Iterative Drawing Reveals Diversity and Change in Student Thinking About Evolution

Spencer Buck[1] & Andrew Martin[1]

[1] Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, 30809, USA

Correspondence: Andrew Martin, Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, 30809, USA. Tel: 1 303-325-1790. E-mail: am@colorado.edu

Received: December 15, 2022 Accepted: January 12, 2023 Online Published: January 18, 2023

doi:10.5430/ijhe.v12n1p17 URL: https://doi.org/10.5430/ijhe.v12n1p17

Abstract

One theoretical framework for learning is that knowledge exists in pieces. Understanding emerges by developing a coherent and stable integration of multiple ideas. When learning gains are evaluated iteratively, over time, it is clear that the mental constructs representing key ideas are dynamic. Here we describe an assessment framework that enables estimation of the dynamic nature of mental constructs as students make gains towards coherency of knowledge and understanding. The framework emphasizes the value of iterative assessment combined with multivariate methods borrowed from ecology for revealing and following gains in student thinking. We applied our framework for monitoring and describing student gains in their abilities to visualize and describe the process of evolution. Our approach was observational. We evaluated 276 drawings and accompanying text-based descriptions of evolution generated by 102 students by implementing the same open-ended assessment question four times in two sections of an upper division evolutionary biology course during spring 2021. Based on a binary rubric of 10 key ideas, students showed evidence of gains and losses of key ideas over time, and their learning trajectories were diverse and dynamic. Our findings revealed students take a multitude of pathways to concept mastery and that they struggled to succinctly construct and communicate comprehensive evolutionary models. Based on our study, we recommend using iterative free-response assessment with an explicit rubric and multivariate non-metric dimensional scale data visualization for revealing student thinking and guiding data-driven revision of curriculum, teaching strategies, and assessment for achieving greater coherency and stability of knowledge.

Keywords: evolution, natural selection, education, undergraduate, student thinking, dynamic, visualization, diversity

1. Introduction

1.1 Pedagogical Framework

Knowledge-in-Pieces (KIP) is, according to its originator, "...a broad theoretical and empirical framework aimed at understanding knowledge and learning (diSessa 2018: 65)." There are several relevant aspects of the KIP framework: knowledge is a complex system with many types of elements; the structure of the complex system involves building and continuous revision of models; the system exists across multiple scales, including short and long periods of time and smaller and larger knowledge elements; the complexity and the dynamic nature of the system means there are many potential and real differences among learners; and knowledge and learning depend on context (diSessa 2018). One key feature of KIP, and other models of learning, is that knowledge and learning are complex phenomena and change can happen across multiple dimensions. Curriculum and assessments designed to advance and estimate knowledge must be tuned to different scales of time and knowledge unit size. Assessment, in particular, should be constructed to reveal change in a students' knowledge of small elements and the larger emergent conceptual structures (e.g. theory) over both short and longer time scales. Ideally, assessment is repeated and gains, both small and large, are estimated in ways that reveal student learning trajectories.

Assessment often reveals student knowledge is variable over time. The variability of student knowledge suggests knowledge exists as dynamic mental constructs (DMCs)(Sherin et al. 2011). Knowledge is an assembly of multiple nodes and is dynamic because the same query implemented at different times can elicit a different combination of nodes (Sherin et al. 2011; Lira and Gardner 2020). Thus, a key aspect of evaluating student learning is repeated assessment over time, ideally using prompts that address the different KIP levels. In general, constructed response questions enable better estimation of knowledge and understanding provided students are capable of effectively communicating what they know (Stanger-Hall 2012). Ideally, students develop visualizations or models of phenomena as part of describing their knowledge (citation).

To gain insight about KIP and DMCs inherent in the process of learning, we adopted a framework and assessment method that reveals the dynamic nature of knowledge and learning. A key aspect is repeating the same (or similar) assessments for estimating students gains and variability. Our framework combines an analogy for problem solving in which there may be many and varied pathways towards success with a statistical method—commonly used in community ecology—for revealing the similarity of DMCs. Stanley and Lehman's (2015) described a provocative analogy for achieving a goal in the absence of a known pathway to success. In their framework, they imagined achieving a complex, unspecified objective was similar to "...the problem of crossing a lake on foot by hopping from stepping stone to stepping stone…" subject to the constraint that "...the lake is covered in mist…" making it impossible to see where the stepping stones lead, and which combination of stones leads to the other side of the lake. Each of "...the stepping stones are waypoints that must be crossed to reach the objective...The fundamental problem...is that we usually don't know the stepping stones that lead to the objective at the outset. After all, if we always knew the stepping stones, then everything we hope to achieve would be easy." (Stanley and Lehman 2015: 29).

The statistical framework is non-metric dimensional scaling (NMDS). The NMDS method enables constructing the "lake with stepping stones" by assuming each stepping stone is a combination of mental constructs. These mental constructs can be identified in student answers from free response questions using a binary rubric. Thus, for understanding that may require successful integration of five different constructs, there are $2^5 = 32$ different stepping stones representing all the combinations of 0 and 1 for the five rubric items. The lake is defined by an x-y Cartesian NMDS space that can include all of the individual combinations of binary rubric scores. A novice is represented as five zeros (i.e. 00000) and an expert as five ones (i.e. 11111). Each of these stones will be on opposite sides of the lake because they are the two most numerically different states across all 32 possible stepping stones. In this model, students can take a large number of different pathways—combinations of stepping stones— when crossing the lake.

Student learning trajectories can vary. Depending on the interaction between student knowledge and teaching, we can imagine that if cognitive constructs are dynamic, students may show evidence of reversals (a change from 1 to 0 indicative of the loss of a cognitive construct) and learning trajectories that may be tangential to the other side of the lake (Figure 1, left). Alternatively, students may conform to a model in which they gradually switch the rubric scores from 0 to 1 without reversals for an increasing number of cognitive constructs. In this scenario, the pathways across the NMDS-defined lake are more direct (Figure 1, right). In both cases there are multiple pathways and the variation among students is easy to see.

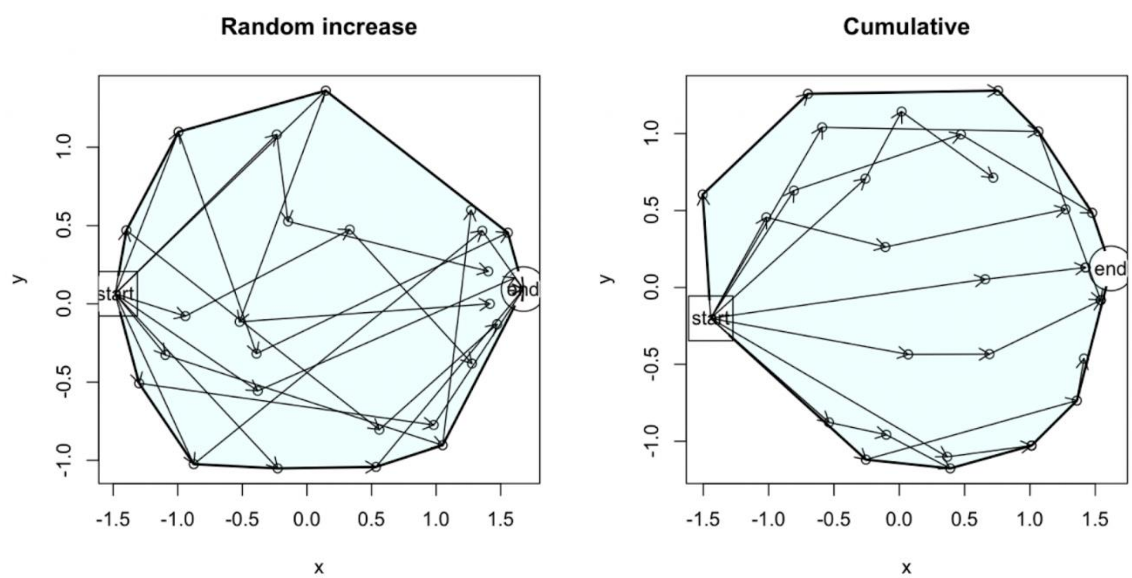


Figure 1. Pictures of the "lake" and possible student trajectories for four iterations of an assessment

