

RESEARCH

Open Access



Computational thinking through the lens of biological evolution learning: enhancing understanding through the levels of biological organization and computational complexity

Dana Christensen^{1*}  and Doug Lombardi²

Abstract

Research on exploring the relationship between computational thinking and domain specific knowledge gains (i.e. biological evolution) are becoming more common in science education research. The mechanisms behind these relationships are not well understood, particularly between computational practices and biological evolution content knowledge. Increased computational complexity (i.e. simple to complex) may support a greater comprehension of scales or levels of biological organization (i.e. micro to macro) within the context of biological evolution learning. We made use of quantitative methods from qualitative work in the form of coding and relational analysis to identify which biological levels of organization students addressed, how students made connections between these levels and the level of computational complexity displayed during evolution learning with the use of two computational interventions. The aim of this study was not only exploring the biological levels and biological level connections made during the computational thinking interventions, but also analysis of the differences between these two interventions. The results illuminated that use of specific biological levels, biological level connections and differences in computational complexity were distinguishable and there were significant differences between the interventions. These factors may contribute to better understanding of biological evolution knowledge gains.

Keywords Biological evolution, Computational thinking, Biological scales, Computational complexity, Evolution education

Introduction

Computational thinking (CT) allows students to derive meaning from data across disciplines (Nardelli 2019), and learning biological evolution is no exception (Christensen and Lombardi 2023). CT has become increasingly integrated within K-12 science curricula across the globe

(Hsu et al. 2019; Sengupta et al. 2013), yet life science educators struggle with its implementation (Shute et al. 2017; Nardelli 2019). We define computational thinking as thought processes involved in situating problems so their solutions can be carried out by an information processing agent (Christensen and Lombardi 2020; Selby and Woollard 2013) and parallels computational science; the intersection between computer science, mathematics and a scientific discipline. The definition is quite vague and varies across and within disciplines. Educators typically describe CT as students' knowledge development about designing computational solutions to problems, algorithmic thinking, and coding (Angeli and Giannakos 2020).

*Correspondence:

Dana Christensen
DanaC7790@gmail.com

¹ Natural and Mathematical Sciences Department, Stockton University,
101 Vera King Farris Dr., Galloway, NJ 08240, USA

² Department of Human Development and Quantitative Methodology,
University of Maryland, College Park, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Most CT based educational research (situated at the level of the learner) involves the identification of increased knowledge gains, problem-based instruction, access, game design, robotics, engagement and or modeling (Wang et al. 2021; Berland and Wilensky 2015). These studies fail to address the relations between specific: (1) CT components (input, integration, output, and feedback), (2) computational complexity and (3) specific domain content (e.g., biological evolution; Christensen and Lombardi 2020). Learning progressions which merge computational thinking with biological evolution content, support and outline the idea that as computational thinking skills become more complex for students, student understanding of biological evolution improves (Christensen and Lombardi 2020). The use of computational thinking to learn biological evolution learning progression (LBECT-LP) has been supported quantitatively by assessments of both evolution and computational thinking knowledge (Christensen and Lombardi 2023). The mechanism and nuances of how computational thinking supports biological evolution learning is less understood. Our study sought evidence for the mechanism of biological evolution learning through exploring the idea that use of computational thinking may result in student engagement with different biological scales of organization (which may have a relationship to previously established quantitative knowledge gains; [Christensen and Lombardi 2023]).

Learning across scales in biological evolution through computational thinking is unexplored. Evolutionary transitions and multiple levels of complexity are often minimized through typical learning dynamics (Vanchurin et al. 2022). Bidirectional information flows between scales, predictive coding, and active inference may support scale-free conceptual tools for learning complex multiscale systems (Fields and Levin 2020). Our research supports life science educators in addressing the teaching of scales to optimize computational thinking integration into discipline specific topics such as evolution. For example, lessons using computational thinking progressions may explicitly address the gap in student understanding between various scales more efficiently than other modalities especially if the used scales are made apparent and discernable. Certain biological levels and biological level connections may attribute to greater biological evolution learning.

As a society, we are harnessing biological data at a much faster rate than we can understand it due to a lack of computational implementation (Chen et al. 2016). The current gap between biological and computational cultures is particularly large (Rubinstein and Chor 2014). In the few CT studies focusing on evolution, students become embodied and immersed in their

models through agent-based approaches or user-friendly interfaces (Guo et al. 2016; Sengupta et al. 2013; Wilensky and Reisman 2006). Students often fail to develop understanding of evolution between organizational units (or scale levels) in biology as evidenced by recent educational research efforts which attempt to address this gap (Jördens et al. 2016; Dauer et al. 2013). Reasoning is often bound to particular scales or levels, and conceptual linkages may be lacking among these scales as crossing these levels is inherently challenging (Nehm 2019) but required in holistically understanding evolution particularly due to emergent phenomena. This emergent phenomenon gives life to the central idea in biology that the whole is greater than the sum of its parts. This inability to span organizational levels may be due to the simplicity of the computation often used during instruction (Guo et al. 2016), the restrictive nature of the lesson, or the scaffolding provided by the instructor. Use of threshold concepts such as spatial or temporal scale are affected by teaching context and require further exploration (Göransson et al. 2020) therefore we sought to explore presence and relationship of biological levels of organization through their entirety and not restrict our scope to specific levels. Technological tools have aided biologists in the past (e.g. microscopes, electronic probes or digital spreadsheets), but oftentimes educators use these tools at one particular biological level of organization. Our purposeful distinction of biological scale distinctions from micro to macro is intuitive as many biologists, respective institutions and specific units within curricula differentiate themselves as microbiology if their primary focus is on what we define as microscales. Computational thinking enables students to transcend these organizational levels in ways previous tools could not. This is particularly due to the nature of computation and its inherent ability to display emergent properties to the user.

In the present study, we investigated student artifacts (submitted samples of student work) in order to identify the level of computational complexity, presence of biological scales of organization as well as explicit connections made between these scales for two computational interventions. In previous quantitative investigation (Christensen and Lombardi 2023) Students' biological evolution and CT knowledge gains were significant and differed between these interventions. These interventions were developed based on the *Learning Biological Evolution through Computational Thinking Learning Progression* (LBECT-LP; Christensen and Lombardi 2020, 2023) which emphasizes these micro and macro scales and has respectively paired with both computational thinking and specific corresponding NGSS standards at levels of increasing complexity. To our knowledge, no previous investigations have robustly explored relations

and mechanisms between: students' knowledge of computational complexity and evolution across various scale levels of biological organization. Our study is a follow up based on previous quantitative results; we quantified qualitative work to address the following question: In what ways do students' computational products (artifacts) constructed during instruction promoting biological evolution concepts and computational processes (i.e., input, integration, output, and feedback) display:

(RQ1) Student identification and understanding of different scale levels of biological organization (i.e., from molecular to ecosystem scales); and

(RQ2) Different levels of complexity in computational thinking (i.e., simple, developing, and complex)?

Results from the current research study support the LBECT-LP; as different discipline specific content (biological scales), and computational complexity resulted in distinct differences among student artifacts related to biological level connections and accuracy. It may be advantageous for practitioners to more deeply implement computation into their biology instruction, particularly to emphasize biological concepts such as evolution across scales. Computational thinking is a novel, uncharted integral part of science learning. It encourages student engagement with emergent properties by providing mechanisms for students to navigate biological levels of organization to explore biological evolution; providing a meaningful platform and framework to explore how in biological systems the whole may be greater than the sum of their parts.

Theoretical framework

The framework represents theoretical perspectives from educational psychology, science education research, and the biological sciences. Our constructivist framework is grounded in the idea that students learn through specific cognitive processes in their experiences with the computational components given the computational contexts. Students use prior knowledge of the biological scale levels of organization to actively engage through computation in general stages (simple through complex). Research exists on the acceptance, epistemological beliefs and cognitive dispositions (Sinatra et al. 2003) around biological evolution. Age appropriate evolution related misconceptions held by both biology teachers and students (Yates and Marek 2014) may require conceptual change (Heddy and Sinatra 2013). Specific topics to support evolution have been explored, such as natural selection (Brumby 1979), and tree thinking (Novick et al. 2014). Specific learning progressions and mechanisms (Gašperov et al. 2024) have been developed in learning biological evolution, including agent-based modeling (Guo et al. 2016), and different scale levels from the

micro scale (Burmeister and Smith 2016) through to the macro scale (Nesimyan-Agadi et al. 2023) and a variety of testing instruments (Perez et al. 2013) have been developed in supporting these efforts and frameworks. Testing the idea that biological evolution learning is supported by computational thinking is relatively contemporary (Aras-toopour Irgens et al. 2019).

Computational thinking interventions may vary in different ways based on the complexity and disciplinary focus (micro versus macro scales). Disciplinary focus of computational interventions may in turn vary by: (1) the general levels of biological organization exhibited by participants and (2) the connections made between the biological levels. This novel understanding of differences between the interventions may shed light on differences between knowledge gains as displayed by previously published quantitative assessments from the same study. In this section we expand upon this framework by describing the biological scale levels as well as the complexity of computation.

Previous study using the LBECT-LP

We grounded this study's framework using the LBECT-LP (Christensen and Lombardi 2020) which specifies unique (1) computational components: input, integration, output and feedback, (2) computational complexity: simple, moderate and complex, (3) computational perspectives: computational context, computational process and computational product. The LBECT-LP frames biological evolution through specific ordered topics which are anchored to biological unity, and diversity, which generally relate to topics at smaller and larger biological scales respectively. In the present study we specifically focus on the levels of biological organization and complexity of computation because the results from our previous quantitative study revealed statistically significant and meaningful knowledge gains in both biological evolution and computational thinking after instruction integrating biological evolution concepts with CT (Christensen and Lombardi 2023). One of the two interventions, which purposefully differed in scale produced significant knowledge gains in biological evolution, therefore exploring the scales and relationships that students used is the next logical step in understanding the mechanism of the knowledge gains as related to the scales. It is important to note that the interventions to test biological evolution (BECKI) and computational thinking knowledge (CTCKS) were designed and specifically so that there were questions at all biological levels and specific to input, integration, output and feedback respectively. These assessments were validated through content and face and validity practices by professionals and rated as having acceptable reliability through

acceptable Cronbach alpha scores. The quasi-experimental research design (pre-test and series of two post-tests) allowed for the testing of computational thinking knowledge as well as biological evolution knowledge for participants in two separate groups whom experienced different interventions; Participants only experienced one intervention of the two. All participants experienced one post-test after a traditional lesson and after one of the computational thinking interventions to teach biological evolution. Based on the inherent curriculum and the needs of the students and teachers, one intervention was designed to lean towards microscales of biology (BLAST intervention) while the other was partial to macro scales of biology (H-W Weinberg intervention). In the present study we quantified qualitative analyses to more deeply explore these previous quantitative results in knowledge gains that were derived after the use of two unique computational interventions, with the aim to provide highly unique insights for both teachers and researchers (Schulze 2003).

Biological scale levels of organization and computation

Specific biological scale levels (often referred to as levels of organization) and associated descriptions differ among the various biological disciplines (e.g., molecular biology, cellular biology, botany, ecology, etc.; Schneeweiß and Gropengießer 2019). The use of computational methodologies at and between these scales (sometimes referred to as systems biology) has tremendously improved fields such as pharmacology, eDNA research, and precision medicine for complex diseases (Tavassoly et al. 2018). Many biological education studies focus on phenomena at one biological scale level (McEntire et al. 2021), yet most biological functions (including evolution) are a result of mechanisms that occur at various biological scale levels of organization (Reece et al. 2014). We consider biological scale levels based on common textbook descriptions: Atoms, molecules, organelles, cells, tissues, organs, organisms, populations, communities, ecosystems and the biosphere (Campbell et al. 2000; Christensen and Lombardi 2020).

Definitions between scale levels may be: (1) the part-whole relationship, (2) the flow of information relationship, (3) the matter-energy relationship, (4) the coevolutionary relationship and or (5) the phylogenetic relationship (Schneeweiß and Gropengießer 2019). Bonabeau (2002) claimed that computation should be used to explore agents and their interactions, which follow specific instructions, allowing emergent properties to become apparent only at the scale level of collective activities. For example, atoms follow specific rules which make up molecules performing certain functions in cells, which in turn, make up tissues within organisms. Both

the concentration of atoms and rules that cells follow affect the physiology of organisms. A common phenomenon that occurs and has implications at all scale levels is variation, an essential component of evolution (McEntire et al. 2021).

Biological evolution is a core idea that is expected to support biology learning as it is a link between concepts and acts as an exploratory mechanism among all biology units. Tibell and Harms (2017) claim that variation, heredity, selection, randomness, probability, spatial scales and temporal scales presented in tandem with visualizations are imperative to understand evolution, as these ideas occur at and between all levels. Contemporary evolutionary computational practices fail to display major organizational transitions due to a primary focus on small populations, strong selection, and use of direct genotype to phenotype mappings (Miikkulainen and Forrest 2021). Computational evolution is an emerging field which has the potential to solve complex biological issues through the production of algorithms that can easily, quickly, and efficiently integrate across all scale levels (Banzhaf et al. 2006).

The act of learning about various biological scales frequently encourages the use of various tools (such as a microscope for the micro levels). Instead of focusing on these tools to explore the levels (and simultaneously removing interference), computational thinking may allow students to identify and focus on the similarities that are common to the levels or the forces and the phenomena that act between them. Wilensky and Reisman (2006) used computational thinking to promote embodiment at one scale (i.e. thinking like a ‘wolf’ in order to properly assign “wolf-like” parameters such as “eating sheep”) however; computational thinking may also facilitate this embodiment at various scales (i.e. biological levels). The cognitive processes associated with computational thinking may allow students to better holistically understand phenomena (such as biological evolution) by using the same mode of thinking to assign parameters to genes while simultaneously observing emergent processes at other levels such as population dynamics. This is the mechanism behind how a whole may be more than the sum of a systems’ parts. Students may similarly assign parameters to the resulting population dynamics (output) if necessary (in the form of feedback).

