Investigating the Efficacy of an Ontological Framework for Teaching Natural Selection Using Agent-Based Simulations

Man Su, Arizona State University, mansu@asu.edu
Michelene T. H. Chi, Arizona State University, mtchi@asu.edu
Jesse Ha, University of California Los Angeles, jesseha@ucla.edu
Yue Xin, University of Maryland College Park, yxin21@umd.edu

Abstract: Integrating agent-based models (ABMs) has been a popular approach for teaching emergent science concepts. However, students continue to find it difficult to explain the emergent process of natural selection. In this study, we employ an ontological framework–the Pattern, Agents, Interactions, Relations, and Causality (PAIR-C)–to guide the design of the ABM simulation module. This study examines the effects of the PAIR-C ABM module versus the Regular ABM module on fostering students' understanding of natural selection. Drawing on pre-posttest data, we found that students in the Intervention group had a better causal understanding when explaining natural selection than the Control group. This paper sheds light on applying an innovative framework to designing effective agent-based simulation modules to teach emergent science concepts.

Introduction

The learning of natural selection has been a challenging concept for learners to grasp. Not only naïve learners, but even advanced learners at the postsecondary level also often hold robust misconceptions in explaining the process of natural selection (Bishop & Anderson, 1990; Gregory, 2009). To address this challenge, many researchers and practitioners have used agent-based models (ABMs) or simulations to teach evolution and natural selection over the past decade (Dickes & Sengupta, 2013; Wagh & Wilensky, 2018).

ABMs are computer simulations used to study individual agents' interactions and how they give rise to unpredictable aggregate patterns (Wilensky & Rand, 2015). A substantial number of studies have explored various approaches toward integrating ABMs into learning evolutionary concepts and natural selection. For instance, researchers have taken a hybrid approach of using multiple external representations (aka. MERs-complemented approach) to complement ABMs in explaining emergent phenomena (Basu et al., 2015; Chi et al., 2012b). Although students' conceptual understanding of natural selection can be positively improved by participating in the MERs-complemented approach, the degree to which learning is improved remains limited. Students continue to have inadequate understanding and robust misconceptions, especially when explaining the causal mechanisms for the process of natural selection (Chi et al., 2012a; Su et al., 2021; Peel, 2019).

Researchers (Chi et al., 2012a; 2012b) thereby propose that developing a correct understanding of natural selection (e.g., how an outcome, such as "darker moths" or "long-neck giraffes", becomes common) requires being able to explain how and why agent-level behaviors could give rise to pattern-level outcomes, referred to as the inter-level causal relationships. Moreover, it is important to distinguish inter-level causal relationships from agent-aggregate complementary relationships or input-consequence relationships. Understanding the agent-aggregate complementary and the input-consequence relationships can be less challenging because neither of them considers all the interactions between agents nor emphasizes the nonlinear, dynamic causal relationship between agent-level interactions and pattern-level outcomes. This paper posits that using ABMs without explaining inter-level causal relationships is not likely to result in a deeper understanding of the emergent process of natural selection.

Building on early works, authors recently proposed an ontological framework - the Pattern, Agents, Interactions, Relations, and Causality (PAIR-C) framework - to explain the root causes for misconceptions and support instructions on inter-level causal relationships (under review). Therefore, this paper aims to show the efficacy of integrating the PAIR-C framework into teaching the emergent process of natural selection using the MERs-complemented ABM approach. Specifically, it compares the effect of the PAIR-C ABM module with a Regular ABM module on facilitating students' deeper understanding of inter-level causal relationships when explaining the emergent process of natural selection. The paper focuses on addressing one major research question: What are the effects of the PAIR-C ABM module versus the Regular ABM module on fostering students' understanding of natural selection?

Theoretical framework

The PAIR-c framework

The PAIR-C framework identifies five dimensions to describe a science process: **P**attern, **A**gents, **I**nteractions, **R**elations among the interactions, and the **C**ausal relationship between the agents and the pattern (Chi et al., under review). A Pattern describes the overall changes by a process that is often visible and meaningful, Agents are elements that participate in the process which produces the pattern, Interactions refer to how the agents of the process interact, Relations compare some agents' interactions with other agents' interactions, Causality refers to the causal relationship between the agents and the pattern.