Description: Each open circle is a combination of zeros and ones for 10 rubric items. Open circles to the left have fewer ones and open circles to the right have more ones. The arrows show 10 example student trajectories. Left. The average score for each iteration increases but the specific rubric items gained is random and ignores past rubric scores. This model allows for the reversal (1 → 0) of rubric items. Right. The average score each iteration increases and the gain is cumulative. This model does not allow for the reversal of rubric items.

1.2 Applying the Framework for Teaching and Learning Evolution

We apply our KIP-grounded conceptual and analytical framework to the problem of teaching and learning evolution. A focus on evolution is particularly useful because, as a phenomenon, coherent understanding requires learning gains for multiple, key ideas. Important key ideas include that individuals reproduce in excess of available resources, there is variation among individuals with effects on function, offspring differ from their parents but are nonetheless similar as a consequence of heritability, there are differences in survival and reproduction among individuals, and the outcome of within generation gain and loss of individuals is influenced by stochasticity (i.e. drift). Additionally, evolution requires mastering more abstract ideas stemming from the fact that the process of evolution is not easily observed; it happens over multi-generational time scales; and the topic may elicit or reinforce anthropocentric, essentialist, teleological, and progressive perspectives about life (Allmon 2011; Coley and Tanner 2012, Evans 2001; Sinatra et al. 2008).

Just as there are many ways to combine the various key ideas to form a pathway from novice and expert, there are a variety of ways to draw visualizations of evolution. Darwin provided an example of an expert visualization in *On the Origin of Species* (Darwin 1859, see figure 2). Aside from being widely used within the scientific community and biology texts (Catley and Novick 2008), tree-like representations of evolution accurately convey key aspects of evolution (Baum et al. 2005). Additionally, Stephens (2012) showed, when properly understood, tree-thinking can also aid in dispelling common misconceptions concerning teleology, progressivism, and intentionality. Another type of drawing shows specific scenarios (e.g., white and black peppered moths against light or dark backgrounds, giraffes with long or short necks feeding on tall trees, and the beak sizes of different species of Darwin's finches). Another commonly constructed visualization is the depiction of evolution as a linear progression, an approach matching the iconography evident in popularizations of evolution (Gould 1989). In many student drawings showing progressions , the starting (ancestral) stage is a single cell, or a more derived organism (fish, reptile, mammal, or primate), and humans are typically the end point. Scenarios and linear progressions are the most common types of drawings evident for individuals with demonstrably novice understanding of evolution. Occasionally, students will draw graphs, phylogenetic trees, or abstract images with different shapes or symbols. All of these types of visualizations can provide the basis for effectively and accurately representing the process of evolution; however, each type of drawing imposes unique constraints on how key principles of evolution are represented.

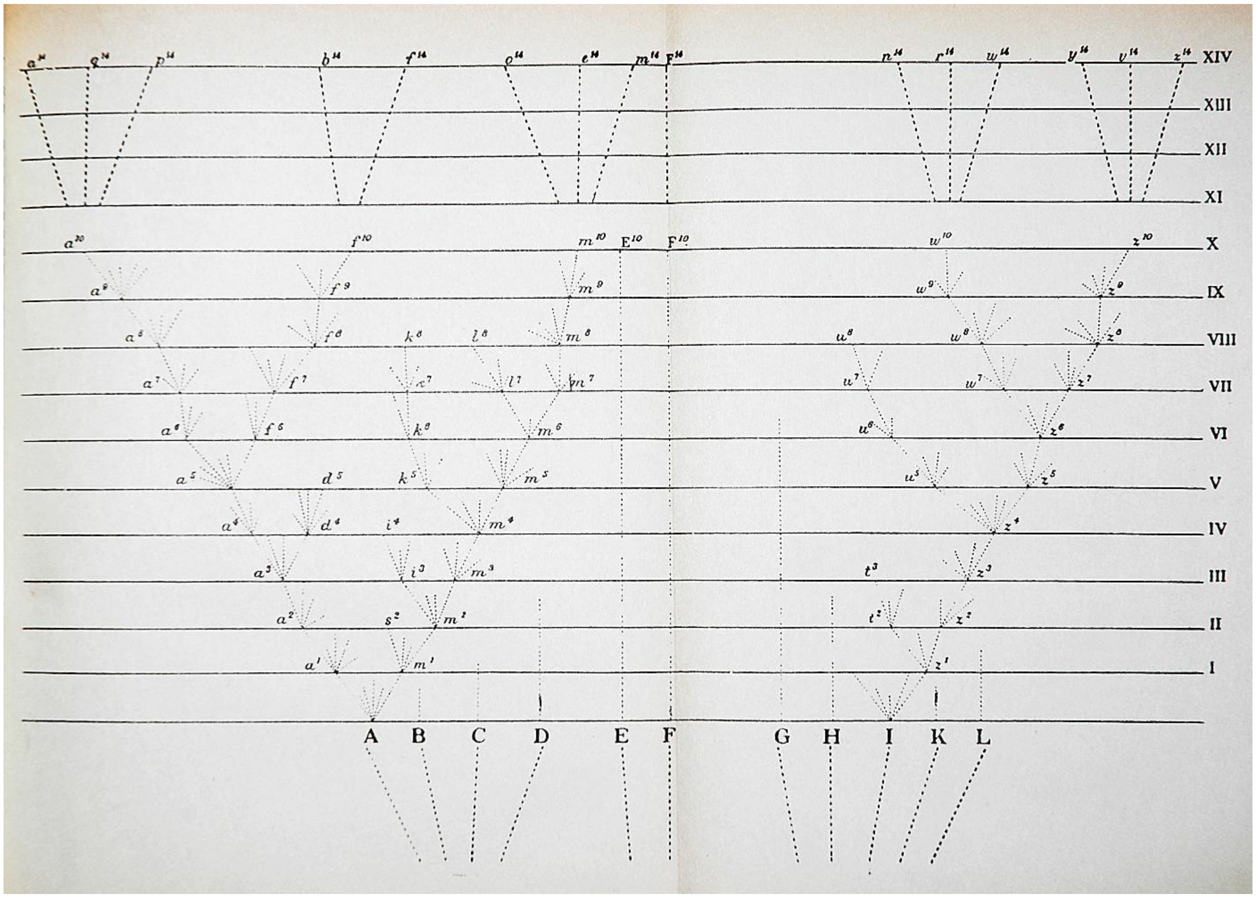


Figure 2. Darwin's canonical illustration of evolution from the original first edition of On the Origin of Species by Means of Natural Selection or The Preservation of Favoured Races in the Struggle for Life (Darwin 1859)

Description: Five fundamental principles of evolution are visible in the drawing: evolution results in the change in the characteristics of individuals comprising populations over time; individuals differ from each other (there is variation among individuals); the variation is heritable because offspring are, on average, similar to the parent); reproduction results in an excess of individuals than can be supported by natural resources resulting in mortality; and there is differential survival and reproduction among individuals. As noted in the text, "The intervals between the horizontal lines… may represent each a thousand generations", while "A to L represent the species of a genus large in its own country."

