Sci Educ 95:217–257. https://doi.org/10.1002/sce.20421.
5. Hofstein A, Dkeidek I, Katchevitch D, Nahum TL, Kipnis M, Navon O, Shore R, Taitelbaum D, Mamlok-Naaman R. 2019. Research on and development of inquiry-type chemistry laboratories in Israel. Isr J Chem 59:514–523. https://doi.org/10.1002/ijch.201800056.
6. Walker JP, Sampson V, Zimmerman CO. 2011. Argument-Driven Inquiry: an introduction to a new instructional model for use in undergraduate chemistry labs. J Chem Educ 88:1048–1056. https://doi.org/10.1021/ed100622h.
7. Cheshire PA. 2018. Another way to skin a cat: Argument-Driven Inquiry in the human anatomy laboratory. PhD thesis. University of Texas, Austin, TX.
8. Wolf SF, Sprague MW, Li F, Smith-Joyner A, Walker JP. 2019. Introductory physics laboratory practical exam development: investigation design, explanation, and argument, p 657–662. Physics Education Research Conference 2019. PERC, Provo, UT.
9. Wolf SF, Sprague MW. 2022. Introductory physics labs: a tale of two transformations. Physics Teach 60:372–375. https://doi.org/10.1119/5.0032370.
10. Callis-Duehl K, Crenshaw T. 2019. BIOL 1101 Principles of Biology Lab I. Hayden-McNeil Publishers, Plymouth, MI.
11. Cole R, Muniz M, Harvey E, Sweeney R, Hunnicutt S. 2020. How should apples be prepared for a fruit salad? A guided inquiry physical chemistry experiment. J Chem Educ 97:4475–4481. https://doi.org/10.1021/acs.jchemed.0c00517.
12. Callis-Duehl K, Brewer M, Marik J, Tate E. 2019. Principles of Biology Lab II. East Carolina University, Plymouth, MI.
13. Scott RJ. 2001. Using a molecular marker to study genetic equilibrium in Drosophila melanogaster. Tested Stud Lab Teach 22:19–42.
14. Bond JE, Platnick NI. 2007. A taxonomic review of the trapdoor spider genus Myrmekiaphila (Araneae, Mygalomorphae, Cyrtaucheniidae). Am Mus Novit 2007:1–30. https://doi.org/10.1206/0003-0082(2007)3596[1:ATROTT]2.0.CO;2.
15. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.
16. Sampson V, Enderle PJ, Walker JP. 2012. The development and validation of the Assessment of Scientific Argumentation in the Classroom (ASAC) observation protocol: a tool for evaluating how students participate in scientific argumentation, p 235–264. *In* Khine MS (ed), Perspectives on scientific argumentation: theory, practice and research. Springer Netherlands, Dordrecht, Netherlands.
17. Hallgren KA. 2012. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 8:23–34. https://doi.org/10.20982/tqmp.08.1.p023.
18. Erduran S, Ozdem Y, Park J-Y. 2015. Research trends on argumentation in science education: a journal content analysis from 1998–2014. Int J STEM Educ 2:5. https://doi.org/10.1186/s40594-015-0020-1.
19. Grooms J, Sampson V, Enderle P. 2018. How concept familiarity and experience with scientific argumentation are related to the way groups participate in an episode of argumentation. J Res Sci Teach 55:1264–1286. https://doi.org/10.1002/tea.21451.
20. Walker JP, Wolf SF. 2017. Getting the argument started: a variation on the density investigation. J Chem Educ 94:632–635. https://doi.org/10.1021/acs.jchemed.6b00621.