PAIR-C also deduces seven features from the Relations and Causality dimensions (first column, Table 1). Among them, there are four *Interaction features* (Feature 1-4, Table 1) identified from the **R**elations dimension. These four *Interaction features* are often perceptible and can provide learners with visual cues to recognize an emergent process (Features 1-4, Table 1). To further illustrate this point, we use the emergent process of ants foraging for food as an example. When ants search for food, they all walk around, emit, and follow pheromones (Feature 1: have the same set of actions). They can follow or stop following any other ant (Feature 2: bi-directional interactions with random others). At any moment, ants can follow other ants without any specific order (Feature 3: occurs simultaneously). Whether an ant follows another is independent of whether another ant follows yet another ant (Feature 4: independent).

In addition to the four *Interaction features*, three other *Inter-level features* (Feature 5-7, Table 1) are identified from the **C**ausality dimension. These *Inter-level features* also cluster multiple attributes as implications of the same feature. For instance, feature 5 specifies some critical attributes of a converging pattern. One is that the initial pattern has no resemblance to the final pattern (e.g., the initial pattern of ants foraging for food is a random distribution of ants. It does not resemble the final pattern of ants forming a single line). Feature 6 clarifies the method to compute the resulting Pattern by adding positive and negative numbers or by averaging out the magnitudes and directions within each unit of time (e.g., the single-line pattern of ants is computed by averaging all the ants' distances and directions - towards and away from the line - at each unit of time; it is the proportion of ants staying on a single line increasing over time, not the absolute number of ants on the single line increasing in each time). Moreover, some of the attributes that describe inter-level causal relationships (i.e., "i. decentralized", "ii. equivalent status", "iii. unintentional", "iv. not teleological", "v. no direct effect", and "non-matching") are classified as implication features for the **C**ausality dimension (see Feature 7a and 7b).

Overall, the PAIR-C framework provides a clear guideline for qualitatively describing and quantitatively computing the causal mechanisms for emergent processes. It is hypothesized that by grasping these PAIR-C features, students can correctly describe and explain inter-level causal relationships for emergent processes.

Table 1
The Seven PAIR-C Features for Describing and Explaining Emergent Processes

Interaction Features	Dimension IV: **R**elations among the agents' interactions
Feature 1	Same set of interactions
Feature 2	Random, bidirectional interactions
Feature 3	Occur simultaneously
Feature 4	Occur independently
Inter-level Features	Dimension V: **C**ausal relationship between the Agents-pattern
Feature 5	Converging change a. The initial pattern has no resemblance to the final pattern b. The changing pattern reflects the interactions of *all* the agents' interactions
Feature 6	Collective summing a. Adding positive & negative numbers within time b. Net effect: Adding magnitudes & directions within each time unit, then comparing across time units (e.g., proportion change)
Feature 7a	All agents are responsible for the pattern i. Decentralized, distributed control ii. Equivalent status iii. Local goal: Unintentional iv. Not teleological/not purposeful v. No direct effect
Feature 7b	No alignment (non-matching) between the agents and the pattern

Using PAIR-c to evaluate ABM integration efforts

In this section, we use the PAIR-C framework as an analytical lens for evaluating previous ABM integration efforts. We also analyze what PAIR-C features were present or absent in previous studies and consider how including/not including such features might affect students' learning outcomes.