1.3 Our Study

In this study, we asked students to construct visual representations of evolution and use the visualization to describe evolution in words. Students were assigned the same drawing prompt multiple times during the semester and the visualizations were evaluated based on whether specific concepts or principles were apparent. Our approach was designed to answer two key questions about how students learn core principles of evolution. Are learned concepts stable or is there instability such that competency for key concepts is lost over time? What is the diversity of pathways from novice to expert? The purpose of the study was to collect data for engaging in data-driven revision of curricula, assessments, and teaching strategies aimed at emphasizing constructing predictive, accurate, and effective representations of evolution.

2. Methods

2.1 Characteristics of the Course

We evaluated student thinking in an upper-level evolutionary biology course (EBIO 3080) at a public tier 1 research University (approved under Institutional Review Board protocol 21-0024). No personal data about the participating students (education level, familial background, gender, age, etc.) was collected or analyzed. We adopted multiple analytic frameworks for more thorough estimation of student thinking (Lira and Gardner 2020, National Research

Council 2012, Pelaez et al. 2005). Importantly, the research team was independent of the two participating course sections. We participated as unbiased observers with no stake in, or influence over, course teaching content, material, or methods.

We enlisted the participation of two professors (and their students). Both professors were experienced at teaching the course and neither were involved in biology education research. One of the sections was a Monday-Wednesday-Friday class with three fifty-minute sessions per week and the other was a Tuesday-Thursday class with two 75-minute sessions per week. Both classes were taught remotely and synchronously using Zoom during the pandemic. There were 59 and 69 students in each of the two sections, respectively. Both sections of the course had the same stated learning goals in the syllabus: “By the end of this course, you will be able to: 1) Construct and evaluate arguments based on evidence, 2) Identify, create, and evaluate alternative hypotheses, 3) Use hypotheses to direct data collection and analysis, 4) Graphically, verbally, or quantitatively represent evolutionary problems and models, 5) Collaborate with others to reach common goals, and 6) Communicate with brevity, clarity, and scientific persuasion.” The course description and structure were identical between the two sections. There were, however, some differences between sections. For instance, only in the larger of the two sections was concept mapping taught as a learning strategy during an instructional module; nonetheless, we did not attempt to quantify the differences between the courses or assess whether there were differences in gains between the two courses. Instead, we focused on estimating the extent learning corresponds with the predictions stemming from the theoretical frameworks of knowledge-in-pieces and dynamic mental constructs.

2.2 Study Design

Students were presented an overview of the study and given the opportunity to sign a Qualtrics digital consent form opting-in to the review, analysis, and potential publication of their visualizations and concept inventory scores. While participation in the study was optional, students received course credit for completion and submission of their answers to the assessment prompts. Students were also afforded the opportunity to exclude their submissions from review yet still receive course credit for submitting the assignment. Specific and identifiable student products were used only for students who completed the consent form. Data was anonymized and were assigned a number identifier.

There were two main sources of data. First, to gauge student understanding of evolution, we implemented a validated concept inventory (DeSaix et al. 2011) at the beginning and end of the semester. Second, we asked students to “Construct a unique and personalized visualization (not a google image search result!) of your understanding of evolution and use the visualization to explain how evolution works. Try to depict evolution as thoroughly and completely as possible.” This free-response prompt was administered at four evenly spaced intervals throughout a semester (weeks 1, 5, 10, 15). The assignments were introduced by the respective section professors during lecture and assigned through the Canvas course websites. Students were allowed to submit any image file type (.jpg, .png, .pdf, etc.) for credit. Importantly, this study was conducted during the Covid-19 pandemic and all student work happened with full access to the internet and other potential resources.

2.3 Collecting the Data

Completed visualizations were collected and categorized into 1 of 5 common drawing types : scenarios, phylogenies, abstractions, graphs, and linear progressions. Scenario-based visualizations generally included an illustration of a common evolutionary example, including but not limited to Darwin’s finches, peppered moths, the length of giraffe’s necks, or predator-prey relationships. Phylogenies varied in complexity, scope, and detail and included basic branching tree structures as well as Darwinian time-scaled illustrations. Abstract drawings included many different forms of visualization from infographic style depictions, concept maps, and artistic renditions of phylogenies or other artistic interpretations of evolution. Graphs were generally simple and typically depicted trait frequencies and disruptive or directional selection. Linear progressions imitated the typical iconography of the ‘chimp to man’ image or a similar linear progression of organisms or objects.

The scoring of the visualizations was based on commonly and generally accepted key concepts in education research for evolution by natural selection as well as from Charles Darwin’s depiction of evolution in *On the Origin of Species* (Darwin 1859, Nehm and Reilly 2007, Moharreri et al. 2014). We included 10 total rubric items and graded for presence or absence (1 or 0) in the student’s visualization and descriptive text. Scores were recorded and tracked with anonymized student IDs by both section and iteration. Students received full course credit for completion of the visualization assignment. To our knowledge, students were not aware of the rubric or grading criteria. There were ten binary rubric items: 1) evolution occurs in a population, 2) there are at least five individuals, 3) there is variation among individuals, 4) the variation is heritable, 5) there is reproduction, 6) there is evidence that reproduction exceeds the available resources, 7) there is evidence resource are limited, 8) there is evidence of competition among individuals, 9) there is evidence of differential survival depending on phenotype, and 10) there is evidence of a change in the population across generations. Figure 3 provides an example of an expert illustration.

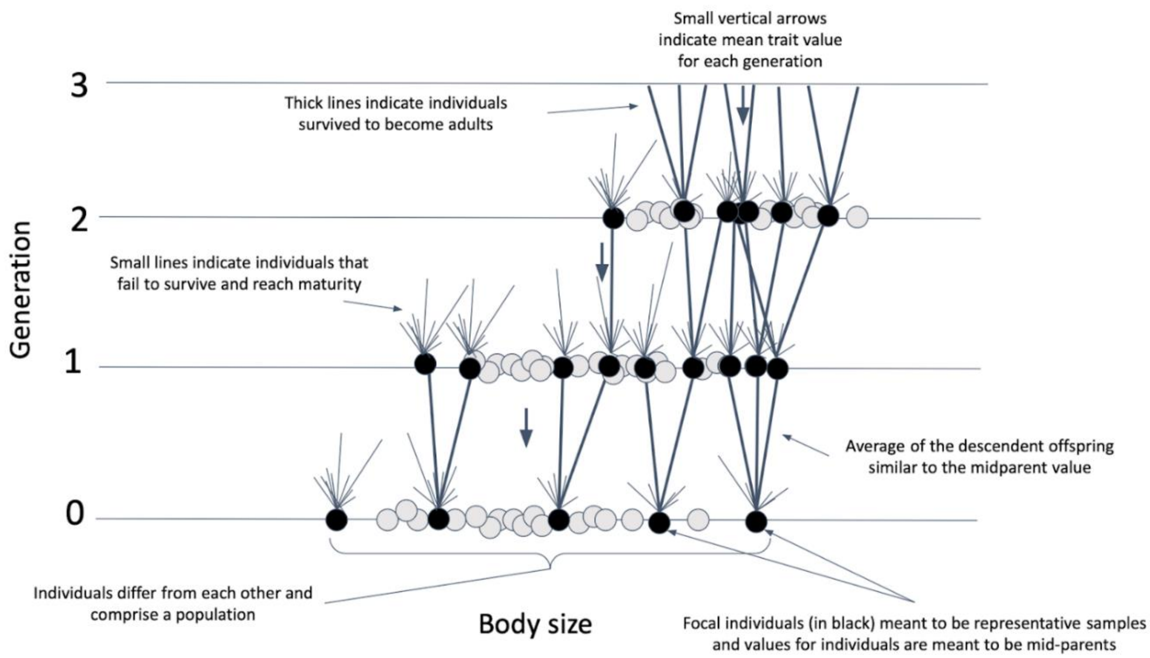


Figure 3. An example of an expert-like drawing with annotations highlighting various key concepts