“Slippage between levels” is sometimes used to describe the disconnects (or student inability) to make micro to macro scale connections (Brown and Schwartz 2009). “Yo-yo” learning encourages students to think backwards and forwards to prevent this slippage (Knippl 2002). Students’ struggle to make connections between micro and macro levels of evolution through the use of computation (Guo et al. 2016), potentially due

to the simplicity of computation scaffolded into a lesson. Generally speaking, the more biological scale levels students can relate properly (e.g. gene, protein and variation within a population), the greater understanding they have of phenomena (Jördens et al. 2016; Wilensky and Resnick 1999; Penner 2000). If students understand that genes can vary, but not that the gene variation affects proteins produced, the physiology of individuals or physical differences among populations which undergo evolution over time, they may not develop complete understanding of evolutionary phenomena. Christensen and Lombardi (2020) emphasized and outlined the importance of unity and diversity, aligned with recent reformed-based science education frameworks, via the LBECT-LP by intertwining biological evolution and computational knowledge across scales through increasing complexity; computation supports biology learning across scales as evidenced by quantitative study (Christensen and Lombardi 2023).

Complexity of computation

Educators tend to be unfamiliar with CT practices (Rahayu and Osman 2019) and there are no clear methods of CT assessment (Mueller et al. 2017). This poses difficulty in integrating CT through daily classroom practices (Basso et al. 2018) particularly for the life sciences. Three aspects of computation within the context of a classroom setting are the: (1) computational context (as provided by instructor), (2) computational product (as produced by student) and (3) the computational process (actual act of student development of and with the computational components) (Christensen and Lombardi 2020). The student's computational process is the variety of ways that students make meaning about content through computational thinking as facilitated by the instructor which includes the student reasoning and implementation of four computational components: input, integration, output and feedback (as modified from Weintrop et al. 2016; Christensen and Lombardi 2020). The computational product is any artifact produced by students which may be in the form of a verbal testimonial, assessment, questionnaire, project, code etc. whereas the computational context are the factors described by the teacher. Similar to scientists, the computational process may allow students to predict phenomena at a variety of biological scale levels, resulting in a higher level of cognitive engagement and deeper understanding of evolution (Sinatra et al. 2015). For example, as students solve problems (in conjunction with the various computational components), they may think more deeply about data (input), relations between various data and variables (integration), their analyses and results (output), and possibly re-modeling these initial results (feedback). Computational thinking fully encompasses this scientific process through use of the

computational components. In a very general research example, input data may be a series of DNA sequences from various organisms, integration may be a formula to relate similarity between the sequences whereas output may be statistical values representing the similarity and feedback re-ordering the sequences in terms of similarity.

The computational context will be similar among students if lectures or assignments are presented to the class as a whole. Assessing computational artifacts based on specific parameters can allow researchers and eventually educators to determine the displayed complexity of the computation. Students are expected to move from a simple toward a complex computational level. A simple computational context may include step-by-step instructions provided by an instructor, whereas a complex computational context would present research questions which are open-ended. Simple computational products may involve output from interface friendly websites, or Google sheets with simple and or incorrect explanations regarding the computation or phenomenon; whereas complex computational products include software development (using a coding language) that merges the biological phenomenon with sophisticated computational tools. The LBECT-LP encourages use of the computational product to determine growth and mastery of computation at the individual level. Although the LBECT-LP is specific in computational components and complexity, it is purposely vague in what evolutionary based content may be represented and assessed through the complexity for both classroom and research use. It is important to note that modeling is a small but an integral part of computational thinking.

The present study

We hypothesized, that infusing computation within evolution teaching could promote knowledge across unique levels of biological organization and through computational complexity as suggested by our two hypotheses: (1) participant artifacts would reveal discernible relationships between computational intervention exposure and levels of biological organization and biological level connections (RQ1). In addition, we speculated that students with stronger biological and or computational knowledge (as derived from Christensen and Lombardi 2023 quantitative analysis) would display use of different biological levels, different amounts of biological levels, make different numbers of biological level connections, and make specifically larger or smaller biological connections (i.e., connections that span a larger range) within their artifacts. (2) Artifacts would reveal discernible relationships between computational intervention exposure and level of computational complexity. It would be reasonable to speculate that students with stronger computational or

biological knowledge (as derived from Christensen and Lombardi 2023 quantitative analysis) would display distinctly different levels of computational complexity or that the interventions may induce different levels of computational complexity from students.

Methods

In the present study we performed a qualitative (coding artifacts) analysis and quantified our analysis to supplement results from our quasi-experimental, within-subjects repeated measures research study (Christensen and Lombardi 2023). The purpose of this study was to further investigate the effectiveness and nuances of the LBECT-LP through exploring complexity of computational thinking and unique use of biological levels of organization. Specifically, to identify participant use and explicit associations of different levels of biological organizations (RQ1) and the levels of complexity of computational thinking (RQ2) while students learned biological evolution.

Setting and participants

We conducted the present research study within the two large public high schools (East and West School, pseudonyms) within the Pine Bay School District (pseudonym), located in the mid-Atlantic region of the United

States. A researcher involved in the study is also an educator within the district, as a form of action research (referred to as researcher teacher). Study participants were enrolled in Advance Placement (AP) Biology (College Board 2019) for the 2019–20 academic school year, within one of four classes taught by one of two teachers. AP courses are the most rigorous courses offered at the high school for each subject, they follow a standard international curriculum through College Board and prepare students for an end of the year exam in which they may earn college credits. Each of the two AP biology teachers resided in one of the schools in the district was responsible for two AP biology classes each. A total of fifty-one student participants participated in the study, with $n=21$ participants at West School and $n=30$ participants at East School. About 42% of the participants ($n=27$) identified as male. Table 1 shows the number of students and demographics at each school.

Procedures

The length of the full administration of both computational interventions was approximately four weeks, and we used a quasi-experimental research design (see Table 2). When limited to quasi-experimental design a counter-balance comparison of interventions among class groups is more robust than comparing changes over

Table 1 School demographic statistics

	East High School	West High school
Number of students	1558	1285
White	84%	77%
Hispanic	7%	13%
Black	6%	7%
Asian	3%	3
American Indian/Native	< 1%	< 1%
Multiracial	< 1%	1%
Female	46%	49%
Male	54%	51%
Free/Reduced Lunch (Economically Disadvantaged)	14%	29%
College and Career Readiness	30th percentile	30.5 percentile
Staff to student ratio	14:01	13:01
Graduation Rate	94%	94%
Math Proficiency	23%	13%
Reading Proficiency	28%	37%
National Percentile on AP exams	83%	75%
Proportion of 12th grade students who took an AP course	43%	35%
Students who score over a 3 on AP test	30%	23%
12th grade students who score over a 3 on AP test	60%	68%
Pass rate (over 3) on AP Exams	59%	54%

Neither high school is entitled to Title I funding. These statistics were obtained through the US News and World report for the 2016–2017 school year (the most recent year in which data were available; Christensen and Lombardi 2023; US News World Report, 2017)

Table 2 Quasi-experimental study design

Class	Time 1 (Pretest)	Lesson/teacher	Time 2 (Posttest 1)	Lesson/teacher	Time 3 (Posttest 2)
A	BECKI and CTCKS	Traditional H-W Lesson with West School Teacher	BECKI and CTCKS	Lesson 2: Computational BLAST Lesson with Researcher Teacher*	BECKI and CTCKS
B	BECKI and CTCKS	Computational H-W Lesson with Researcher Teacher *	BECKI and CTCKS	Lesson 2: Traditional BLAST Lesson with West School Teacher	BECKI and CTCKS
C	BECKI and CTCKS	Traditional H-W Lesson with East School Teacher	BECKI and CTCKS	Lesson 2: Computational BLAST Lesson with Researcher Teacher*	BECKI and CTCKS
D	BECKI and CTCKS	Computational H-W Lesson with Researcher Teacher *	BECKI and CTCKS	Lesson 2: Traditional BLAST Lesson with West School Teacher	BECKI and CTCKS

*Indicates when artifacts were collected (Christensen and Lombardi 2023)

time in one class group, who did the computational interventions only, to another class group, who did traditional lessons only when classroom placement has already dictated the independent groups (Cook and Campbell 1979). These limitations were unavoidable because we could not limit student learning nor dictate truly random groups, however, other indices of homogenization have been addressed and the test design involves a uniform series of testing and artifact collection.

All four classes were given pre-tests and a series of post-tests after each biological evolution lesson. Two lessons were traditional in nature and taught by their teacher and two were the computational interventions taught by the researcher teacher; see Table 2). The artifacts collected and analyzed in the present study were collected after the computational interventions only with the intention to support the previous qualitative analysis on computational and evolution knowledge gains as derived from the pre- and post-test analysis (Christensen and Lombardi 2023). There were no significant knowledge differences between class grouping before presenting the interventions through calculations of Mahalanobis distances and z-scores of the pre-tests. Artifacts include samples of submitted student work which were open ended questions designed to elicit reflections of interactions with the computational tools. These artifacts were researcher made based on the AP lab manual, the biological content and the computational tools and submitted digitally to the respective teachers through an online classroom management tool.

Computational interventions

We developed two computational interventions, which were modified lessons from the AP Biology Lab Manual (College Board 2019), in order to maintain the integrity of the AP biology curriculum and to reduce stress for the instructors and student participants while maintaining the integrity of the study design. It is important to note that these interventions modified the original content

and objectives and made them more computational in nature as per specific guidelines in the LBECT-LP, and we assumed some predictive power of these materials due to their original publication within the AP curriculum. The biological content and associated computational context differed between intervention groups and there was no true control group. The shared feature of both interventions was biological evolution taught through computational thinking. These activities naturally incorporate biology and computation yet many activity descriptions found in standard educational manuals are unclear for instructors and the provided examples are outdated (Moreno-León et al. 2017).

For each intervention we developed a novel electronic slide presentation incorporating the computational components with the appropriate biological evolution concepts, classroom activities, instructional sheets, and respective worksheets with questions to be completed during the interventions (used as artifacts). During the computational interventions, the researcher teacher described the four components of computational processes (input, integration, output and feedback), within the appropriate biology context; Intervention 1 focused on the Hardy Weinberg Lesson and Lesson 2 focused on the development of phylogenetic trees through DNA analysis. The researcher teacher then passed out an assignment sheet to each participating student and allowed them to complete the assignment on a Google Doc which included open ended questions and a location to link their computation. Students submitted work individually to their instructors on their Google Classroom page, and it was forwarded to the researcher teacher. Students worked during class and were allowed to finish the assignment after school hours. Researcher teacher to participant interactions and participant interactions were not prompted, but also not discouraged. Aside from biological and associated computational context the interventions and associated questionnaires were as comparable as possible. The two

lessons focused on primarily micro versus macro scales; however, they were not restricted to those scales. The levels that students engaged with (and identified) were somewhat emergent and identified in the results. Within each intervention we explored the scales identified and connected by students as well as the associations with knowledge gains (from previous study) and novel understanding.

Computational Hardy–Weinberg (H-W) Lesson (Computational Intervention 1)

Computational intervention 1 consisted of a lesson around the Hardy–Weinberg (H-W) Law of Genetic Equilibrium. After the lesson students were shown an example Google doc with specific biological features of a hypothetical population and how the computational components related to this population. Participants were prompted to relate evolution and the change in allele frequency through the use of a computational tool (in Google Sheets) to demonstrate evolutionary phenomena over many generations. Participants then independently developed a hypothetical population with specific features, then used the H-W formula to display the allele frequency of the population in evolutionary equilibrium (i.e., not evolving). Participants then modified ratios based on a hypothetical event that would affect the environment (i.e. climate change, drought) by generating an appropriate algorithm. Participants observed if and how their population was evolving or not based on their working computation and application of the H-W equation. In cases when the population was evolving, the questionnaire sheets prompted students to write why (i.e., environmental changes, advantageous adaptations, human impacts etc.). Participants worked with computer programs such as Microsoft Excel or Google Sheets to develop a spreadsheet with all computational components that mimicked two successive generations without environmental change and then two generations with an induced environmental change. Participants designed their computational products to display an output so that there was a single input value to influence the rest of the model. Participants were prompted to recognize input (single allele frequency value), integration (H-W equation and other written algorithms), output (resulting generations and graphs) and feedback (successive generation influence) within their models. Ten questions for students included items such as: “Identify what you observed in your populations and how it related to the H-W equation and evolution” and “Overall what did your results display; and how do these results display evolution?”.

Computational BLAST Lesson (Computational Intervention 2)

During computational intervention 2 the researcher teacher taught a lesson around phylogenetic trees and showed participants how to navigate the National Center for Biotechnology Information (NCBI) website (including BLAST, which is a basic local alignment search tool) (Geer et al. 2019). Participants compared DNA and protein sequences to understand evolutionary relationships using computational tools (computational intervention 2). Participants were shown separate examples of input (Google doc of sequences), integration (NCBI features) and output (phylogenetic tree and associated statistics), then created their own phylogenetic trees based on a protein sequence of their choice. Participants were prompted to recognize input, integration, output, and feedback within their analysis procedure to support student awareness and interaction with computational thinking the way it is defined in the learning progression. Input involved the participant selecting either an amino acid or DNA sequence for their selected protein, while the integration was done by the BLAST tool within NCBI itself (code hidden from students). The output was in the form of a unique phylogenetic tree coupled with additional statistics. For this intervention, there was no form of feedback. Ten questions for students included items such as: “How do the DNA relationships compare to the morphology or what you initially thought about the relationships between these organisms?” or “What other questions might you ask about your protein or the evolutionary relationships?”.