In Dickes & Sengupta (2013), after interacting with a Birds & Moths ABM simulation, all students could provide agent-aggregate complementary explanations in which they explained aggregate-level outcomes using the agent perspective. For example, in reasoning the aggregate-level outcome of darker moths becoming more common, students state "the dark moth population will go up because they will have babies because they're not eaten." (p.932, Dickes & Sengupta, 2013). However, this causal statement was not necessarily correct even though it explains an aggregate-level outcome (i.e., dark moth population goes up) in terms of agent-level behaviors (i.e., dark moths have babies, dark moths are not being eaten). Similar statements could be found in students' utterances sampled in other studies (Dickes et al., 2016; Wagh & Wilensky, 2018). According to the PAIR-C framework, these agent-aggregate complementary explanations tend to focus on a subgroup of species (i.e., the dark moths) without considering all interactions between all species at the agent level (i.e., the dark moths, the light moths, and the birds who prey on moths). The causal statement seemed to distinguish dark moths' behaviors from other moths as they have special abilities to have babies and avoid being eaten. Moreover, these statements tend to attribute the aggregate-level outcome as static rather than explain how the agent-level behaviors give rise to the aggregate-level outcome by considering all the interactions among the agents. Therefore, agent-aggregate complementary explanations do not necessarily subsume a correct understanding of inter-level causal relationships. Compared to these student explanations, a more sophisticated causal explanation should state "the dark moth population will go up because in most generations dark-colored moths survive from being spotted by birds and those survived can reproduce, compared to light-colored moths who have lower survival and reproduction rates (due to industrial pollution, trees became darker with soot and thereby dark-colored moths could blend in the environment more easily than the light-colored moths). Over many generations, dark-colored moths will become more common." Since previous studies already used models with multiple breeds of agents governed by the predator-prey relationship (e.g., both birds and moths are shown in the ABM simulations), explicit instruction on the *Relations among the agents' interactions* (i.e., having the *same set* of interactions: all moths can reproduce with each other and can be eaten by birds, see Table 1 Feature 1) should be available to students.

Existing efforts also reveal that researchers often used ABMs to teach the simple idea of "interactions or relations between agents" rather than using ABMs to teach "Relations among the agents' interactions". For example, Basu et al. (2015) used a Saguaran ecosystem model to teach students about the relations between agents (e.g., "Doves eat seeds of the cacti", "Rats eat pods of the ironwood trees", "Hawks prey on rats", "Hawks prey on doves", etc.) without mentioning the relations among the interactions, such as whether one interaction can occur at the same time (*i.e. simultaneously*) as another interaction (missing Feature 3 from Table 1). Their study showed that understanding the interactions between agents did not help students reason in causal chains unless they were scaffolded to notice the *simultaneous and bidirectional* nature of interactions. When asked, "Knowing that hawks eat rats and rats eat pods, what would happen if hawks were removed from the ecosystem?", all students in their study initially stated that "If there were no hawks to eat the rats, rats would increase. So, pods would decrease and disappear soon." What was missing in students' responses was the ability to reason further about the consequences of the lack of pods on the population level of rats (missing Feature 2 from Table 1). To remedy this issue, Basu et al. (2015) later introduced an external representation tool (i.e., the causal map) as a complementary approach to scaffold students' understanding of the bidirectional nature of the food chain causal relationships. By visualizing the bidirectional interactions between pods and rats, hawks and rats using the causal mapping tool, students showed significant improvement in their causal understanding of the Saguaran ecosystem.

Recent efforts used multiple external representations (MERs) to illustrate PAIR-C features and complement the use of ABMs (Su et al., 2021). A pilot study modified ABMs by adding "links" as visual cues to represent agent interactions within the model and later generated enhanced visual representations in the form of videos, animations, and screenshots to show and prompt students about all four *Relations among the agent's interactions* features. However, results showed that there was no significant difference in pre-post test scores between the Control and the Intervention group. Although most of the PAIR-C features were illustrated throughout the simulation module in the pilot study, one limitation was that the PAIR-C instruction was not explicit in teaching ideas about inter-level causal relationships. Therefore, to facilitate students' deeper understanding of natural selection, there is a need to explicitly teach the *Inter-level features*, especially the "converging change" and "collective summing" causal mechanisms through the MERs-complemented ABM approach.

Methods

Participants and settings

The study took place at a Southwestern University in the United States. Participants were a diverse group of online students enrolled in a 7-week technology literacy course, majoring in Education (n = 29, 58%), Social Studies (n = 13, 26%), and Arts & Humanities (n = 8, 16%). Participants were predominantly juniors (52%), sophomores (22%), and seniors (16%). The gender distribution was 76% female students, 22% male students, and 2% non-binary. The average age of the participants was 27.4 years with a standard deviation of 8.91. Participants reported their ethnicity with the following proportions: 68% as White, 14% as Hispanic/Latino, 6% as Asian or Asian American, 6% as Black or African American, 2% as Native American, and 4% as Other. 86% of participants reported having taken 1-2 biology classes at high school, 12% had taken 3 or more biology classes at high school, and only 2% had taken no biology classes at high school. Only two participants reported having heard about and used NetLogo simulation before the study. This paper included a total of 50 participants: 26 in the Control group and 24 in the Intervention group.