Description: The visualization provides evidence that evolution happens in a population; there are many offspring produced by parents and most die (indicated by the short lines emerging from the black circles); the offspring body sizes are, on average, similar to the parents (indicative of heritability); there is natural selection (in this case, larger individuals survive and reproduce); and there is change in the average body size over generations (indicated by the downward-pointing short vertical arrows).

Scoring of the visualizations and text descriptions, including the categorization of drawing type, was performed by two graders. Inter-rater reliability (IRR) kappa was 0.85, and a test-retest reliability kappa of 0.82. IRR values were calculated using the IRR package in R (Gamer et al. 2019). Inconsistencies or indeterminate rubric items or drawing type categories were discussed and a consensus reached for all questionable rubric items or drawing types.

2.4 Data Analysis

Using basic functionality within R Studio (R Core Team, 2020) we ran a linear mixed model analysis with ‘student’ as a random variable to assess the effect of iteration on overall cumulative score. Using a similar mixed linear model, we also estimated differences in mean summed score among different drawing types.

Because students completed the same assessment multiple times, we could assess the extent individuals changed what and how they communicated in both words and illustration. We evaluated whether students changed the type of drawing they constructed and the gain (or loss) of rubric items evident in their work. We estimated concept instability as the number of rubric items that changed from one to zero (presence to absence) over the course of multiple iterations. We plotted concept instability relative to the sum of the number of rubric items assigned a one across all iterations as an estimate of the gains and stability of student understanding of evolution.

We adopted and implemented Stanley and Lehman’s (2015) analogy of the stepping-stones on a mist-shrouded lake using nonmetric multidimensional scaling (NMDS). NMDS uses data on the presence or absence features to create a distance matrix between all pairs of assessment rubric scores. The distance data were subjected to NMDS analysis in a manner that created two vectors of numbers useful for describing the similarity of individuals’ 10-item presence-absence rubric scores. A bivariate plot of the NMDS scores was used to visualize the similarity of understanding among students; points closer in the 2D space are more similar. The two sides of Stanley and Lehman’s (2015) mist shrouded lake were defined using scores with all zeros and all ones to represent novice and expert, respectively. Each point in the two-dimensional space represented a different combination of binary scores across the

10 rubric items and the different combinations evident in student work defined the stepping-stones. Individual student pathways "across the lake" were visualized by joining NMDS scores across iterations with arrows (see figure 1 for two examples). We used the Vegan (Oksanen et al. 2019) package in R Studio version 1.3.1073 (R Core Team, 2020) to implement the method.

Finally, we calculated the diversity of student answers based on the composite rubric scores (e.g. the concatenated binary scores) and followed the changes in diversity across the four iterations of the assessment.

3. Results

We evaluated student gains in their understanding of evolution using the same assessment questions multiple times during a semester. Our multiple-choice questions were from a validated concept inventory and our free response question was developed over several years that emphasized students' ability to make their thinking visible and evident through drawings and text.

3.1 Student Participation

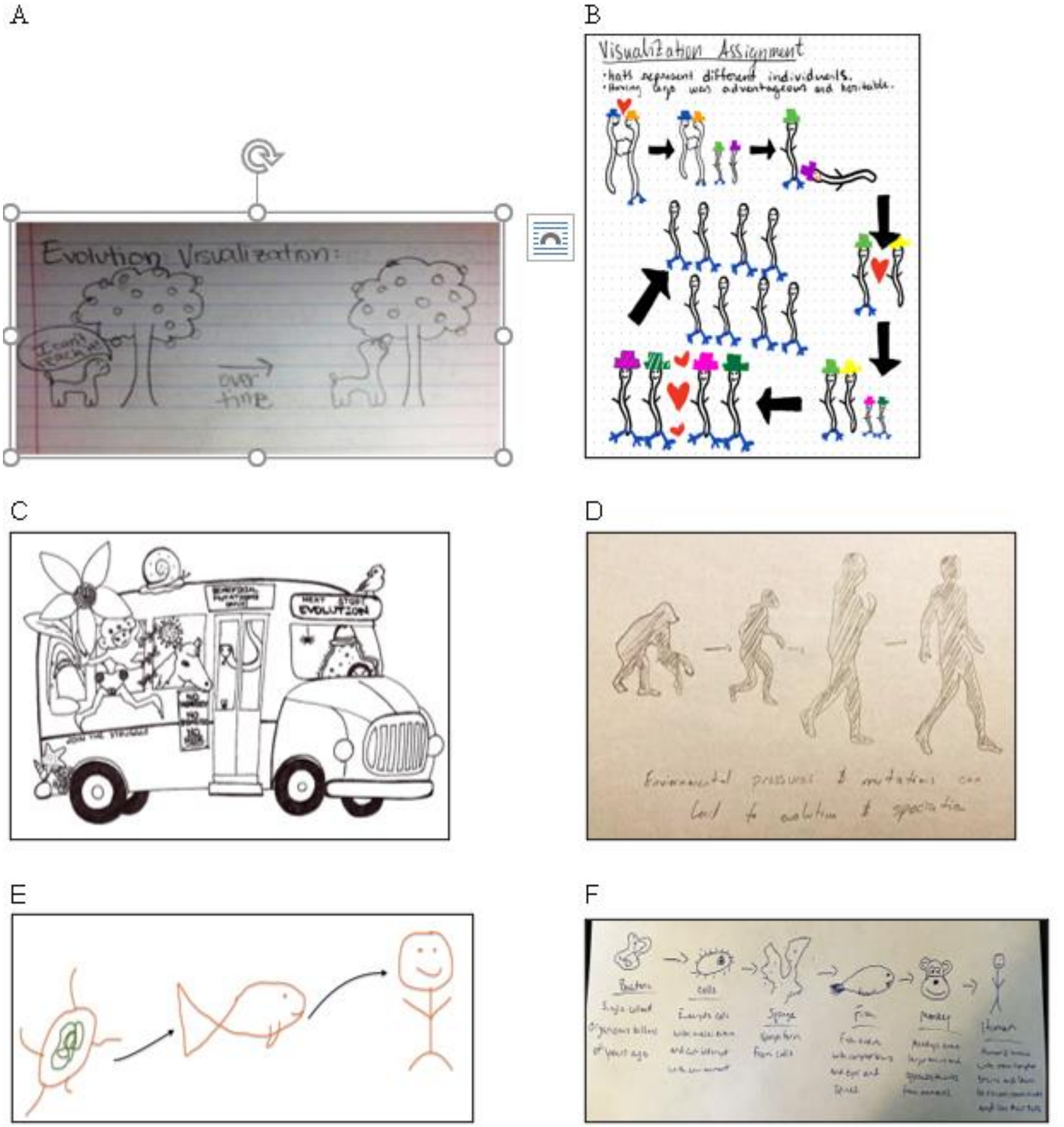
Records indicate that 155 students started the course and 27 (17%) students dropped or withdrew at some point during the semester, leaving 128 students who completed the course. Of the 128 total students across both courses, 65 students completed the pre-course CI and 57 completed the post-course CI. We collected at least one visualization from 107 different students and 276 total visualizations across four separate iterations of the assessment. There were 49 students in the two sections who completed at least three of the four visualization assignments. In general, there was significant attrition: more than 100 students completed the first iteration while fewer than 40 completed the fourth iteration. Additionally, participation differed markedly between the two different instructors.