Qualitative data collection: situated

Results from previous quantitative study (Christensen and Lombardi 2023) indicate that there was a main effect for overall computational thinking knowledge and evolution knowledge due to the computational lessons based on pre- and post-test analysis of two verified assessments; however, these gains were only significant for intervention 1 (H-W lesson). Participants involved in this lesson experienced a significant increase in computational knowledge, specifically at time 2 (which was directly after their computational lesson intervention) (Christensen and Lombardi 2023). These participants experienced a small non-significant decrease in computational knowledge between time 2 and 3 indicating that knowledge was not retained through quantitative analysis of the second post-test. The present study used qualitative analysis on student artifacts collected after the computational interventions with the aim of probing more deeply into these relations and to better understand and enhance prior quantitative results (Bryman 2017). Quantitative analysis in science education research is often overshadowed

by quantitative, as it is still favored by many researchers (Stanley and Robertson 2024), yet it may provide unique insight, particularly when done to supplement quantitative analysis (Borrego et al. 2009, Eyisi 2016).

Coding analysis

In this section, we present the development and results of our coding analysis, then discuss the results of the biological level scoring and associations that emerged from the coding analysis (RQ 1). We follow this with the development of the computational complexity rubric. We used the rubric to score computational complexity for the participants' written artifacts based on the LBECT-LP. We then present and discuss the associated computational complexity results (RQ 2).

To ensure the accuracy and reliability of coding, the author completed two rounds of analysis and revision, the second author verified the utility of the coding scheme and process. An educator of biology at the secondary level within the Pine Bay District coded 10% of the artifacts which resulted in an 93% agreement, which is considered a fairly high consensus for intercoder reliability (Stemler 2004). Coders discussed the areas of disagreement and difficulty in coding at various points in the coding process, with the following sections providing more detail on how we arrived at our final coding scheme.

Conceptual and relational analysis of written explanations: biological levels (RQ1)

Content analysis has two major forms, conceptual analysis and relational analysis both of which we used to address RQ1. In conceptual analysis, a concept is chosen

and the analysis involves quantifying its presence (Hsieh and Shannon 2005). Conceptual analysis is particularly important in science education research because it allows various terms (used by student participants) to represent a larger construct (identified by researcher). This allows the researcher to develop an appropriate concept definition (i.e., the concept of "cell") with specific borderlines for rich results (i.e., neurons and eosinophils are two very different terms for the construct of "cell"). Explicit terms representing biological organization were the items we selected for within the participant data to answer RQ1 and range from level 1 (L1) through L14 (see Table 3). These levels were selected prior to coding participant artifacts. Participant identifiers of these biological levels varied greatly due to the open ended nature of the intervention questions.

Relational analysis is a specific form of content analysis which explores explicit relationships identified (by participants) within texts. It provides higher levels of statistical rigor as compared to other qualitative methods used in educational research because it allows researchers to make specific inferences from participant artifacts (Robinson 2011). It explores the instances in which participants made connections, and not solely identifying the presence of a concept for example (i.e. allele changes as it relates to populations versus the identification of allele changes or populations within artifacts alone). Artifacts were coded using relational analysis to explicitly identify if the participants made connections between biological levels. This was more meaningful than identifying if the two levels occurred together through statistical means (i.e., a bivariate correlations). We also performed a relational analysis to identify and better understand

Table 3 Biological levels identified from participant artifacts

Level	Biological level	Participant examples	Concept definition
14	Biosphere	No participants recognized this level	The earth and interactions of its regions
13	Ecosystem	The habitat, the environment	Community of organisms and their physical environment
12	Community	Biotic factors, all living organisms	Interacting group of species in common location
11	Species	<i>Mus musculus</i> , all dogs	Group of organisms that can produce viable offspring
10	Population	Bird population, the next generation	Organisms of a species in a particular location within a given time
9	Organism	The fish, the organism	Discrete and complete individual living thing
8	Organ System	Central nervous system, the bones	Group of organs working together to perform function
7	Organ	Heart, skin	Group of tissues which perform a specific function
6	Tissue	Blood, muscle tissue	Similar cells which carry out a specific function
5	Cell	Red blood cell, nerve cell	Smallest structural and functional unit of an organism
4	Organelle	Ribosome, mitochondria	Specialized structures within a cell which perform a function
3	Macromolecule	Protein, DNA	Large complex molecule
2	Molecule	Glucose, gene sequence	Two or more atoms covalently bonded
1	Atom	Oxygen, hydrogen ion	Smallest unit of matter that contains its properties

Concept definitions modified from Campbell et al. 2000

the nature of relationships participants made between these biological levels. There is a major gap in educational research around conceptual and relational coding analysis to quantify student use of biological levels and we have suggested in our framework that these levels are essential to learning biological evolution.

Defining biological levels of organization We examined the participant artifacts that were generated from questions in both interventions. The questions were designed to be answered during and directly after the interventions (unanswered questions completed for homework). Questions were specific to the intervention and there was one shared computational question between the interventions (“Identify the input, integration, output and feedback”). Most questions within the interventions prompted students to identify and describe computational and biological components in an open-ended form (for example: “Identify what exactly you observed occurring in your populations” or “How does BLAST [computational tool] actually compare these [molecular] sequences?”). Due to participant choice to work with infinite biological scenarios (i.e., pick a protein of interest), We had to examine participants’ answers in detail to identify which biological levels were present based on participant use of key biological words. We read through the participant artifacts multiple times until no new biological words (and associated biological levels) emerged.

Biological words within participant artifacts were used to classify the presence of biological levels based on the LBECT-LP. We derived concept definitions for the biological levels (Table 3) from *Biology in Focus*, a textbook written by Campbell, Mitchell and Reese (2000) which is accepted by AP biology curriculums and many college level biology courses.

We counted biological words if they “fit” one of the 14 biological concepts. For example, if a participant wrote “neuron” (a type of nerve cell), that was counted as a biological word, classified and quantified as one attempt at the cellular level (L5). If a participant wrote “the environment”, we classified it at the ecosystem level and it counted as one attempt at the L13 level. We defined these concepts based on important biological properties associated with each level. For example, molecules (L2) comprise macromolecules (L3) or environments are the interaction of the living and nonliving components of an ecosystem (L13). We recognize that some of our defined ‘levels’ may be imperfect in terms of classification. For example, technically the way oxygen is found in nature (within air for example) it is a molecule (O_2), but when broken apart and used in the body it may become (or act as) an individual atom. However, oxygen in the way participants use it, one atom vs two, is not the focus here. In the instances that participants used it, oxygen’s overall

behavior in biological systems better aligns with the atom level; therefore, we classified oxygen at the atom level (L1). Usage was not always explicit within the artifact and in those (few) cases we did not quantify the construct.

There are inconsistencies among scientific fields between how many bonded atoms comprise a molecule versus a macromolecule (i.e., a macromolecule is a large molecule). Biology courses spend time on this differentiation, the typical macromolecules presented to biology students are: proteins, nucleic acids (DNA or RNA), carbohydrates and lipids, therefore we defined molecules of this size at the macromolecule level. Some of the macromolecules participants identified were “proteins” and “DNA”. If portions of these macromolecules were referenced (monomers), we classified them at the molecule level (“amino acid” or “nucleotide”). These words came from the individual conditions participants developed during instructional activities. These discrepancies would naturally occur in typical teacher grading and identifications of accurate use of these levels (by the students) was on par with typical grading expected by an AP level biology teacher.

Other levels that were important to distinguish were organism (L9), population (L10) and species (L11). Organisms are individual living beings. Populations are groups of organisms, usually living in the same region, and are composed of multiple generations. Species are all of the organisms that have the ability to viably reproduce (which is determined by DNA similarity) and are assigned specific scientific names. All three of these definitions are blurry, even to those within the scientific community. For example, *Canis lupus* (wolf) and *Canis familiaris* (dog) are different species due to the fact they normally would not be able to interbreed given where they typically reside, however they do have the ability to produce viable offspring and are sometimes referred to as *Canis lupus familiaris*. Biology texts claim that a species is a group of interbreeding natural populations or one can say it is a reproductively cohesive assemblage of populations; making the emphasis on the genetic relationship as it is a property of populations not individuals nor based on morphological difference (Mayr 2000). The evolutionary process itself dictates that species may be units of evolution if they are made up of organisms related by descent. Depending on the definition species possess all the characteristics of individuals however there is a difference between organisms and species in that individuals contain a fixed genetic makeup whereas species do not (Hull 1976). We felt it important to make these distinctions based on our data particularly because they differed from textbook levels of organization. We classified biological levels based on the ways participants used them contextually.

We counted and classified all of the biological words within each participant's artifact that would distinctly represent an organizational level. If participants referenced a specific protein such as "keratin" it was classified at the macromolecule level (L3) and if the participant referenced the word "protein" again in a different context it was counted as another word at the L3 level. It was possible for participants to identify multiple biological words at each biological level. If participants used the word "keratin" more than once however, this second instance was not counted again. If it was not clear that the participant identified the level (i.e., there was no context given or example was too vague) it was not counted. We quantified the total number of "biological words identified" as well as the "biological levels attempted" which were identified for each participant from their digitally submitted artifact. This type of counting and quantification of the biological levels and biological level connections was a meta-analysis above what standard AP teachers would be responsible for; as educators typically grade on the individual level and generally make note of issues at the class level.

We also identified "levels correctly identified" by excluding incorrect attempts in which the participant used the inappropriate context with the biological word. In theory each biological word represented a biological level, however not all representations were accurate. For example, if a participant used the word "organism" but used it in the context of a species, we counted it as an incorrectly identified instance at the organismal level (L9). It is also important to note that not all participants turned in an artifact and not all participants completed all questions for the interventions. If students did not answer all questions within artifacts but there was enough meaningful information present within answered portions for scoring (greater than 80% of questions responded to) they were scored. This on average may have reduced the reported averages of biological levels and biological level connections. There were a total of 39 participants who turned in artifacts with 80% of the questions completed ($N=39$) therefore less artifacts were scored as compared to those who completed the interventions and respective pre- and post-tests from the previous study. The mistakes we identified were on par with what AP biology teachers would correct in their students' work.

Relational analysis In order to address RQ1 more robustly, we wanted to determine how participants understood the relationship between the biological levels. We addressed this by identifying how many explicit connections were made between the identified biological levels. For example, if participants claimed that "allele frequencies" (molecule level, L3) "influenced"

"generations" (population level, L10) we counted it as a connection attempt with 7 levels in-between. We read through the artifacts until no new biological level connections (BLC) emerged. We counted explicit connections that participants made, the number of levels in-between and if the connections were "accurate". These connections were prompted by various questions (not explicitly asked about). The nature of the biological level and biological level connection identification are on par with what AP biology teachers are capable of; but analysis of these items at the class or district level is not a standard practice.

We selected Level 9 as a pertinent distinction between "micro" and "macro" levels within this study. Level 9 (organism level) was selected because students use vastly different methods to study levels below L9 as compared to levels above L9 (i.e., microscopes vs ecosystem models). Facets of biology that are smaller than L9 include cellular biology or anatomy. Facets of biology that are larger than L9 include ecology or population dynamics. Fewer biological studies (and corresponding biology units within classrooms) tend to bridge these levels. Biological unity (micro) and diversity (macro) are frequently associated with smaller or larger biological levels, although they may be seen at and between all scales. Both concepts of unity and diversity contribute to the understanding of biological evolution. We suspected that students require connections between this level to fully grasp evolution at all scales. The following table is what was used to identify the biological levels and connections, there were 14 rows one for each level (Table 4). A completed example with a student artifact is provided in Appendix A.

Analysis of computational complexity (RQ2)

We scored complexity for all of the computational components (input, integration, output and feedback) as either "simple", "developing" or "complex" based on participant description (as classified according to the LBECT-LP). Each intervention had at least one question which explicitly required students to "identify each of the computational components within their activity". Accuracy and complexity definitions that we used to code artifacts are outlined in Table 5. Participants scored no points for absence of the component or a 1 within the simple category, developing category or complex category for each component (if it was mentioned).

For example, if a participant described input as "The FASTA sequence" the participant scored a 1 in the input category because this is how the researcher teacher described the input during the lesson. If participants explicitly identified their FASTA sequence accurately and identified what it represented biologically they scored a 1 in the developing category for input (i.e., "The input was

Table 4 Example coding diagram for participant artifacts

	Biological words student uses to identify biological level(s)	Which biological levels (1–14) and accuracy	How many properly identified level(s), supporting explanation present	How many connection attempts made between levels?	How many levels between these attempts?	Was correct connection made between levels?	Other Notes
List all biological levels 1–14 (even if students do not use all levels)	List the words that student participants use to represent biological levels within their artifacts	For each level identified specify the accuracy	Identify how many levels and the presence of an explanation and or proper use of term/s	Identify how many connections were made between levels and use arrows to connect levels	Identify how many levels were between each connection attempt	Identify the accuracy of the level connections	List notes about connections within student artifact

The instruments used had 14 rows, one for each biological level

Table 5 Scoring rubric for computational component complexity within participant artifacts

Component	No points	Simple	Developing	Complex
Input	Absence of the word input	Participants describe input as provided by the instructor	Participants uniquely decide on the input data for their computational component, recognize what it is and identify correctly in writing	Participant able to find and modify input data on their own
Integration	Absence of the word integration	Participants describe integration as a formula or program	Participants accurately describe which variables are interacting within the integration	Participants describe which variables are interacting and how they are doing so explicitly within their computational representation
Output	Absence of the word output	Participants describe as a graph or model	Participants correctly identify their output and make the connection between the visual output and the integration	Participants have multiple representations of output
Feedback	Absence of the word feedback	Participant recognition that the model re informs itself	Participant describes how output rein forms model	Participant description of how output rein forms model and provides alternate ways
<i>Model</i>				
Model Present	No Model (0 points)	Model Present (1 point)		
Model Complexity	No Model	Present and similar to instructor examples	Working and distinctly different from instructor examples	Participants use alternate computational tools to generate models. (No examples provided for participants at complex level)

Classified according to the LBECT-IP (Christensen and Lombardi 2020)

the coded amino acid sequence of keratin from a common mouse which I found from the NCBI database”). If participants used an alternate database to find DNA or amino acid sequences on their own and described this as their input, they would have received a 1 in the complex category for input. If participants failed to mention input, they received zero points in the input category. We used this same process to categorize complexity for integration, output and feedback for each participant based on the parameters in Table 5.