Study design and procedure

The study adopted a pretest-posttest randomized block design (RBD). The blocking factor was the participants' scores on the pretest True or False (T/F) questions. The pre-posttest and the surveys were distributed via Qualtrics. This study designed two simulation modules: the Regular ABM Module used in the Control group and the PAIR-C ABM Module used in the Intervention group. To deliver the simulation modules in an online environment, we used NetLogo Web to show the simulation models and Qualtrics to present relevant instructional materials. Participants in both groups received tutorials on how to navigate the NetLogo Web and the Qualtrics-supported instructional page before starting the simulation modules.

Simulation modules

Both simulation modules were about the same length and used the same types of multiple external representations (MERs) including simulation models, explanatory texts, images, screenshots, and videos. We also used prompts to help students actively make connections between MERs by asking them to generate new inferences beyond observing or manipulating MERs. More importantly, students in both groups had the same sequence of activities which allow them to observe, explore, and investigate questions using interactive simulations and MERs. Table 2 presents a detailed comparison between the Regular ABM module and the PAIR-C ABM module and points out the main differences between them.

Table 2

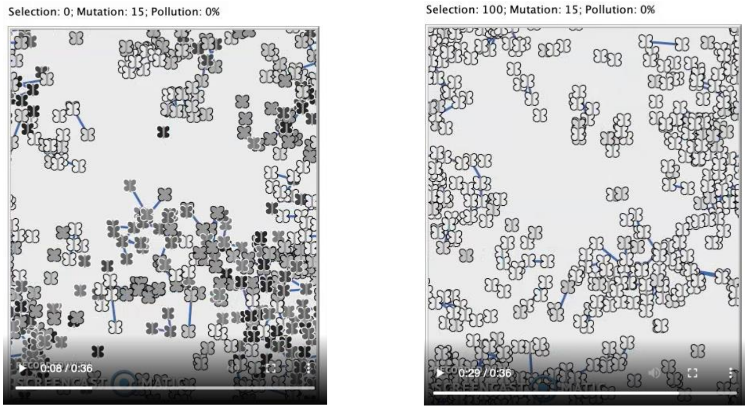
Comparison of the Two ABM Simulation Modules

	Regular ABM Module (Control group)	PAIR-C ABM Module (Intervention group)
Overall Use of Multiple External Representations	Both modules use the same amount of MERs at the same location.	
Simulation models:	Both modules use the same interactive simulations - the Peppered Moths model and the Rock Pocket Mice model- and allowed learners to observe, explore, and investigate these simulations at their own pace.	
Explanatory texts:	Provide texts on the five Darwinian principles (such as genetic determination, adaptation, reproductive advantage), conventional definitions, and common misconceptions but withhold information on the emergent properties shown in the simulations.	Provide texts on the seven PAIR-C features and how to apply these features to explain the emergent properties of natural selection caused by mechanistic details of agent interactions (such as inheritance, predation, and reproduction).
Images, screenshots, videos:	All representations were generated from the agent-based models directly and did not contain external visual	Most of the representations were modified and provided visual cues (e.g., blue links were used to

	cues to represent different types of agent-level interactions.	represent mating interaction. See example in Figure 1)
Overall Use of Prompts	Both units use the same number of identical generic prompts that contain the same types of questions.	

Figure 1

Screenshots of the Simulation Video with Different Selection Pressure



Note. The screenshots were taken from the link-visible ABM video used in the PAIR-C intervention group. The video in the control group did not show any links.

Pre-posttest instrument

Understanding of Natural Selection was assessed by the pre-posttest instrument. The pretest and the posttest are identical despite the sequence. Most of the questions were revised from the AAAS conceptual inventory and past AP Biology Exams. The revised items were validated by a high-school biology teacher who was a collaborator with our team and previously taught the PAIR-C framework in his natural selection lessons. The test contained three sets of True or False (T/F) questions with five statements for each set. These fifteen T/F statements were designed in the format of two-tiered questions. The first tier simply asked students to decide whether a statement was True or False. The second tier asked students to explain the statement they picked as False. In our previous experience with T/F explanations, students tend to rephrase or copy the True statement without giving an explanation. Therefore, the pre-posttest in this study only asked students to explain the False questions to reduce the likelihood of guessing while remaining a relatively shorter test. For the True/False selection part of the question, if students correctly determined whether the statement was True or False, they were given a score of 1, otherwise a 0 point. For the open-ended part of the T/F items, we created scoring rubrics for each false item (on a scale of 0, 0.5, and 1).