3.2 Concept Inventory Data

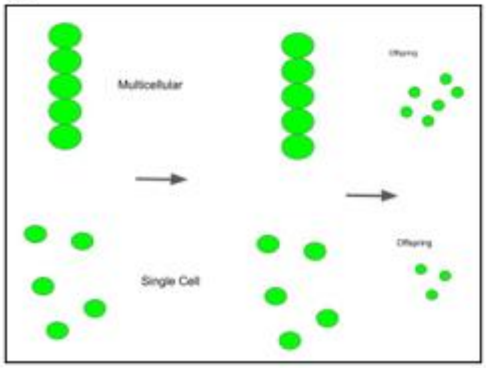
Average student scores on the concept inventory were high at the beginning (9.2 of 12 possible) and end (10.5 of 12 possible) of the course. Based on published work, there were four notable and common misconceptions. The responses indicated 37% of respondents believe that "natural selection gives organisms/species what they need, natural selection involves a will, effort, or intent on the part of the organism/species"; 18% answered in ways that showed "evolution results in progress; organisms are always getting 'better' or more complex through evolution"; 28% of answers were consistent with the view that "acquired characteristics can be inherited"; and 22% revealed "species can evolve the traits necessary for survival and reproduction no matter what."

3.3 Analysis of Students' Constructed Responses

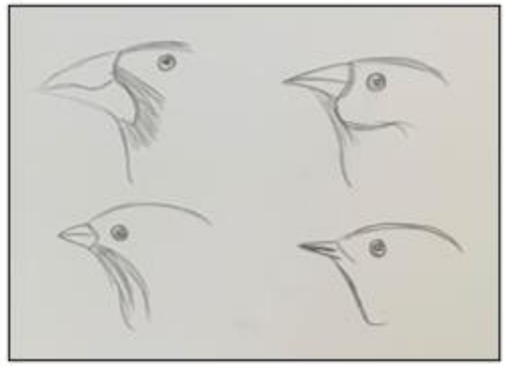
Students drew a wide variety of pictures of evolution (Figure 4). Overall, we classified their drawings into four common types: abstract, phylogeny, progressions, and scenarios. The frequency of the different types of drawings changed over the four assessment iterations; scenarios were most common and the frequency of phylogenies increased concomitantly with a decrease in scenarios. For the 47 students who completed 3 or more drawings, we tallied whether they drew the same type of drawing or switched between different types of drawings. There were 51 different patterns: 10 did not change their drawing type, 26 changed once, 11 changed twice, and 4 changed every time (the latter group includes students who changed their drawing twice and three times depending on whether they completed 3 or 4 assessments, respectively). Figure 5 shows the variability of drawing type over time for the 29 students who completed all four assessments. These data revealed the type of drawing students use for illustrating the process of evolution is dynamic, and more often changes (23 out of 29 students) than stays the same (6 out of 29 students).


A
Evolution Visualization:
I can't reach it!
over time
B
Visualization Assignment
·hats represent different individuals.
C
NEXT STOP EVOLUTION
D
Environmental pressures & mutations can lead to evolution & speciation
E
F
Bacteria
Cells
Sponge
Fish
Monkey
Human

G


Multicellular
Offspring
Single Cell
Offspring

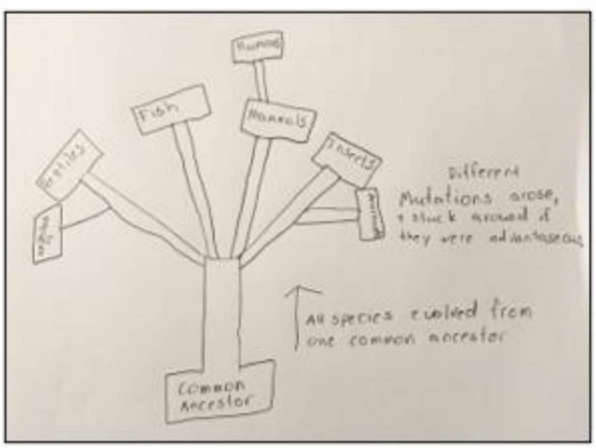
H



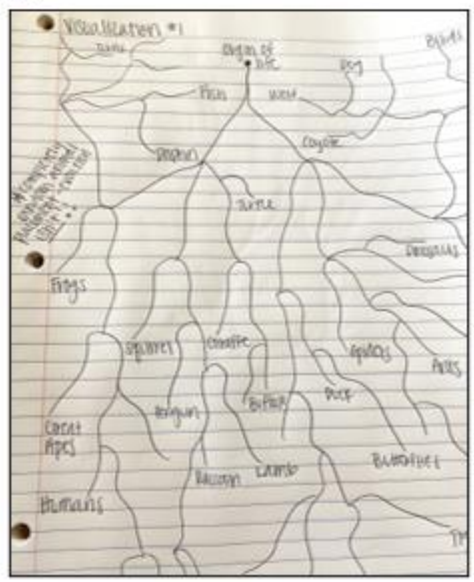
I


Eukaryotes Emerge
A complex web of species evolving from others that came before them
As you go organisms tend to become more complex
Bacteria
millions of years ago
Now

J


Fish
Mammals
Insects
Reptiles
Different Mutations arose, + stuck around if they were advantageous
All species evolved from one common ancestor
Common Ancestor

K


Visualization #1
Origin of life
Fish
Coyote
Turtle
Frogs
Giraffe
Spiders
Ants
Penguin
Duck
Great Apes
Humans
Lamb
Butterfly

L

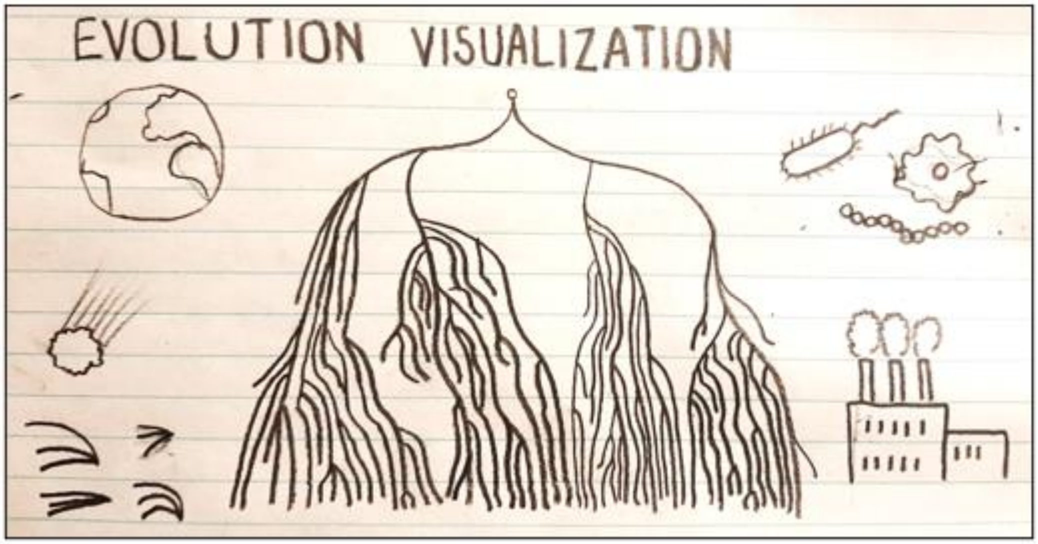


Figure 4. Examples of students' illustrations

Description: These were a few of the many different illustrations students produced in response to the assessment prompt. A-C and H can be categorized as scenarios, D-F as progressions, G as an abstract representation, and I-L as phylogenies.