Along with complexity, we also identified if participant models were present within their artifacts and how complex the models were (under the parameters of integration within the LBECT-LP). Instructions indicated that participants should either supply links or screenshots of their working computational models which they developed during the interventions, however not all participants followed these directions. Simple models varied very little from the provided examples. Developing models showed participants manipulating their models based on computational instructions. Model level ranged from 0 (absent) through simple or developing based on parameters in Table 5. According to the complex category from the learning progression, participants would need to display multiple unique and accurate models with little teacher scaffolding to score in the complex category (none of which did so).

Most participants who mentioned the computational components scored within the simple or developing categories. After receiving one computational lesson, the learning progression predicts few or absent instances of complex computational components. Teachers and researchers can use the learning progression similar to the way we did to develop rubrics for specific lessons and select the applicable components to include based on classroom instructional activities and situate them for their level and range of students. Table 6 represents an example of the diagram used to code student artifacts for computational complexity and computational components. Student example found in Appendix A.

Results

We initially present results for *RQ1* as related to biological levels of organization and biological level connections followed by a comparison of the interventions. We then present results for *RQ2* and present levels of computational complexity, followed by a comparison of the interventions. The results from this study are quantitative, from the qualitative coding of artifacts only in an attempt to support and explain the knowledge differences from the biological evolution and computational thinking assessments as participant knowledge gains between the

two computational interventions significantly differed (Christensen and Lombardi 2023).

Biological levels of organization (*RQ1*)

Research question 1 specifically seeks to answer how artifacts developed during computational interventions display different levels of biological organization. The average number of biological words per participant was 5.75 ($SD=2.38$) with a minimum of zero and maximum of 12. The average biological levels attempted per participant was 4.28 ($SD=1.64$) with a minimum of 0 and maximum of 8 levels. The average of correctly identified levels per participant (excludes incorrectly identified biological levels) was 3.38 ($SD=1.53$) with a maximum of 7 correctly identified levels (and minimum of 0).

We calculated the proportion of participants who attempted each level as well as the proportion of participants who accurately identified each level. We represent these proportions as percentages throughout the rest of the study for readability (“percent attempted” and “percent correct”). Approximately 89.7% of participants identified L2 (molecule), 59.0% of participants identified L3 (macromolecule) and 56.4% of participants identified L11 (species). The levels attempted least frequently were L12 (community; 10.3%), L6 (tissue; 10.3%), L8 (organ system; 2.60%), and L14 (biosphere; 0%).

We also calculated the percent correct (for each biological level) from the difference between the attempted percent and the accurate percent which represents the mean for overall participant accuracy. For example, 30.8% of participants attempted L7, and 23.1% identified it accurately resulting in 90.3% correct for L7. “Percent correct” reveals the biological levels participants had difficulty representing most often within their artifacts (and this is not necessarily representative of the levels that were used most often). Biological levels in which less than 70% of participants identified correctly were L12 (community; 50.0%), L11 (species; 32.1%), L8 (organ system; 0%) and L4 (organelle; 57.2%). Whereas students most often correctly identified L1 (atom; 100%), L2 (molecule; 100%) and L7 (organ; 90.3%). Table 6 displays the percentage of participants who attempted each level (along with accompanying *SDs*), the percentage of participants who correctly attempted each level (and *SDs*) as well as the percentage correct for each level (representing overall participant accuracy). Table 7 also includes Kendall’s *W* rank which are values we used to determine ordinal associations between the biological levels.

The resulting *W* value of 0.295 was low. This low *W* value indicates that there is little agreement in the frequency of the biological levels used among the participants (Legendre 2010). In other words, there was a large variation of biological levels used among the participants

Table 6 Table representing coding scheme for computational complexity and computational components

	Key biological words mentioned with term	Simple	Developing	Complex	Notes
Computational Construct: Input, Integration, Output, Feedback	Words listed from artifact as associated with computational component	Construct mentioned	Construct mentioned Properly with biological construct	Construct mentioned clearly, description provided and correct	Additional notes
Computational product: Model/ Output display (present)	Words listed from artifact within model	Construct mentioned	Construct mentioned Properly with biological construct	Construct mentioned clearly, description provided and correct	Additional notes
Computational process: Description of process	Is a description of process present	Not present or incorrect	Correct output	Correct description provided	Additional notes

Table 7 Conceptual analysis: biological level totals including attempted (SD) percent, accurate percent (SD), overall percent correct and Kendall's W Rank

Level	Attempted (%)	SD	Accurate (%)	SD	(%) Correct	Kendall's W Rank*
14	0	–	0	–	–	–
13	25.6	4.42	20.5	4.09	80.7	6.53
12	10.3	3.07	5.10	2.23	50.0	5.53
11	56.4	5.02	17.9	3.98	32.1	8.53
10	48.7	5.06	38.5	4.93	77.6	8.03
9	43.6	5.02	35.9	4.86	81.8	7.69
8	2.60	1.60	0	–	0	5.03
7	30.8	4.68	28.2	4.56	90.3	6.86
6	10.3	3.07	7.70	2.70	80.0	5.53
5	30.8	4.68	23.1	4.27	74.2	6.86
4	14.0	3.51	7.70	2.70	57.2	5.36
3	59.0	04.98	51.3	5.06	86.4	8.69
2	89.7	3.07	89.7	3.07	100	10.7
1	12.8	3.39	12.8	3.39	100	5.69

*Kendall's $W = 0.295$, $\chi^2(12, N = 39) = 138$, $p < 0.001$

(this pool of participants was exposed to either intervention 1 or 2). A chi-square test of independence was used to show significance of W with, $\chi^2(12, N = 39) = 138$, $p < 0.001$. Figure 1 displays the percentage of attempts of biological levels in decreasing biological size order along with the correctly identified percents at corresponding biological levels. This figure explicitly displays which levels were identified most by participants and which they may have had trouble correctly identifying.

A Kendall's tau-b (τ_b) correlation was used to identify the strength and associations for all biological levels within each artifact. We used this correlation because it is robust against outliers (as compared to Spearman correlations; Schober et al. 2018) and the data were not normally distributed. Effect size (τ_b) determines the strength of the correlations or how strong the relationship is between the variables. Positive effect sizes

indicate the biological levels that participants tended to use together, whereas negative effect sizes indicate levels that participants tended to avoid using together. Effect sizes less than 0.3 are considered "small", effect sizes between 0.3 and 0.5 are considered "medium" and effect sizes greater than 0.5 are considered "large" (Schober et al. 2018). All of the Kendall's tau-b correlation effect size (τ_b) values can be found in Table 8.

These correlations partially answer RQ 1 through indicating which variables participants tended to identify within the same artifact (positive) or which variables participants tended to avoid using together (negative). These correlations are indicative solely of biological level identification, not to be confused with biological connections intentionally and explicitly made by participants (found in the relational analysis).

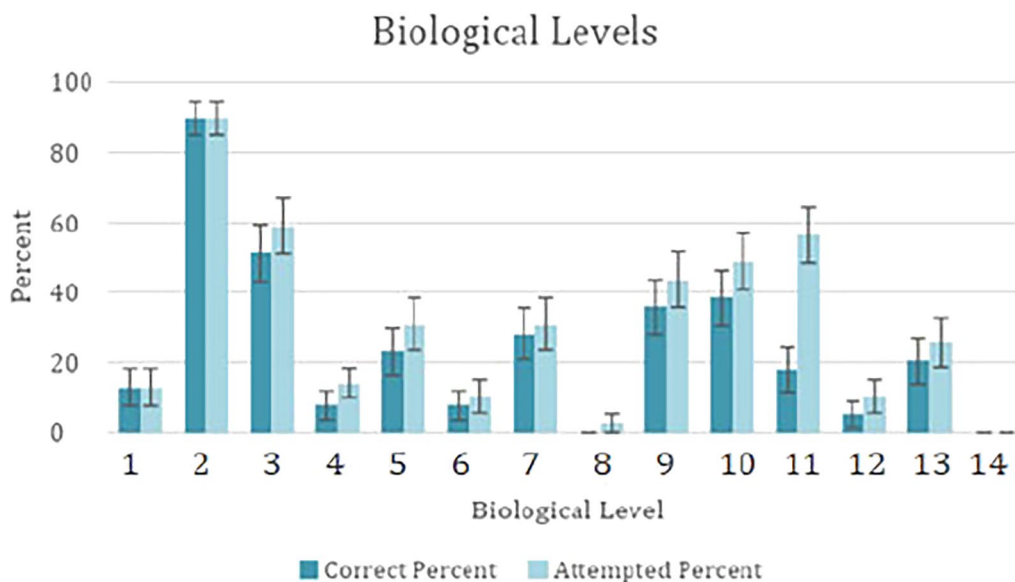


Fig. 1 Biological level percentages attempted by participants and percentage of correct instances, with bars showing ± 1 standard error. Biological levels represented 1 = atom, 2 = molecule, 3 = macromolecule, 4 = organelle, 5 = cell, 6 = tissue, 7 = organ, 8 = organ system, 9 = organism, 10 = population, 11 = species, 12 = community, 13 = ecosystem, 14 = biosphere

Table 8 Kendall's tau-b correlations for the biological levels within participant artifacts

Level	1	2	3	4	5	6	7	8	9	10	11	12	13
1	–												
2	0.130	–											
3	0.320*	–0.110	–										
4	–0.111	–0.220	0.241	–									
5	–0.890	–0.141	0.217	0.016	–								
6	0.376*	0.114	0.282	–0.098	0.141	–							
7	0.243	0.225	–0.122	–0.192	0.157	0.507*	–						
8	–0.062	0.055	0.135	–0.047	0.243	–0.055	–0.108	–					
9	0.127	0.127	–0.108	–0.254	–0.138	0.044	0.086	–0.143	–				
10	–0.374	0.330*	–0.543*	0.089	–0.094	–0.160	0.239	–0.156	0.0741	–			
11	0.182	–0.127	0.633*	0.254	0.250	0.297	–0.086	0.143	–0.166	–0.385*	–		
12	0.123	0.114	0.282	–0.098	–0.042	–0.114	–0.225	0.480*	0.214	–0.330*	0.127	–	
13	–0.255	0.199	–0.585*	–0.170	–0.010	–0.199	0.117	–0.095	0.431*	0.368	–0.313	–0.00511	–

Biological levels represented 1 = atom, 2 = molecule, 3 = macromolecule, 4 = organelle, 5 = cell, 6 = tissue, 7 = organ, 8 = organ system, 9 = organism, 10 = population, 11 = species, 12 = community, 13 = ecosystem, 14 = biosphere (L14 was not identified by participants). (N = 39) All significance listed at * $p \leq 0.05$

Relational analysis: biological level connections

We calculated the means and standard deviations of biological level connections (BLCs). There was an average of 2.56 “BLC attempts” per participant ($SD = 1.25$), while “the average levels between attempts” was 5.59 ($SD = 2.22$). Because participants made multiple BLC attempts, the minimum number between levels “levels between min” ($M = 3.46$, $SD = 3.11$) as well as the maximum number between levels “levels between max” ($M = 7.87$, $SD = 2.28$) was also identified for each

participant. The means and standard deviations of the BLC constructs are displayed in Table 9. Accurate BLCs on average were smaller (in biological scale) than those for participants which attempted more levels in-between.

We further quantified and classified the types of BLCs that participants made into one of 3 groups from our qualitative analysis. It is important to note that we only counted instances of these three types of connections if they were accurate within the participant artifacts. The first type of connection were connections

Table 9 Relational analysis: summary of total biological connections means and standard deviations

BLC construct	<i>M</i>	<i>SD</i>
BLC attempts	2.56	1.25
Accurate connections	1.64	1.01
Levels between min	3.46	3.11
Levels between max	7.87	2.28
Average levels between attempt	5.59	2.22
Average between accurate attempts	3.98	3.08
Micro level connections (L1–L9)	1.03	0.986
Macro level connections(L9–L13)	0.10	0.384
Micro to macro connections(L1–L9)	0.44	0.598

(*N*=39) Values represent attempts per participant. Levels between refer to biological levels 1–14

which were made below L9 (connections between L1 through L9, which we refer to as “micro levels connections”) such as a connection between macromolecule (L3) and cell (L5). Eleven participants (28.2%) identified 1 micro level connection, 10 participants (25.6%) identified 2 micro level connections, 3 participants (7.7%) identified 3 micro level connections and 15 participants (38.5%) did not identify this type of connection. Twenty-four students of the total thirty-nine (*N*=39) correctly made at least 1 micro level connection with an average of 1.03 micro level connection per participant (*M* = 1.03, *SD* = 0.986).

The second type of connection were the connections above L9 (connections between L9 through L14, which we refer to as the “macro level connections”). An example of a macro level connection would be between populations (L10) and ecosystems (L12). Two students made 1 macro connection (5.1%), 1 student made 2 macro connections (2.6%) and 36 students (92.3%) did not make any macro connections. Three students correctly made at least one macro level connection with an average of 0.10 micro level connection per participant (*M* = 0.10, *SD* = 0.384).