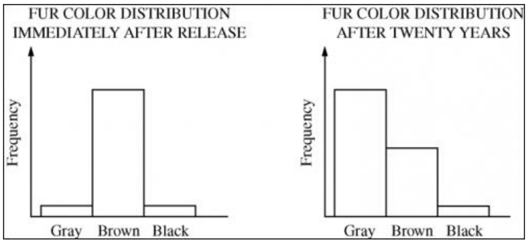
Seven out of the 15 T/F questions were labeled as Shallow which assessed basic understandings of Darwinian principles or knowledge of either agent-level or pattern-level understanding while eight out of the 15 T/F questions were Deep questions that assessed understanding of inter-level causal relationships. For example, the following T/F statement *"Changes you observed in the mice's fur color patterns from the start of the experiment to the end were wholly due to random factors"* was considered a shallow question. The reason was that it did not require an explanation or understanding of how the mice agents' interactions produce the pattern of fur color change, and 64% of the participants answered it correctly in the pretest. In contrast, another T/F statement *"The number of gray mice in the population increases slightly each year for 20 years, which adds up to the pattern of gray fur becoming more common in the population"* exemplified in Table 3 was considered a deep question because it requires a deeper understanding of the inter-level causal relationships. A correct causal explanation of this statement would imply that there was no alignment between the agents and the pattern as well as no continual increase in the number of agents for each generation. Only 20% of participants correctly determined this statement as a false statement in the initial pretest.

The test also contained two open-ended (OE) questions assessing students' understanding of the natural selection process in different contexts (to be referred to as context transfer items). These two context transfer questions were adapted based on Peel et al. (2019). Similarly, we also created scoring rubrics for each context transfer item to assess students' transfer performance on a scale of 0 to 5. For each open-ended item, more than 50% of student responses were scored by the first author and the third author for inter-rater reliability over three rounds (Kappa for the open-ended part of the T/F questions was an average of 0.843, $p < 0.001$, Kappa for OE context transfer questions was 0.925, $p < 0.001$, indicating strong scorer agreement). Table 3 provides examples of the pre-posttest items, the scoring rubrics, along with sample responses from students.

Table 3
Pre-posttest Items

True or False (T/F) Question

Q1. Decide True or False for each of the following statements that explains the change in the frequency distribution of fur color in the mouse population after 20 years, as shown in the figures below. Provide explanations for the false statements you identified.



A. The number of gray mice in the population increases slightly each year for 20 years, which adds up to the pattern of gray fur becoming more common in the population.

Rubric for explaining statement A:
This statement is False.

- ***Score 1*** *if the explanation mentions that the increase is not gradual or continuous by every generation OR refers to the "not align or non-matching" PAIR-C feature.*
- ***Score 0.5*** *if only refers to not enough data or information.*
- ***Score 0*** *if the answer only describes the bar chart.*

Sample responses for explaining statement A:

- S39: "One year there could be a surge of gray mice and one year there could be many dying instead. It is *not a continual increase*."
  Score 1
- S42: "While the grey mice population did increase significantly, *without more data we don't know* that it increased slightly for 20 years."
  Score 0.5
- S38: "The population of mice with grey fur dramatically increases."
  Score 0

Open-ended Context Transfer Question

Q4. How would biologists explain how a living mouse species with claws evolved from an ancestral mouse species that lacked claws?