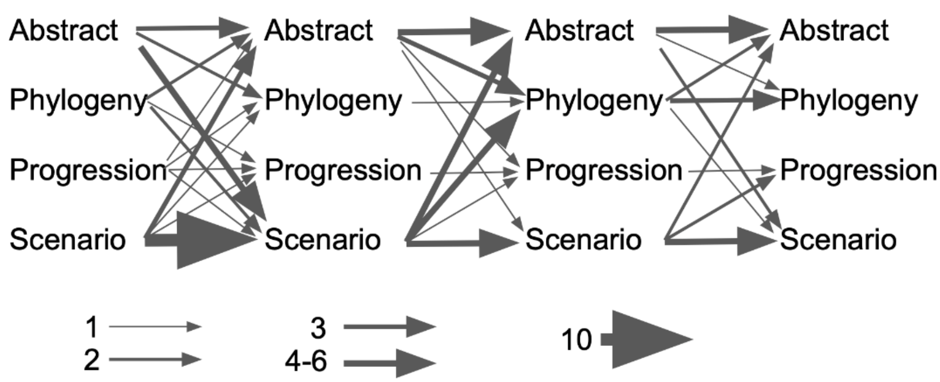


Figure 5. Change in the types of drawings across the four assessments

Description: Each column of drawing types represents a separate assessment. Arrows are proportion to the number of students. The data are from the 29 students who completed all 4 assessments.

There was little change in the mean summed rubric score across the four iterations of the assessments; however there was considerable variation in the individual summed scores (Figure 6). Additionally, there were many cases in which students' scores declined from one iteration to the next.

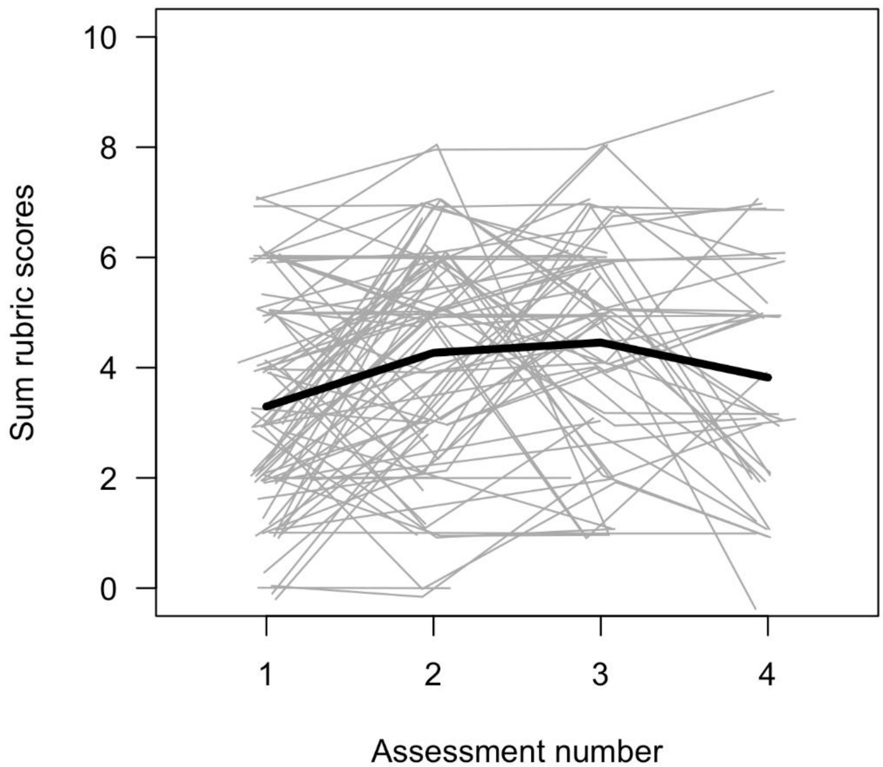


Figure 6. Scores for the four iterations of the assessment for each student (grey lines) showing variability over time in student understanding based on the sum of the rubric scores

Description: The data are the summed rubric scores for 107 students for each of the four assessments. A linear mixed model analysis, with student as a random variable, revealed there was a small, positive effect of iteration (0.24 ± 0.10, t = 2.29, intercept = 3.3). The solid black line shows the mean scores for each iteration.

Student competence varied across rubric items over the four assessment iterations (Figure 7). Rubric item 10 (change over time) was evident in ≈90% of students' first answers and the proportion of student answers with evidence of this rubric item declined over time to less than 75%. Some rubric item scores increased from assessment iteration 1 to 2 (e.g. 1, 2, 3, 4, 5 & 9) but failed to increase in subsequent assessment iterations. One item (item 4, heritable variation) was evident in less than 40% of student answers in the pre-assessment, the value increased to a maximum of approximately 75% in assessment iteration 3, and then decreased to the value evident in the pre-assessment. Reproduction in excess of carrying capacity (encoded by rubric items 6 & 7) was the least included concept and did not change appreciably over the semester. Finally, the only key idea to consistently increase throughout the iterations was reproduction (rubric item #5). In general, these data revealed ideas (pieces of knowledge) were dynamic over the course of the semester.

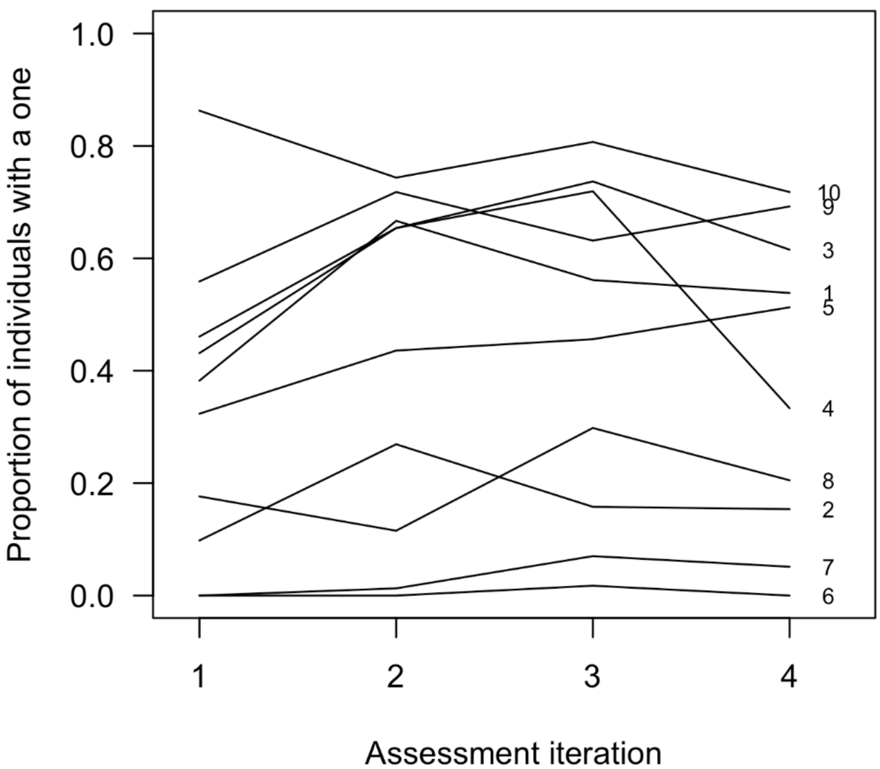


Figure 7. The proportion of students with a one for each of the ten rubric items across the four assessment iterations

Description: The thin lines show the proportion of students with evidence of competence for the ten rubric items.