The third type of connection were the connections that were made directly through level 9 such as connections between molecule (L2) and population (L10) (“micro through macro level connections”). Thirteen students (33.3%) made 1 micro through macro level connection, 2 students (5.1%) made 2 micro through macro connections and 24 students (61.5%) did not make any micro through macro level connections. Fifteen students made at least one accurate micro through macro level connection with an average of 0.44 micro through macro level connections per participant (*M* = 0.44, *SD* = 0.598).

Table 10 Means and standard deviations for biological words and levels identified between intervention

Category	Intervention	<i>M</i>	<i>SD</i>	Min	Max
Biological words identified	1	5.29	2.20	0	9
	2	6.45	2.44	2	12
Biological levels attempted	1	3.65	1.73	0	6
	2	4.77	1.41	2	8
Levels correctly identified	1	3.00	1.50	1	7
	2	3.68	1.52	0	7

(*N*_{Total} = 39), (*n*_{Intervention1} = 17) and (*n*_{Intervention2} = 22), No statistical significance between interventions

Comparing computational interventions

Interventions resulted in different gains in computational and evolution knowledge (Christensen and Lombardi 2023) therefore we also compared the interventions separately. Computational intervention groups will be referred to as intervention groups 1 (Classes B and D; Table 1) and 2 (Classes A and C; Table 1) for simplicity. There were 17 participants who turned in artifacts within intervention group 1 (*n* = 17) and 22 participants who turned in artifacts within intervention group 2 (*n* = 22). Means, standard deviations, minimum, and maximum identifications for each of these constructs are listed in Table 10 for each intervention group. A Kruskal–Wallace test did not identify any statistical significance between the intervention groups ($p \geq 0.0761$), it was used because the data did not meet the assumptions for normality (Martin et al. 1993).

Scope of biological levels: comparing computational intervention groups There were significant differences between computational intervention groups for the identification of 5 biological levels (L1, L3, L10, L11 and L13) based on a Kruskal–Wallace test. Table 11 depicts the percentage of participants who attempted each level (percentage attempted) and the percentage of participants who accurately represented each level (accurate percentage) with respective standard deviations for each of the intervention groups. Significance was gauged at $p \leq 0.025$ to account for familywise error.

There was a significant difference between intervention groups for L3 with, $H(1) = 34.2$, $p < 0.001$ where participants within intervention group 2 identified L3 more frequently (100%) as compared to intervention group 1 (6%). There was a significant difference between intervention groups for, L10 with, $H(1) = 13.3$, $p < 0.001$ where participants within intervention group 1 identified L10 more frequently (82%) as compared to intervention group 2 (23%). There was a significant difference between intervention groups for, L11 with, $H(1) = 17.9$, $p < 0.001$ where participants within

Table 11 Percents and SD of biological levels comparing interventions

Biological level	Intervention	Attempted %	SD	Accurate %	SD	% Correct
14	1	0	–	0	–	–
	2	0	–	0	–	–
13*	1	53	5.14	41	5.07	78
	2	5	2.13	5	2.13	100
12	1	0	–	0	–	–
	2	18	3.95	9	2.94	50
11*	1	18	3.51	0	–	0
	2	86	5.02	32	4.77	37
10*	1	82	3.93	76	4.37	93
	2	23	4.29	9	2.94	39
9	1	47	5.14	35	4.93	74
	2	41	5.03	36	4.92	88
8	1	0	–	0	–	–
	2	5	2.13	0	–	0
7	1	41	5.07	35	4.93	85
	2	23	4.29	23	4.29	100
6	1	0	–	0	–	–
	2	18	3.95	14	3.51	78
5	1	24	4.37	18	3.93	75
	2	36	4.92	27	4.56	75
4	1	0	–	0	–	0
	2	14	3.51	8	2.70	57
3*	1	6	2.43	0	–	0
	2	100	0	91	2.94	91
2	1	94	2.43	94	2.43	100
	2	86	3.51	86	3.51	100
1*	1	0	–	0	–	–
	2	23	4.29	23	4.29	100

($N_{Total}=39$), ($n_{Intervention1}=17$) and ($n_{Intervention2}=22$)

*Significant at $p \leq 0.025$

intervention group 2 identified L11 more frequently (86%) as compared to participants within intervention group 1 (18%). There was a significant difference between intervention groups for, L13 with, $H(1) = 11.5$, $p < 0.001$ where participants within intervention group 1 identified L13 more frequently (53%) as compared to intervention 2 (5%). Significant differences between intervention groups 1 and 2 were also found at the same biological levels for correct attempts ($p \leq 0.038$). A visual representation of the average percent identified of each biological level (differentiating intervention) is depicted in Fig. 2.

Relational analysis: comparing computational intervention groups We calculated the percent of correctly made BLC attempts between intervention groups. For intervention 1, 41.8% of participants correctly made BLCs and 27.0% of participants correctly made BLCs in intervention 2. A Kruskal–Wallace test was used

to identify statistical significance between the groups because the data did not meet normality assumptions as displayed in Table 12.

The average BLC attempts were significantly different between intervention group with, $H(1) = 10.3$, $p < 0.001$. Participants significantly made more connection attempts in intervention 2 ($M = 3.14$, $SD = 1.08$) as compared to intervention 1 ($M = 1.82$, $SD = 1.07$). The average minimum levels between attempts was significantly different between intervention groups with, $H(1) = 8.69$, $p = 0.003$. Participants within intervention group 1 significantly had more levels between their minimum attempts ($M = 5.24$, $SD = 3.41$) as compared to participants within intervention 2 ($M = 2.09$, $SD = 2.04$). There was no significant difference between the intervention groups for the average maximum biological levels between connections ($p = 0.672$).

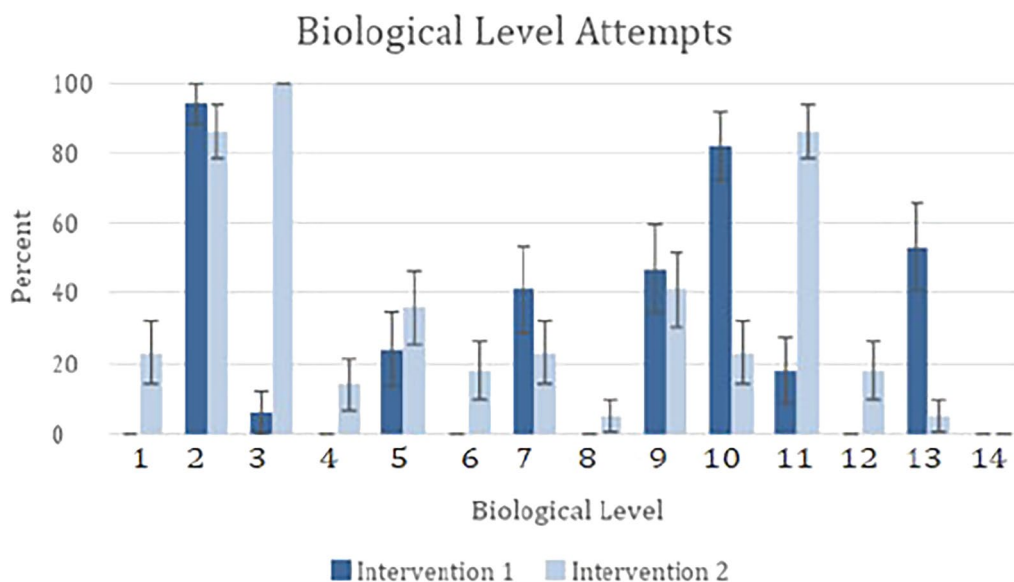


Fig. 2 Biological level attempt percentage between intervention 1 and 2, with bars showing ± 1 standard error. Biological levels represented 1 = atom, 2 = molecule, 3 = macromolecule, 4 = organelle, 5 = cell, 6 = tissue, 7 = organ, 8 = organ system, 9 = organism, 10 = population, 11 = species, 12 = community, 13 = ecosystem, 14 = biosphere

Table 12 Relational analysis comparing intervention groups: summary of total biological connections means and standard deviations

Category	Intervention	M	SD
Connection Attempts*	1	1.82	1.07
	2	3.14	1.08
Levels Between Connections Min*	1	5.24	3.41
	2	2.09	2.04
Levels Between Connections Max	1	7.76	2.88
	2	7.95	1.75
Levels Between Average Attempt*	1	6.69	2.53
	2	4.74	1.51
Correct Connections*	1	1.18	1.01
	2	2.00	0.870
Correct Average Between Connections	1	4.98	3.70
	2	3.21	2.10
Correct Micro Level Connections*	1	0.180	0.393
	2	1.68	0.780
Correct Macro Level Connections	1	0.180	0.529
	2	0.050	0.213
Correct Micro Through Macro Level Connections*	1	0.760	0.664
	2	0.180	0.395

($N_{Total} = 39$), ($n_{Intervention1} = 17$) and ($n_{Intervention2} = 22$)

*Significant at $p \leq 0.025$ to account for familywise error

The average number of individual biological levels between connections was significantly different between intervention groups, with $H(1) = 19.23$, $p = 0.002$.

Participants in intervention 2 had made BLCs which were significantly closer together ($M = 4.74$, $SD = 1.41$) as compared to intervention 1 ($M = 6.69$, $SD = 2.53$). In other words, participants in intervention group 1 had more individual biological levels between their BLC attempts on average. There was a significant difference for the amount of correct connections between interventions with, $H(1) = 6.31$, $p = 0.012$. Participants in intervention group 2 ($M = 2.00$, $SD = 0.870$) had significantly more correct BLCs as compared to participants in intervention group 1 ($M = 1.18$, $SD = 1.01$).

When considering the three types of connections, there was a significant difference for micro level connections (L1–L9) between intervention, with $H(1) = 24.1$, $p = 0.01$. Participants in intervention group 2 made more micro level connections ($M = 1.68$, $SD = 0.780$) as compared to participants within intervention group 1 ($M = 0.18$, $SD = 0.393$). There was also a significant difference in the average number of participants making connections between the micro through macro levels between intervention groups with $H(1) = 24.1$, $p = 0.01$. Participants part of intervention 1 ($M = 0.76$, $SD = 0.664$) significantly made more connections through biological L9 as compared to participants part of intervention group 2 ($M = 0.18$, $SD = 0.395$). There was no significant difference for connections at the macro level ($p = 0.391$) between the interventions groups. Significance was gauged at $p \leq 0.025$ to account for familywise error. All of the BLC means and standard deviations between the intervention groups are listed in Table 11.

Computational complexity (RQ2)

Table 13 displays the means and standard deviations for each of the computational complexity categories summed for all participants (participant scores within each category ranges from 0–4). It also includes the percentage of participants who included models and the level of their model complexity average was between simple and developing ($M=1.23$, $SD=0.842$; 0 for model absent, 1 for simple model present, and 2 for developing model present). On average participants' complexity of the computational components (input, integration, output and feedback) were not broken down further due to limited data (resulting in inconsequential results). Complexity of the separate computational components was outside the scope of this study.

It is important to note there were major differences in the content of these interventions. There was technically no concrete feedback within intervention 2 for participants to identify (although participants were prompted to hypothetically recognize it, none of them did). We ran a Kruskal–Wallis Test and results indicated there was a statistical difference between intervention groups for the computational developing category with $H(1)=4.89$, $p=0.027$. Participants in intervention 2 ($M=1.68$, $SD=1.21$) scored significantly higher in the developing category as compared participants in intervention 1 ($M=0.82$, $SD=1.02$). There was also a significant difference in intervention groups for the model present category with, $H(1)=10.1$, $p<0.001$. The means and standard deviations comparing interventions for computational complexity are also shown in Table 13.

Participant scores for each of the complexity categories ranged from 0 (all components absent) through to 4 for each of the computational components (input,

integration, output and feedback; i.e. participants had four opportunities to score within the simple category [0 to 4 comes from a combination of complexity scores of the computation components; input, integration, output and feedback]). For example, a participant may have scored in the simple category for input, the developing category for integration and output, then failed to mention feedback (Participant score: simple=1, developing=2, complex=0).

Discussion

Conceptual and relational analysis of biological levels (RQ1)

The total biological levels identified by participants ranged from L1 to L13 with a majority of participants identifying levels at the micro (L2 and L3) scales and the macro scales (L10, L11 and L13); however, there was a lower frequency of identified levels in-between (middle scales). Atom, macromolecule and tissue tended to occur together, as did tissue and organ. The items that tended to occur together which were further apart in scale were organ system and community as did molecule and population. Macromolecule tended to not occur with population and ecosystem and population tended to not occur with species and community. These results support the idea that some scale levels tended to occur together more readily than others based on the intervention, even though students were not restricted to specific levels.

It is also important to note levels which participants most frequently identified incorrectly: L8, L11 and L12. For example, 56% of participants identified L11 but less than 40% of participants who recognized it identified it correctly. Participants hypothesized about organisms at L11, because usually organisms here were a component of phylogenetic trees (output, intervention 2), or they were depicted in the hypothetical populations (integration, intervention 1). Participants had difficulty identifying species (all organisms which have the ability to viably reproduce) and identifying that individual organisms were part of populations (organisms within a local area). Increasing knowledge in these areas may be of importance for biology and biological evolution learning. Although individual organisms are important to distinguish within biological systems, participants frequently referred to them incorrectly within their artifacts. Biological levels 1, 2 and 3 had greater than 80% accuracy in correct participant identification. In most cases, based on artifacts participants used resources directly to define these levels (i.e., to search and find the function of a specific protein). The students that had accuracy in their artifacts due to searching, may not have learned the concepts as reflected by the knowledge gains within the previous quantitative study (Christensen and Lombardi

Table 13 Means of computational complexity comparing interventions

Category	M	SD	Intervention	M	SD
Simple	0.920	1.22	1	1	1.41
			2	0.86	1.08
Developing*	1.31	1.20	1	0.82	1.02
			2	1.68	1.21
Complex	0.030	0.160	1	0	-
			2	0.050	0.045
Model Level	1.23	0.842	1	1.47	0.514
			2	1.05	0.999
Model	%	SD	Intervention	%	SD
Model Present*	74.0	44.2	1	100	0
			2	55	5.10

Intervention 1 ($n=17$), Intervention 2 ($n=22$),

*indicates significance at $p<0.05$ between interventions.