Rubric:

- ***A score of "5"*** *should be given to responses that mention:*
  - *trait variation present before selection pressure*
  - *random mutations result in trait variations*
  - *mice survive and reproduce*
  - *influence of selective pressure or environment on survival and reproduction rate*
  - *changes occur over time or multiple generations*
- ***A score of "1-4"*** *should be given to responses that contain 1 to 4 key points mentioned above.*
- ***A score of "0"*** *should be given to responses that contain no key point.*

Sample response from S5:

"If one mouse was born with *a mutation in their DNA* which produced claws [score 1 for random mutations]/, once they reproduce there is a *good chance the mutated gene will present* in their offspring. [score 1 for trait variation]/ As the trait is passed down *through generations*, the population of the mutated mouse species would increase [score 1 for over time/generations]/. The claw likely provided a form of defense toward their predators, *making the mice without the claws an easier prey* [score 1 for the influence of selective pressure]. As those with claws *survived at a higher rate*, they also *reproduced at a higher rate* due to the overall lack of mouse without claws in the population [score 1 for survive and reproduce]/."

(A total of five points was given to S5's response).

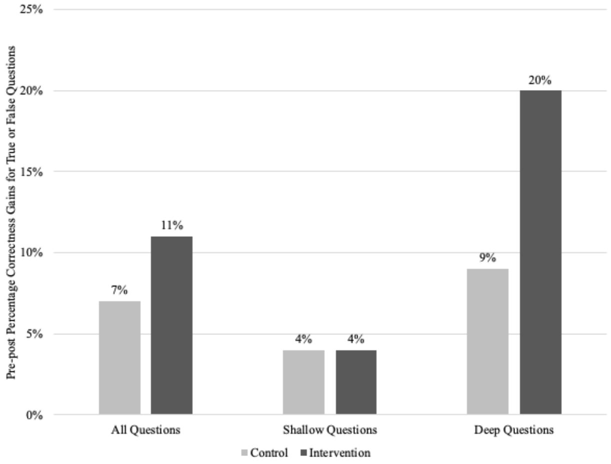
Results & discussion

Results show that student performance on the T/F items improved significantly from the pretest to the posttest for both groups. The control group (N =26) improved from a mean score of 8.94 (SD = 3.09) to a mean score of 10.89 (SD = 2.87), t =2.60, p = .016, with a medium effect size d = 0.51. The intervention group (N = 24) improved from a mean score of 9.31 (SD = 2.78) to a mean score of 12.73 (SD = 4.36), t =4.42, p < .001, with a high effect size d = 0.90. These results indicate that both the Regular ABM simulation modules and the PAIR-C ABM simulation modules used in this experiment could successfully improve students' overall understanding of natural selection. To investigate group differences attributed to the different simulation module treatments, ANCOVA was conducted using pretest scores as covariates. The result shows that there was a marginally significant difference between the two groups in the overall understanding of natural selection (F = 2.94, p = .093, partial η^2 = .059). This difference is also reflected in the first two columns of Figure 2. As shown in the figure, the percentage of correctness gains was 7% for the control group whereas the percentage of gains was 11% for the intervention group. This result indicates that the intervention group who used the PAIR-C ABM simulation modules had a better performance on the overall understanding of natural selection compared to the control group who used the Regular ABM simulation modules.

Results also show that students improved significantly in answering deep questions from the pretest to the posttest for both groups. The control group improved from a mean score of 2.54 (SD = 2.16) to a mean score of 3.73 (SD = 2.52), t =2.08, p = .048, with a medium effect size d = 0.41. The intervention group improved from a mean score of 2.52 (SD = 2.15) to a mean score of 5.56 (SD = 3.95), t =3.90, p < .001, with a high effect size d = 0.80. These results indicate that both the Regular ABM simulation modules and the PAIR-C ABM simulation modules used in this experiment could successfully improve students' deep understanding of natural selection. ANCOVA analyses show that there was a statistically significant difference between the two groups in answering the deep questions of natural selection (F = 4.16, p = .047, partial η^2 = .081). This difference is reflected in the last two columns of Figure 2. As shown in the figure, the percentage of correctness gains for answering deep questions was 9% for the control group whereas the percentage of correctness gains was 20% for the intervention group. These results indicate that the intervention group had a significantly better performance in answering deep questions of natural selection compared to the control group after the simulation modules. In other words, the PAIR-C ABM simulation modules were more effective in fostering students' deeper understanding of natural selection compared to the Regular ABM simulation modules. In contrast, students' performance did not differ between groups in answering shallow questions (F = .01, p = .934, partial η^2 = .000). This pattern is shown in the middle two columns of Figure 2.