We estimated the stability of concepts over time by counting the number of times a rubric item changed from 1 to 0 relative to the maximum score across all rubric items for each student (Figure 8). If students retained knowledge, there would have been no recorded losses of rubric items that were evident from earlier assessments and concept instability = 0. By contrast, if students retained none of the key ideas from one assessment to the next, concept instability is one. The data across all students and iterations indicated that, on average, about 37% of the rubric items were unstable (subject to loss).

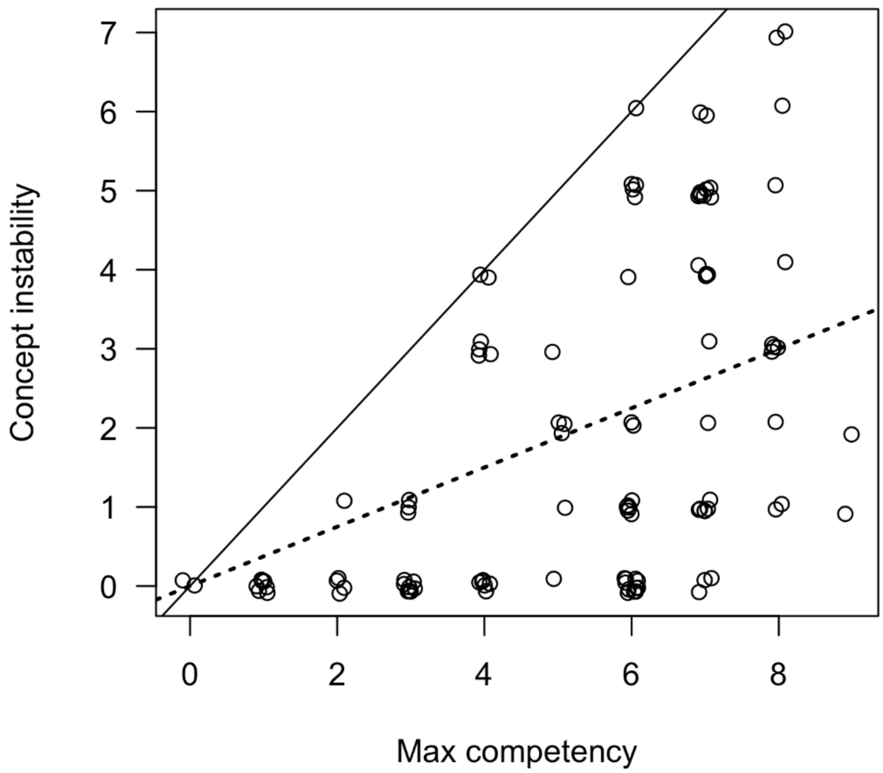


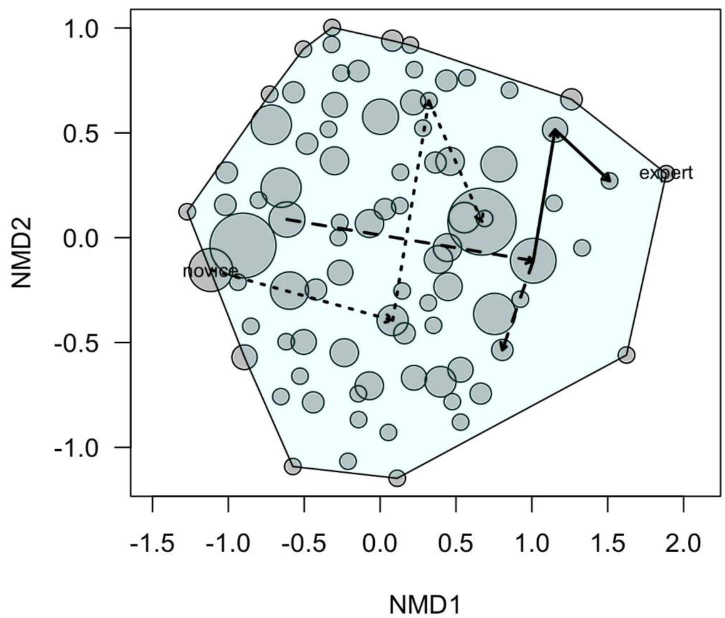
Figure 8. Estimate of concept instability relative to the maximum summed rubric score (max competency) for the 107 students

Description: Concept instability is the number of times a rubric item switched from 1 to 0 (an estimated loss of a key concept) from rubric scores across multiple iterations of the assessment relative to the sum of the number of rubric items present across all iterations for each student. Maximum competency is the maximum score across the ten rubric items for each student. Maximum instability (switching from a one to a zero) is indicated by the solid line. The dashed line is the predicted value of concept instability based on a linear model constrained with an intercept of zero (B1 = 0.37 ± 0.03). Points were jittered to better show the data. Note that the points with y = 0 indicate individuals that did not have any 1 to 0 reversals .

We evaluated whether the summed rubric score depended on the type of drawing. Overall, a mixed linear model revealed differences in the average summed score among the different drawing types, although the differences were small. The most notable difference was that progression tended to result in lower scores and scenarios with higher scores relative to the other visualization categories.

3.4 NMDS Analysis

There were 72 different answers based on the ten-item rubric from a total of 276 illustrations of evolution; these different answers formed 72 stepping stones in the hypothetical lake following the analogy of Stanley and Lehman’s (2015) model (Figure 9). We plotted the trajectories of selected students to illustrate the various and divergent pathways from novice to expert answers. Overall, student trajectories were rarely straight from novice towards expert. There were several students whose scores increased across most or all iterations (Figure 9, top); however, most students had trajectories that were tangential and often back-tracked, taking “stepping stones” towards where they started (Figure 9, bottom). Additionally, none of the 29 students who completed all four iterations took the same path “across the lake”.