2023). It is important to note that although differences between interventions as well as evolutionary knowledge gains are assumed from previous study our results do showcase the difference between intervention 1 and 2. Most of our results are not dependent on if the students actually learned evolution, but focus on the differences between the interventions (biological scales and content) through computational thinking. Student participants also spent the beginning of the year learning about these microscales, which may have contributed to accuracy at these levels. Curricula from previous years emphasizes microscale topics (i.e. chemistry) and the biological scales of organization. Results indicate that additional testing or exploration around these specific biological scales are warranted by both researchers and educators.

Participant identification of Levels 1 and 3 (micro scales) as well as 10, 11 and 13 (macro scales) differed on average based on intervention group. There were significantly more level identifications focused at levels 1, 3 and 11 for intervention group 2 on average, and at levels 10 and 13 for intervention group 1 on average. This supports the idea that intervention 1 focused on macro levels while intervention 2 focused on micro levels, but that these levels were not mutually exclusive. Participants part of intervention 2 frequently identified L11 incorrectly. Although participants in intervention 1 identified levels 10 and 13 more frequently, they did not identify L2 and L3, the smallest scales, as frequently. These differences may be due to the differences in nature of the interventions, objectives and questions asked in the interventions. These levels were the areas participants selected to identify based on the interventions.

Participants in intervention 1 had both more growth in computational thinking as compared to intervention 2 and had significantly more growth in biological evolution knowledge (interpreted from the qualitative analysis; Christensen and Lombardi 2023). The relational analysis in the present study reveals, participants in intervention group 2 (the group with less knowledge gains from previous study) significantly made more connection attempts between biological levels, more correct connections, smaller numbers of biological levels between attempts, and had more micro level connections (between L1-L9). Participants part of intervention group 1 (the group with greater knowledge gains from previous study) made more connections through micro and macro levels (between L9) and larger BLCs on average. Between both intervention groups, the BLCs that were smaller tended to be accurate more often. This presents evidence that making larger connections and connections between the micro and macro levels may be important to biological evolution learning when comparing these biological level differences to quantitative knowledge gains of previous

study. The connections with less levels in between may require mastery before attempting larger levels. Understanding the reason for student accuracy may also shed light, it may have been easier for students to make connections at levels closer because this occurs more often in biology resources and texts (that participants may have used) as compared to levels which are not often connected. The resources participants were required to use were specific, in that they directly searched for the function of their proteins (which in many cases related directly to another microscale level) and they did not need to make any inferences about these relationships.

In biological systems the explanations of mechanisms of phenomena (such as biological evolution) apparent at one scale often lie at a different scale (Parker et al. 2012). Intervention 1 may have better contributed to students' sense-making skills and perceptual fluency (Rau 2018) allowing students to holistically engage with the topic. It may also support the idea that micro scale levels and connections are not necessarily influential on biological evolution learning when considering previous results. For example, a microbiologist studying protein sequencing and seeing the outcomes at the level of the cell or organism only may be less clear on biological evolution as compared to a microbiologist who also understands these implications above the organism level. This is evidenced by many universities that have courses, majors and even departments broken up into the microbiology and or cell biology departments that are distinguishable from their ecology and evolution counterparts.

This problem of student inability to make micro to macro scale connections is sometimes referred to as "slippage between levels" (or disconnects between levels) and is also associated with fragmented and compartmentalized knowledge (Brown and Schwartz 2009). This problem has not received much attention in the literature on evolution education (Jördens et al. 2016). Fluidity between these levels allows students to reason across them and contribute to biological literacy (Brown and Schwartz 2009). Student understanding of these levels and BLCs is important in evolution learning but may also provide educators understanding of where their students struggle with biology concepts. Curricular materials and assessments often focus on concepts through objectives and not explicitly at the levels or connections being pursued.

This finding is particularly interesting because it emphasizes one of the 5 chief strategies that encourage thinking across levels in biology (Parker et al. 2012). Thinking across levels in biology consists of: (1) distinguishing different levels of organization, (2) interrelating concepts at the same level of organization (horizontal coherence) (3) interrelating concepts at different levels

of organization (vertical coherence), (4) thinking back and forth between levels (yo-yo thinking) and (5) meta-reflection about the question which levels have been transcended (Jördens et al. 2016). Our finding presents the importance of point 3 (vertical coherence) and 4 (yo-yo thinking) in biological evolution learning, which may have been supported by computational thinking within intervention 1 based on previous study as explored here. More emergent phenomena occur for students as their BLC distances increase, and these phenomena may have become more apparent for participants in intervention 1. Previous studies have considered the importance of macro and micro level connections (Jördens et al. 2016); however, there is a gap in research explicitly identifying multiple levels within the micro and macro level ranges and connection distances in the ways that we have; which can be applied to other learning scenarios. Computational thinking can emphasize learning across specific scales, including those which tend to be problematic or pose misconceptions (Chi et al. 2012).

We suspected that the interventions may have emphasized different biological levels of organization or prompted different degrees of computational complexity. Our results also support our hypothesis that there would be a discernible emergent relationship between participant understanding of biological levels of organization in response to computational intervention exposure. The difference in knowledge gains and in essence effectiveness (Christensen and Lombardi 2023) may have a relationship with the participants' use of biological levels of organization, biological level connections and or computational complexity as these were the areas which significantly differed between the interventions. It also supports that the group with greater knowledge gains (intervention group 1) more frequently identified scales at the macro level. Participants within intervention group 2 identified more micro scale levels and showed a greater ability to identify them accurately indicating that accuracy within artifacts at the micro levels in this case was not a contributor toward knowledge gains gleaned from previous study.

There were discernible differences between intervention 1 and intervention 2 considering how many BLC attempts were made, the numbers of levels between attempts and the type of attempts that were made. This also supports our hypothesis regarding the relationship between unique participant understanding of biological levels of organization in response to the different computational interventions. These differences may have contributed to the significant knowledge gains present for intervention group 1 (Christensen and Lombardi 2023). Participants part of intervention group 2 (the less effective intervention) made more connection attempts

overall and more connection attempts at the micro scales. Participants in intervention 1 (the effective intervention) made more successful biological connections through level 9 (organism level) and their BLCs significantly spanned larger ranges. Because intervention group 1 made more significant knowledge gains based on previous study, our exploration provides support that connection attempts through level 9, macro level identifications, and larger BLCs on average (encompassing more scales) may have an impact on student learning biological evolution whereas the number of connection attempts for example did not. Larger BLCs might indicate that participants are making relationships more holistically (encompassing both unity and diversity concepts; Jördens et al. 2016). Students making successful connections through level 9 are showing evidence of combining micro level concepts with macro level concepts. These connections may have been better supported through various computational aspects of intervention 1.

Many factors may have contributed to this finding: including use of scales students are comfortable with, scales that were most often prompted by biological evolution or the nature of the interventions themselves. Regardless of the intervention differences, the student's understanding and use of the scales alone may provide insight as to which scales are most difficult for students or require emphasis before or while learning biological evolution. Artifacts were not collected during the traditional lessons, therefore there was no control group and these results are applicable to the two versions of the computational thinking interventions.

Computational complexity analysis (RQ2)

Most participants exhibited simple and developing complexity among the computational components mentioned in their artifacts. It is important to note that participants explicitly wrote their answers identifying the computational components and answers were scored based on accuracy. On average more participants exhibited scores in the developing category ($M=1.31$, $SD=1.20$) as compared to the simple category ($M=0.93$, $SD=1.22$) and the difference was significant between intervention groups. For most of the computational components, a simple score was given if the component was mentioned at all, while a moderate score was given if the component was given alongside a biological context (See Table 5).

When comparing intervention groups 1 and 2, participants in intervention 2 significantly had more participants score in the developing category ($M=1.68$, $SD=1.21$) as compared to the simple category ($M=0.82$, $SD=1.02$). Significantly more participants within intervention group 1 (100%) had their models present within their artifacts as compared to intervention group 2 (55%).

These differences may be attributed to the specific activities and components emphasized within the interventions coupled with participant's ability to learn from them. Quantitative results (Christensen and Lombardi 2023) indicate participants in intervention 1 produced greater knowledge growth in both computational and evolution knowledge constructs, however qualitative results regarding computational complexity in the current study indicate model presence was the only construct that significantly measured higher for this group, indicating that attempts at computation contribute to understanding. It was interesting to note that overall computational knowledge growth was not considerable (only at time 2), and this could be because it was the participants first time interacting with computational constructs. There was a shared aspect of the computational lessons and as well as unique computational aspects between the interventions. Both intervention groups were exposed to computational learning, and proof of computation as displayed significantly more often for intervention group 1, may be responsible for the knowledge gains of the group. The computational context was also not considered here, as the computational context provided for intervention 1 was not a preconstructed tool with an interface (as this was the case in intervention 2). To describe an analogy, a preconstructed tool in order to 'save your work' may be clicking the floppy disk icon (and potentially being redirected) when typing up a document, while an alternative more complex way to complete this task (requiring higher computational complexity) would be selecting save as, and selecting where and how to save the file.

There were discernible differences between the interventions when considering computational complexity within participant artifacts (computational products). Interestingly, significantly more participants in intervention 2 (less successful intervention) scored in the developing category (as compared to intervention group 1). Significantly more participants turned in representations of their computational models in intervention group 1, which may indicate that computational interactions themselves (and the specific computational context) may contribute to overall gains as compared to the computational complexity alone, supporting the idea that computational thinking exposure and type is beneficial within itself.

Based on LBECT-LP (Christensen and Lombardi 2020) the computational products (the artifacts in this case) are a representation of the combination of the instructional context and computational process experienced by participants. We suspected that increased computational complexity (exhibited by artifacts of intervention group 2) would have a relationship with computational

knowledge gains or evolution knowledge gains, however this is not the case. The results of the study did not support this relationship between increased complexity and knowledge gains. It is important to note that participant artifacts (computational product) were assessed to discern computational complexity. The LP indicates that computational product, computational process and instructional context should all be considered when assessing complexity (Berland and McNeill 2010). For the purposes of the study we made the assumption that computational products (artifacts) would accurately reflect a combination of complexity of the computational process and instructional context.

When considering the instructional context, the first computational intervention itself had more computational complexity as compared to the second intervention (google sheets, vs. BLAST program) however the integration within intervention 2 was essentially invisible to the student participants due to a more complicated interface (i.e. save icon versus file save setup). A follow up to the study might compare a group of student participants who had not been exposed to computational learning interventions at all or develop artifact questions to provide pre-intervention to use as a baseline comparison to discern student computational complexity, though it is reasonable to hypothesize that students had very little understanding of computation to start. The LBECT-LP is designed to expose students to computation and evolution over a much longer period of time than the study allowed and the computational context as it related to complexity was not considered. In future studies, the instructional context, computational process and computational product may need to be considered simultaneously to truly assess success (and complexity) of computational instruction. These findings partially support the hypothesis that there would be a discernible relationship between student level of complexity in terms of understanding of computational thinking (simple, developing or complex) in response to the different computational interventions, although the complexity differences did not align with our predictions based on differences in content knowledge.

Limitations

One limitation relates to the degree to which the findings are generalizable based on the subset of participants. The study sample is representative of advanced biology students in a predominantly white middle class population. Participants were randomly assigned at the class level and the number within each class differed although results did not indicate significant classroom or teacher effects allowing us to discern randomness. There was no true control group of participants (who did not experience

computation in order to compare their tests or artifacts with) which might be a consideration for future studies. Artifacts were collected after the computational interventions, and not the traditional lessons due to inequality between the artifacts and attempts to reduce the burden on the teachers to try and equate the assignments. The LBECT-LP provides some structure for use with additional age, content and ability levels. The first intervention may have targeted unique aspects of evolution or computational thinking as compared to intervention 2. Addressing and assessing the instructional context and computational process along with the computational product (artifact) would provide more insight on computational complexity. These items may include consideration of the lesson structure itself and or social interactions during lessons. Certain students are more versed in coding than others and novelty of the researcher teacher may have influenced the results. The interventions prompted student use of interface friendly computational tools, partially due to the fact that the learning progression is not used in the district, prompting students to start at the lowest levels. Computation at a higher complexity would have required additional learning time for students. Ideally, in a more computationally complex version of this lesson (when presented at an appropriate ability level), students could write simple programs and be provided less scaffolding. Researchers versed in biology education and computational thinking assisted educators in this study. The researcher teacher also designed the study to reduce burden on the teachers, provide relatively equal opportunities to all participants and maintain the integrity of the AP curriculum. Extending the research or teaching implications may require less intricate computation, objectives that target specific levels or computation or pre-made lessons (with infused computational thinking) to support educators.

It is also important to note the open-ended nature of the study, as students could use an infinite number of biological words to discern a biological level. Although the interventions were both teaching content related to biological evolution, there were differences between the resulting artifacts. This novel type of relational coding analysis may be beneficial to complete at the conclusion of other units in order to accurately assess student knowledge of biological evolution, and may be applied to other biological education studies. Student computational thinking may have encouraged the use and flow of multiple biological levels of organization; as this was the tool used for both interventions. The design of the study was based on infusing as much testing and artifact collection as possible while minimizing stress on teachers and maintaining integrity of the curriculum that was already in place. Although computational thinking constructs are used within the curriculum, the use of the computation in the interventions was at a higher level

due to the exclusiveness, the definitions presented by the LBECT-LP and explicit tool use. The sample size was small, and computational tools were limited (and simple) particularly due to lack of prior computational knowledge. The use of free online resources (such as Google sheets and BLAST) are available and accessible to teachers and districts, but often require training and exposure in order for educators to feel comfortable with them. This study assumes that evolution knowledge gains were made in a previous study. Our purpose was to ask additional research questions that would provide a deeper dive into the previously published quantitative results, as this may be a limitation but also an item to consider for repeatability.