Figure 2

Pre-post Percentage Correctness Gains for All, Shallow, and Deep T/F Questions



To see whether the PAIR-C ABM simulation module could further impact students' abilities to explain the process of natural selection, analyses were conducted based on their responses to the two context transfer questions. Results show that there was no significant improvement in answering the two context transfer questions from the pretest to the posttest for both groups. The control group improved from a mean score of 1.73 (*SD* = 1.34) to a mean score of 1.77 (*SD* = 1.75), $t = 0.09$, $p = .928$, with a small effect size $d = 0.02$. The intervention group improved from a mean score of 2.04 (*SD* = 1.68) to a mean score of 2.38 (*SD* = 2.36), $t = 0.78$, $p = .445$, with a small effect size d = 0.16. For testing group differences in answering the post-test context transfer questions, ANCOVA was conducted. There was no statistically significant difference between the two groups ($F = 0.70$, $p = .409$, partial $\eta^2 = .015$). Similar results were also manifested in students' responses to the simulation prompt question.

The above findings suggested that the ABM simulation module integrated with the PAIR-C features could contribute to deeper learning of emergent process concepts, such as natural selection. Nevertheless, students in both groups were not able to demonstrate significant improvement in explaining the process of natural selection across different contexts. One reason could be that students did not receive enough instructions on how to use PAIR-C features or Darwinian Principles to elaborate on the process of natural selection in the simulation module. More explicit instructions should be provided to scaffold students learning.

Conclusion

In conclusion, this study shows the potential of using the PAIR-C framework in designing agent-based simulation modules for learning complex emergent phenomena. Using natural selection as the exemplary concept, the PAIR-C framework has its ecological validity as it underscores the importance of understanding the inter-level causal relationships for emergent science processes taught in school curricula. Moreover, our approach of integrating the PAIR-C framework using MERs-complemented ABM seems promising to be operationalized for science instruction broadly. We believe it can impact how science teachers produce curricula, create assessment rubrics, and use computer technologies to help students reach a deeper understanding of emergent causal mechanisms in science processes.

References

Chi, M. T. H., Kristensen, A. K., & Roscoe, R. (2012a). Misunderstanding emergent causal mechanism in natural selection. Evolution challenges: Integrating research and practice in teaching and learning about evolution, 145-173. New York, NY: Oxford University Press.

Chi, M. T. H., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012b). Misconceived causal explanations for emergent processes. Cognitive science, 36(1), 1-61. https://doi.org/10.1111/j.1551-6709.2011.01207.x

Chi, M. T. H., Su, M., VanBibber, B., Cho, J., Ding L., Boucher, N. (under review). Individualistic versus collective causal structures.

Basu, S., Sengupta, P., & Biswas, G. (2015). A scaffolding framework to support learning of emergent phenomena using multi-agent-based simulation environments. *Research in Science Education*, *45*(2), 293–324.

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, *27*(5), 415–427.

Dickes, A. C., & Sengupta, P. (2013). Learning natural selection in 4th grade with multi-agent-based computational models. *Research in Science Education*, *43*(3), 921–953.

Dickes, A. C., Sengupta, P., Farris, A. V., & Basu, S. (2016). Development of Mechanistic Reasoning and Multilevel Explanations of Ecology in Third Grade Using Agent-Based Models. In *Science Education* (Vol. 100, Issue 4, pp. 734–776).

Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. *Evolution*, *2*(2), 156–175.

Peel, A., Sadler, T. D., & Friedrichsen, P. (2019). Learning natural selection through computational thinking: Unplugged design of algorithmic explanations. *Journal of Research in Science Teaching, 56*(7), 983-1007.

Su, M., Cho, J., Chi, M. T. H., Boucher, N., & Vanbibber, B. (2021, June). Designing Simulation Module to Diagnose Misconceptions in Learning Natural Selection. Long paper presented at the 15th International Conference of the Learning Sciences, Online.

Wagh, A., & Wilensky, U. (2018). EvoBuild: A quickstart toolkit for programming agent-based models of evolutionary processes. *Journal of Science Education and Technology*, *27*(2), 131–146.

Wilensky, U., & Rand, W. (2015). *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*. MIT Press.

Acknowledgments

This research was funded by the Institute of Education Sciences (R305A150336).