Implications for teaching and research

The implications for teaching and research can be both applied and theoretical. In a practical sense the complexity of computational thinking presented and expected of students was relatively simple in this study. Although it is feasible for AP biology students to write their own (very simple) code in various programs (i.e. R and Python) in a more computationally complex task, this is more likely achieved with high levels of scaffolding that would be difficult to impose on teachers. This may be taxing for some schools, as most integration of computational thinking tends to be, which lends towards questions of equity, access, teacher preparation and scaffolding through grade levels (as addressed in the LP). Multiple simple computational tasks scattered throughout biological learning during a school year may have positive impacts on student learning, particularly for topics that span biological levels such as biological evolution. In a theoretical sense, we make assumptions about the relationship between computational thinking and biological evolution learning as supported by the results of Christensen and Lombardi (2023). The current study supported the idea that multiple biological levels and BLCs emerged through the use of computational thinking differently between the two interventions. Our two computational interventions differed in learning gains (as was the focus on in previous studies), either micro (unity) or macro (diversity) scales and again lent to identification of differences biological levels addressed by students, biological level connections made, and size and types of the BLCs. Student development of models without an interface (intervention 1) was associated with knowledge gains and may provide insight on the fact that computation itself may provide benefit, even if it is simple as compared to not engaging with computation at all. Understanding student identification of biological levels and relationship building among them is a novel yet important idea for both researchers and teachers which can be applied to other methods of learning.

AP biology students are unique in that they tend to be high achieving students at the high school level who are

expected to learn (and are tested on) a standardized college level biology curriculum. Student navigation and interaction with a spreadsheet (no interface) versus a user-friendly website (with interface) may have contributed to the success of intervention 1 and the differences in the use of biological levels and biological level connections. Computational complexity may have been achieved more often by students in intervention 2 because they were able to identify and use biological words to describe their computation and the integration portion was essentially written for them within the interface. Students in intervention 2 did not start with a modeled template which may be a reason for the absence of the student models in intervention 2. Interestingly, an accurate description of more complex computation was not as valuable as the evidence of the models itself (even if it was inaccurate and or less complex). Lending to considering additional factors (such as computational context) when considering artifacts and or testing around biological evolution or computational thinking for both research and teaching purposes.

The interventions were imperfect in providing this scaffolding both from a biological evolution and computational thinking perspective. We claim that computational learning is beneficial for biological evolution learning due to its inherent ability to span levels and facilitate emergence. We identified which and if these levels were present as well as the student connections that differed based on computational intervention. Often in certain types of qualitative studies, ideas and results emerge from the data. The *best* practices for helping students to think computationally are still unknown. This becomes especially important when considering differentiation to account for various ability levels found within classrooms. Such research would include testing other lessons (i.e., interventions), instructional units, and modified (or new) assessments based on the LBECT-LP. Researchers may better understand impacts on classroom practice, specific NOS processes, and overall effectiveness. The results from new interventions or assessments may be compared to groups of students who have not received computational interventions to better identify if and how computational thinking contributes to learning specific constructs around evolution. Additional groups of students (ages, demographics or ability levels) should also be considered.

Further exploration of the specific biological levels or computational components is necessary, especially since there were distinct differences between the interventions within the study. For example, specific participant answers of the qualitative study may be paired with specific quantitative biological levels, computational complexities, computational components, and or NGSS standards, not just evolution or computational knowledge constructs as a whole. The evolution knowledge assessment was

developed specifically to assess all levels and the computational thinking assessment was designed specifically with each question tailored to input, integration, output and feedback. Pairing these specific questions with biological levels, connections and complexity would shed light on more specific benefits of computational interventions around biological evolution. Our results indicated that biological scales (especially micro through macro scale connections, or identification of larger levels) may be important for biological evolution learning because of the significant difference in evolution knowledge and biological organizational level scale use between the interventions associated with greater knowledge gain from previous studies. Further exploration of specific biological levels (and BLCs) and their interaction (and emphasis) with computation (among other methods) lends for future research. Analogous to our assessment of biological levels, another scale to consider may be time (just as evolution occurs across biological scales, it also occurs over time). Improved assessments or additional interventions may be used to further explore the relationship between biological evolution learning and computation.

Conclusion

Our study serves as a deep dive into exploring knowledge gains from the use of CT based on the idea that CT encourages students to interact with various levels of biological organization. The interventions both involved teaching biological evolution through computational means; however, they involved distinctly different activities that may have emphasized different biological aspects of biological knowledge and complexity of computational processes. It is reasonable to hypothesize that these levels, connections and computational complexity differences may have some relationship with evolution or computational knowledge based on previous quantitative results (Christensen and Lombardi 2023). The purpose of this study was to explore the differences in knowledge gains (between the interventions) by quantifying our qualitative analysis through participant artifacts around (1) biological levels of organizations and biological level connections as well as (2) computational complexity. Participants identified a range of biological levels with the most frequency occurring at lower biological scales (L2 and L3) and higher scales (L10, L11 and L13) with less incidence of scales in-between as prompted by both computational based interventions. When considering the 5 chief strategies that encourage thinking across levels in biology (Parker et al. 2012) this presents a deficiency of the first strategy, identifying the levels of organization. Sometimes micro and macro scales are presented alongside certain concepts and methods within biology classrooms, however some scales (the scales in-between) may not be presented as often

with evolution. Certain levels were easier for participants to identify correctly (L1 and L2) whereas other levels (L11, L10 and L8) presented more difficulty. These findings present that certain levels and connections may need more assessment and attention within biology classrooms in learning biological evolution. Participants in intervention 1 identified L11 and 13 more frequently whereas participants in intervention 2 more frequently identified L2, L3 and L10. These differences were likely due to the nature of the content within the interventions.

The BLC count also differed significantly between the interventions. Although biological evolution growth for participants in intervention 2 was not significant, on average these participants made more connection attempts, had smaller numbers of biological levels between attempts, and had more micro level attempts (between L1-L9). Participants a part of intervention 1 made more connections through micro and macro levels (between L9) which may pose greater implications for biological evolution knowledge growth as indicated from the quantitative analysis. Facilitation of learning around these specific biological level connections can be achieved through computation and particularly what items are most meaningful.

On average more participants generated developing level computation as compared to simple computation, however this was only based on the computational product and did not include scoring of the computational context or computational process. When comparing interventions, significantly more participants part of intervention 2 scored in the developing category, yet less participants submitted their computational model which may have indicated a higher difficulty or complexity. For most participants it was the first time that participants had interacted with computational components, especially in the way they were defined in the LP. Developing strategies to get students towards complex computation, especially coupled with specific biological evolution content objectives is an important topic for further research.

Overall, the results support the use of the LBECT-LP because it may assist in student exploration of different biological scales as a contemporary method to get students thinking about the connections between biological levels of organization in learning biological evolution. More importantly this type of thinking supported by our framework can span across other topics and disciplines. It is unknown how computational processes will help students to better understand evolution and in turn how this might strengthen their knowledge within the domain of biology as a whole. Using computation should strengthen student knowledge and NOS processes, but it is unclear in what ways. Continued quantitative analysis may identify if knowledge

gains are being made, while supplementary qualitative analysis can provide domain specific insight to direct future research in these areas. Students may become more comfortable developing new computational tools or applying these skills to other disciplines. It is also unclear how computational processes explored through this type of learning progression may relate to overall student achievement, collaborative learning outcomes, or its application to other scientific disciplines.

Proper scaffolding would allow students to make connections across various levels of biological organization, our research may aid in resolution of scaffolding for specific students. With the assistance of the appropriate experts, we suspect this learning progression can be modified and applied to other domain areas. These types of scales exist across disciplines to emphasize biological evolution. It is known that learning across scales is central in science, technology, engineering and mathematical learning. Relational reasoning (basic cognitive mechanism involved in the formation of conceptual categories and encompasses the ability to detect similarities and differences in patterns among objects, concepts and situations) has been presented to alleviate misunderstandings across biological scales however very few specific learning activities have been developed on the topic, and none associated with biological evolution knowledge (Resnick et al. 2017). The act of computational thinking may encourage relational reasoning and activate unique cognitive processes that may be beneficial to learning biological evolution. These ideas expand to include subfields of the major sciences such as biochemistry, physiology, marine science, or environmental science (Christensen 2022). Therefore, developing learning progressions incorporating computational thinking with fundamental concepts within other scientific fields (in addition to biology) should be considered for future study to expand evolution knowledge and application. Additional lessons which infuse biological evolution and computation may be an avenue to further explore in a practical sense (Gallagher et al. 2011). Computational thinking continues to expand within science curricula, and educators need assistance in developing feasible lessons for their students to more effectively blend computational thinking with concepts such as biological evolution.

Future study may also lend to the discussion of *why* computational thinking might be effective for learning across scales (i.e. biological levels). These ideas relate to the concept that is presented in many biology texts to describe emergent properties and interestingly also applied to many other domains: The whole is more than the sum of its parts (Campbell et al. 2000) as can be modeled through computational thinking. Computational thinking may provide students the tools and thought processes which allow them to account for all of the moving parts (i.e. biological levels)

alongside the big picture (i.e. biological evolution across scales). Understanding of these mechanisms and benefits of these thought processes are essential for advancement of biology within a technological world.

Appendix A
 Example of coding analysis
 See Tables 14 and 15.

Table 14 Biological level coding from artifact below

		Biological words student uses To identify biological level(s) (instances)	Which biological levels (1-14)	How many properly identified level(s), supporting explanation present	How many connection attempts made between levels?	How many levels between these attempts?	Was correct connection made between levels? (at least once)	Other Notes
14	Biosphere							
13	Ecosystem	Environment	13	1, no		12	no	
12	Community							
11	Species							
10	Populations	Populations	10	1, yes		9	yes	
9	Organism	Bird, offspring (2)	9	1, no		8	yes	
8	Organ system							
7	Organ	Beak	7	1, no		6	yes	
6	Tissue							
5	Cells	Gametes, Zygotes (2)				4	no	
4	Organelle							
3	Macromolecule							
2	Molecules	Allele, Allele A, Allele B (3)	2	3, yes				
1	Atoms							
	totals	5 levels used	2-13	5 total, 2 explanations	2-5, 2-10, 2-7, 2-9, 2-13 (5)	4,6, 8, 9, 12	2-5 no, 2-10 yes, 2-7 yes, 2-9 yes, 2-13 no	

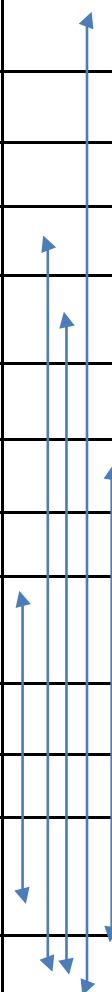


Table 15 Computational complexity coding from artifact below

	Key Biological Words Mentioned with term	Simple	Developing	Complex	Notes
Input	Allele		x		
Output	Allele	x			
Integration	–	x			
Feedback	allele		x		
Model/Output Display	Present? Yes		x		
Description of process		Vague or absent	No mathematical construct, incorrect use description of model	Includes proper mathematical construct, provides proper other uses of model/concepts, brings up new ideas	

Participant Artifact Coding

Student typed in red.

Identified areas to be analyzed and quantified in table above

Computational components

Biological Levels

Connections

overlap

Participant Artifact

*Student typed in red

H-W Activity

What do you need to turn in?

(Each section only needs to be a few sentences).

1. Document that includes:
 - a. Part 1
 - i. General procedure and procedural information (A summary, be sure to define what you did, computational components, biological aspects)
 - The general procedure is to open up the excel sheet and put in the input value for the frequency of allele A. The result of changing this value is a cascade effect in which the value for the frequency for allele B is changed and the random gametes and zygotes for the chart is produced.
 - ii. Results (Be sure to identify what you observed in your populations and how it related to the H-W equation and evolution, be sure to include the micro through the macro scales and include time)
 - Within the populations, the larger the value for the frequency of allele A becomes, the smaller the frequency for allele B becomes. This also affects the chart of random zygotes because, for example, if the value of allele A's frequency is 0, then there will only be B's present for offspring. This relates to the H-W equation and evolution because what's produced relates to the p and q above the equation. (Allele A is p and Allele B is q).
 - iii. Conclusion (Overall what did you find, how does this display evolution, what computational processes did you use, what is the input, integration, output and feedback for this model)
 - Overall, I found that this displays evolution because the less of allele A present (the frequency), the more of allele B is present. In the environment, that would mean that the trait for allele A would die out eventually. I plugged in 0.6 for the frequency which resulted in more A alleles present for the chart or integration section. The output is considered the 0.4 frequency for allele B and the feedback is the p and q values towards the bottom of the section (which relate to the allele frequencies).
 - b. Part 2
 - i. General procedure (what did you do to your model in part 1, provide the scenario)
 - I plugged in 0.6 for the frequency of allele A which represented the long beak (and medium beak) trait. This resulted in a 0.4 frequency for allele B.
 - ii. Results (How did it change your model)
 - This changed my model by having for A alleles present in the chart and making the graph make this representation too.
 - iii. Conclusion (Overall what did you find from manipulating your population, how does this relate to evolution, what computational processes did you use, what is the input, integration, output and feedback for this model)
 - I found that by manipulating the population, there was a cascading effect in how the other allele was presented and expressed. This relates to population because if i were to make allele A frequency equal 0, then there would be no birds with long beaks left.

*Your document should include screen shots

 2. Google or Excel spreadsheet
 - a. Attach the computational model that you used during the activity

Acknowledgements

Acknowledgement goes to the participating school district and associated teachers and administrators for making this research possible.

Author contributions

Dana Christensen, Doug Lombardi contributed to the development of this manuscript. Dana Christensen, Doug Lombardi wrote the manuscript. Dana Christensen developed the tables and figures. Dana Christensen, Doug Lombardi reviewed the manuscript.

Funding

There were no funding sources.

Availability of data and materials

Availability of data and materials are available through contact with the corresponding author.

Declarations**Ethics approval and consent to participate**

Ethics approval and consent to participate was found under Temple University's IRB 26250. All procedures outlined within this protocol were followed including consent and ascent forms provided to all participants. All human participants part of the study provided written consent.

Consent for publication

We consent this document for publication.

Competing interests

There are no competing interests or external funding sources.

Received: 17 September 2023 Accepted: 1 June 2024

Published online: 24 June 2024

References

- Angeli C, Giannakos M. Computational thinking education: Issues and challenges. *Comput Hum Behav.* 2020;105: 106185.
- Arastoopour Irgens G, Dabolkar S, Bain C, Woods P, Hall K, Swanson H et al. Modeling and measuring students' computational thinking practices in science. *J Sci Educ Technol.* 2019.
- Banzhaf W, Beslon G, Christensen S, Foster JA, Képès F, Lefort V, Ramsden JJ. From artificial evolution to computational evolution: a research agenda. *Nat Rev Genet.* 2006;7(9):729–35.
- Basso D, Fronza I, Colombi A, Pahl C. Improving assessment of computational thinking through a comprehensive framework. In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research*, 2018. pp. 1–5.
- Berland LK, McNeill KL. A learning progression for scientific argumentation: understanding student work and designing supportive instructional contexts. *Sci Educ.* 2010;94(5):765–93. <https://doi.org/10.1002/sce.20402>.
- Berland M, Wilensky U. Comparing virtual and physical robotics environments for supporting complex systems and computational thinking. *J Sci Educ Technol.* 2015;24(5):628–47.
- Borrego M, Douglas EP, Amelink CT. Quantitative, qualitative, and mixed research methods in engineering education. *J Eng Educ.* 2009;98(1):53–66.
- Brown MH, Schwartz RS. Connecting photosynthesis and cellular respiration: pre-service teachers' conceptions. *J Res Sci Teach.* 2009;46(7):791–812. <https://doi.org/10.1002/tea.20287>.
- Brumby M. Problems in learning the concept of natural selection. *J Biol Educ.* 1979;13(2):119–22. <https://doi.org/10.1080/00219266.1979.9654240>.
- Bryman A. Quantitative and qualitative research: further reflections on their integration. In *Mixing methods: qualitative and quantitative research*. Routledge; 2017, pp. 57–78
- Burmeister AR, Smith JJ. Evolution across the curriculum: microbiology. *J Microbiol Biol Educ.* 2016;17(2):252–60.
- Campbell NA, Mitchell LG, Reece JB, Taylor MR. *Biology: concepts and connections*. Menlo Park, California: Benjamin Cummings; 2000.
- Chen Y, Argentinis JE, Weber G. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther.* 2016;38(4):688–701.
- Chi MT, Roscoe RD, Slotta JD, Roy M, Chase CC. Misconceived causal explanations for emergent processes. *Cogn Sci.* 2012;36(1):1–61.
- Christensen DM, Lombardi D. Biological evolution learning and computational thinking: enhancing understanding through integration of disciplinary core knowledge and scientific practice. *Int J Sci Educ.* 2023. <https://doi.org/10.1080/09500693.2022.2160221>.
- Christensen DM. Computational thinking to learn environmental sustainability: a learning progression. *J Sci Educ Technol.* 2022. <https://doi.org/10.1007/s10956-022-10004-1c>.
- Christensen DM, Lombardi D. Understanding biological evolution through computational thinking: a learning progression. *Sci Educ.* 2020. <https://doi.org/10.1007/s11191-020-00141-7>.
- College Board. AP[®] Biology lab manual for teachers, New York, NY: The College Board. 2019.
- Cook TD, Campbell DT. *Quasi-experimentation: design and analysis issues in field settings*. Boston, MA: Houghton Mifflin; 1979.
- Dauer JT, Momsen JL, Speth EB, Makohon-Moore SC, Long TM. Analyzing change in students' gene-to-evolution models in college-level introductory biology. *J Res Sci Teach.* 2013;50(6):639–59.
- Eyisi D. The usefulness of qualitative and quantitative approaches and methods in researching problem-solving ability in science education curriculum. *J Educ Pract.* 2016;7(15):91–100.
- Bonabeau E. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci.* 2002;99 Suppl 3:7280–7287. <https://doi.org/10.1073/pnas.082080899>
- Fields C, Levin M. Scale-free biology: integrating evolutionary and developmental thinking. *BioEssays.* 2020;42(8):1900228.
- Gallagher SR, Coon W, Donley K, Scott A, Goldberg DS. A first attempt to bring computational biology into advanced high school biology classrooms. *PLoS Comput Biol.* 2011;7(10): e1002244.
- Gašperov B, Đurasević M, Jakobovic D. Leveraging more of biology in evolutionary reinforcement learning. In *International conference on the applications of evolutionary computation (Part of EvoStar)* (pp. 91–114). Cham: Springer Nature Switzerland.
- Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. The NCBI BioSystems database. *Nucleic Acids Res.* 2019;38(1):D492–6. <https://doi.org/10.1093/nar/gkp858>.
- Göransson A, Orraryd D, Fiedler D, Tibell LA. Conceptual characterization of threshold concepts in student explanations of evolution by natural selection and effects of item context. *CBE—Life Sci Educ.* 2020;19(1): ar1.
- Guo Y, Wagh A, Brady C, Levy ST, Horn MS, Wilensky U. Frogs to think with: improving students' computational thinking and understanding of evolution in a code-first learning environment. In *Proceedings of the 15th International Conference on Interaction Design and Children*; 2016. pp. 246–254.
- Heddy BC, Sinatra GM. Transforming misconceptions: using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. *Sci Educ.* 2013;97(5):723–44.
- Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res.* 2005;15(9):1277–88.
- Hsu YC, Irie NR, Ching YH. Computational thinking educational policy initiatives (CTEPI) across the globe. *TechTrends.* 2019;63(3):260–70.
- Hull DL. Are species really individuals? *Syst Zool.* 1976;25(2):174–91.
- Jördens J, Asshoff R, Kullmann H, Hammann M. Providing vertical coherence in explanations and promoting reasoning across levels of biological organization when teaching evolution. *Int J Sci Educ.* 2016;38(6):960–92.
- Knippels MCPJ. Coping with the abstract and complex nature of genetics in biology education: the yo-yo learning and teaching strategy. 2002.
- Legendre P. (Ed.). (Vols. 1–0). SAGE Publications, Inc., 2010. <https://doi.org/10.4135/9781412961288>.
- Martin P, Bateson PPG, Bateson P. *Measuring behaviour: an introductory guide*. Cambridge: Cambridge University Press; 1993.
- Mayr E. The biological species concept. *Species concepts and phylogenetic theory: a debate.* 2000; 17–29.

- McEntire KD, Gage M, Gawne R, Hadfield MG, Hulshof C, Johnson MA, et al. Understanding drivers of variation and predicting variability across levels of biological organization. *Integr Compar Biol*. 2021;61(6):2119–31.
- Miikkulainen R, Forrest S. A biological perspective on evolutionary computation. *Nat Mach Intell*. 2021;3(1):9–15.
- Moreno-León J, Román-González M, Hartevelde C, Robles G. On the automatic assessment of computational thinking skills: a comparison with human experts. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2017; pp. 2788–2795.
- Mueller J, Beckett D, Hennessey E, Shodiev H. Assessing computational thinking across the curriculum. In *Emerging research, practice, and policy on computational thinking*. Springer, Cham; 2017. pp. 251–267
- Nardelli E. Do we really need computational thinking? *Commun ACM*. 2019;62(2):32–5.
- Nehm RH. Biology education research: building integrative frameworks for teaching and learning about living systems. *Discip Interdiscip Sci Educ Res*. 2019;1(1):1–18.
- Nesimyan-Agadi D, Assaraf BZ, O. How can learners explain phenomena in ecology using evolutionary evidence from informal learning environments as resources? *J Biol Educ*. 2023;57(1):170–83.
- Novick LR, Schreiber EG, Catley KM. Deconstructing evolution education: the relationship between micro-and macroevolution. *J Res Sci Teach*. 2014;51(6):759–88.
- Parker JM, Anderson CW, Heidemann M, Merritt J, Merritt B, Richmond G, Urban-Lurain M. Exploring undergraduates' understanding of photosynthesis using diagnostic question clusters. *CBE—Life Sci Educ*. 2012;11(1):47–57.
- Penner DE. Explaining systems: investigating middle school students' understanding of emergent phenomena. *J Res Sci Teach*. 2000;37(8):784–806. [https://doi.org/10.1002/1098-2736\(200010\)37:8%3c784::AID-TEA3%3e3.0.CO;2-E](https://doi.org/10.1002/1098-2736(200010)37:8%3c784::AID-TEA3%3e3.0.CO;2-E).
- Perez KE, Hiatt A, Davis GK, Trujillo C, French DP, Terry M, Price RM. The EvoDevoCI: a concept inventory for gauging students' understanding of evolutionary developmental biology. *CBE—Life Sci Educ*. 2013;12(4):665–75.
- Rahayu T, Osman K. Knowledge level and self-confidence on the computational thinking skills among science teacher candidates. *Jurna Ilm Pendidik Fis Al-Biruni*. 2019;8(1):117–26.
- Rau MA. Making connections among multiple visual representations: how do sense-making skills and perceptual fluency relate to learning of chemistry knowledge? *Instr Sci*. 2018;46(2):209–43.
- Reece JB, Urry LA, Cain ML, Wasserman SA, Minorsky PV, Jackson RB. *Campbell biology*, vol. 9. Boston: Pearson; 2014.
- Resnick I, Davatzes A, Newcombe NS, Shipley TF. Using relational reasoning to learn about scientific phenomena at unfamiliar scales. *Educ Psychol Rev*. 2017;29(1):11–25.
- Robinson OC. Relational analysis: an add-on technique for aiding data integration in qualitative research. *Qual Res Psychol*. 2011;8(2):197–209.
- Rubinstein A, Chor B. Computational thinking in life science education. *PLoS Comput Biol*. 2014;10(11): e1003897.
- Schneeweiß N, Gropengießer H. Organising levels of organisation for biology education: a systematic review of literature. *Educ Sci*. 2019;9(3):207. <https://doi.org/10.3390/educsci9030207>.
- Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg*. 2018;126(5):1763–8.
- Schulze S. Views on the combination of quantitative and qualitative research approaches. *Progressio*. 2003;25(2):8–20.
- Selby C, Woollard J. Computational thinking: the developing definition. 2013.
- Sengupta P, Kinnebrew JS, Basu S, Biswas G, Clark D. Integrating computational thinking with K-12 science education using agent-based computation: a theoretical framework. *Educ Inf Technol*. 2013;18(2):351–80.
- Shute VJ, Sun C, Asbell-Clarke J. Demystifying computational thinking. *Educ Res Rev*. 2017;22:142–58.
- Sinatra GM, Heddy BC, Lombardi D. The challenges of defining and measuring student engagement in science. *Educ Psychol*. 2015;50(1):1–13.
- Sinatra GM, Southerland SA, McConaughy F, Demastes JW. Intentions and beliefs in students' understanding and acceptance of biological evolution. *J Res Sci Teach*. 2003;40(5):510–28.
- Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Pract Assess Res Eval*. 2004;9(1):4.
- Stanley SD, Robertson WB. Qualitative research in science education: A literature review of current publications. *Eur J Sci and Math Educ*. 2024;12(1):175–199. <https://doi.org/10.30935/scimath/14293>
- Tavassoly I, Goldfarb J, Iyengar R. Systems biology primer: the basic methods and approaches. *Essays Biochem*. 2018;62(4):487–500.
- Tibell LA, Harms U. Biological principles and threshold concepts for understanding natural selection. *Sci Educ*. 2017;26(7):953–73.
- US News and World Report. Jackson Township School District. 2017. <https://www.usnews.com/education/best-high-schools/new-jersey/districts/jackson-township-school-district-100094>. Accessed 13 Sep 2019.
- Vanchurin V, Wolf YI, Katsnelson MI, Koonin EV. Toward a theory of evolution as multilevel learning. *Proc Natl Acad Sci*. 2022;119(6): e2120037119.
- Wang C, Shen J, Chao J. Integrating computational thinking in stem education: a literature review. *Int J Sci Math Educ*. 2021;1–24.
- Weintrop D, Beheshti E, Horn M, Orton K, Jona K, Trouille L, Wilensky U. Defining computational thinking for mathematics and science classrooms. *J Sci Educ Technol*. 2016;25(1):127–47.
- Wilensky U, Reisman K. ConnectedScience: learning biology through constructing and testing computational theories—an embodied modeling approach. *InterJournal Complex Syst*. 1999;234:1–12.
- Wilensky U, Reisman K. Thinking like a wolf, a sheep, or a firefly: learning biology through constructing and testing computational theories—an embodied modeling approach. *Cogn Instr*. 2006;24(2):171–209.
- Yates TB, Marek EA. Teachers teaching misconceptions: a study of factors contributing to high school biology students' acquisition of biological evolution-related misconceptions. *Evol Educat Outreach*. 2014;7:1–18.

Publisher's Note

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