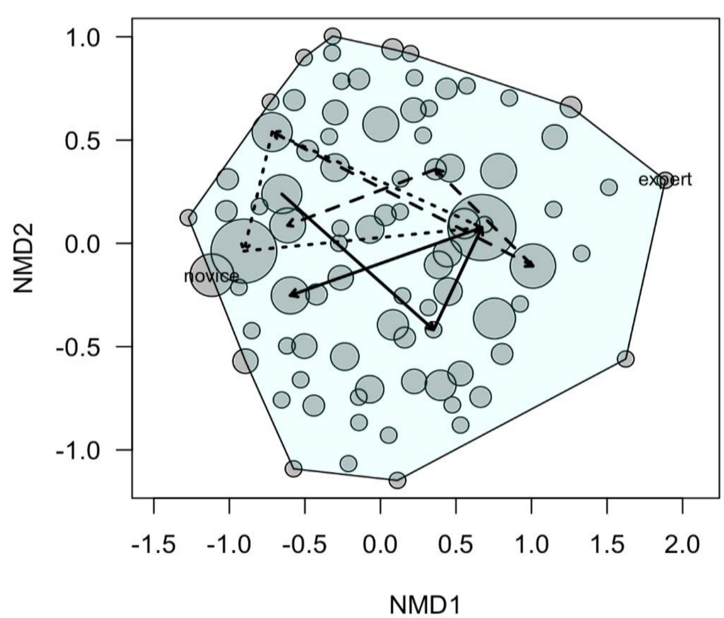


Figure 9. Examples of student trajectories using the “stepping stones to cross the lake” analogy

Description: For both graphs, the circles are different student answers based on the presence or absence of the 10 rubric items. The size of each circle is proportional to the square root of the number of students with the specific combination of rubric scores across the 10 rubric items. There were 72 different student answers (72 “stepping stones”) and 276 total answers across the four iterations. The novice and expert indicate scores of all 0s and all 1s, respectively. Top graph. Examples of the students with mostly positive gains across all four iterations Bottom. Examples of students that gained and lost rubric items illustrating knowledge is variable over time. The different line types represent different students. There are three students' trajectories in each picture.

3.5 Diversity of Student Answers

For each assessment iteration, we compiled composite binary scores for the 10 possible rubric items. There were 34, 38, 35, and 28 different combinations of rubric scores for the four ordered assessment iterations, respectively.For the pre-assessment, of the 34 different composite scores across the 102 students who completed the assessment, the most frequent score (14 instances out of 102 submissions) was one in which all rubric items were scored as absent (0) except for item 10 (change over time). The rank order of abundance of different composite scores was log-normal, similar to what is commonly observed for species in ecological communities. For the last assessment iteration, there were 28 different composite scores across the 39 students who completed the assignment. The most frequent score on the last assessment (6 instances) was one in which all rubric items were scored as absent (0) except for item 10 (change over time); this is the same as the most common composite score from the pre-assessment. Across all four assessment iterations, the diversity of student answers increased over time, in contrast to the general expectation that students would converge on a single correct answer.

4. Discussion

In this study, we implemented the same free-response assessment prompt four times over 15 weeks in two sections of an upper-division course focused on evolutionary biology to better understand the trajectories of student learning. The assessments were implemented independent of the instructor and course material as a means of objectively monitoring student understanding.

4.1 Adopting a Multivariate Perspective When Analyzing Assessment Data

Typical approaches for evaluating a free response assignment involve creating a rubric, scoring the work based on whether students demonstrated each rubric item, and generating a score based on a rubric sum. This results in a unidimensional ranked score. Graphical representations of student performance are often represented as a histogram showing the distribution of scores across all students from which the range and central tendency of the data are evident. While this serves the purpose of generating a quantitative basis for student grades, it obscures variation among individuals useful for data-driven revision of curriculum, teaching strategies, and intervention characteristic of effective teaching that align with the American Association for the Advancement of Science (AAAS) Vision and Change (2011).

In our study, there were many students who received the same score for vastly different answers. By using the composite binary presence and absence scores from the rubric items, we discovered 72 unique variations of student answers out of the 276 total scored. The diversity of student answers was visualized by calculating the distance between individuals based on the presence or absence of rubric data. The distance data were plotted in two dimensions such that the proximity of two points is indicative of similarity: the closer two points were, the greater the similarity based on the composite binary rubric scores. This method, referred to as non-metric multidimensional scaling (NMDS), is widely used in community ecology to describe diversity (Oksanen et al. 2013; Paulson et al. 2021). Education shares many essential features with ecology, including the abstract idea that each individual can be characterized by the presence and absence of particular rubric items in the same way ecological communities are characterized by the presence and absence of species. Additionally, we used a standard diversity index for characterizing ecological communities to estimate whether diversity of student thinking increased or decreased as they gained mastery of more key ideas and concepts. We discovered the student diversity of thinking steadily increased over time instead of converging on total concept mastery. Additionally, for the 29 individuals that completed all four iterations of the assessment, none of the pathways among the students was the same.

4.2 Iteration and Monitoring Dynamic Mental Constructs

There are a variety of models of learning and understanding. The knowledge-in-pieces (diSessa 1988) model posits that student knowledge concerning a singular phenomenon is a collection of mostly disconnected ideas and concepts. For novices, knowledge is often not organized hierarchically or structured; instead, the information is "flat" (Lira and Gardner 2020), like disconnected puzzle pieces spread out on a table. An important aspect of knowledge-in-pieces (also referred to as conceptual dynamics [Lira and Gardner 2020] and dynamic mental constructs [Sherin et al. 2012]) is that student knowledge is variable due to the inclusion and exclusion of ideas and concepts over time. Indeed, the term dynamic refers to the particular assembly of knowledge-in-pieces into a temporary structure composed of multiple ideas or concepts (Sherin et al. 2012). The extent knowledge exists as dynamic mental constructs (DMCs) can be revealed through iterative assessment involving the same or similar challenge (Lira and Gardner 2020). For students, iteration can be viewed as opportunities to expand and change answers in ways that reflect their changing understanding of the topic. For teachers, this can provide valuable insight into students' concept mastery, inform decisions on which concepts may require more classroom instruction to affect more complete and expert understanding, and provide information about the developing coherence of student knowledge in a particular discipline.

Students' visual depictions of evolution changed over time as a consequence of gains, losses, and recombination of represented ideas and concepts which matches the description of DMCs set forth by Sherin et al. (2012). The student-generated visualizations that we collected generally did not represent the entirety of student knowledge and suggest students need guidance in assembling and communicating a comprehensive and accurate mental model. This was evident in our study as most students failed to aggregate more than four of the ten key concepts laid out in our rubric during any given iteration. As such, while the students gained knowledge, their pathway to expert understanding seemed to change frequently and drastically. The dynamic nature of their apparent knowledge may be due to a combination of complex contextual factors such as their current level of understanding, harbored misconceptions, cognitive construals, recent classroom experiences, exposure to different types of illustrations, lack of motivation, or other unknown factors - all of which would be fascinating future explorations.

We chose the lake analogy developed by Stanley and Lehman (2015) as it provided a useful framework for visualizing the dynamic nature of student thinking captured in this study. Students' trajectories across the lake (from novice to expert) were highly variable, often included backtracking due to the loss of previously scored rubric items, and most students finished their education journey stranded somewhere in the middle of the lake without reaching or coming close to an expert answer. Thus, this approach successfully (and, in our opinion, effectively) revealed that knowledge is composed of dynamic mental constructs. Additionally, iteration revealed that, on average, students' overall scores, and by inference their understanding of evolution, improved but quickly reached a relatively low ceiling. In the context of the analogy of crossing the lake it seemed as if many students were lost in the mist. It would be useful to remove some of the mist so that students can see there are multiple pathways across the lake, that many of the different pathways are utilized, but that there are nonetheless few different expert answers such that there should be some convergence and ultimately a decline in the diversity of concepts contained within constructed responses. One way to achieve this in the classroom could be to ask students to evaluate multiple, different visual representations of evolution with a rubric and have students determine whether rubric items are evident. After explicit evaluation of visual representations, iteration becomes an opportunity for engaging in a core epistemic scientific process of revising and refining written and visual modes of communicating (Elliot 2012).

Typically, iteration of assessment questions is rarely implemented in undergraduate courses (although see Zingaro and Porter 2015), except when done for estimating learning gains using pre- and post-instruction assessments. We argue that a guided approach to repetition and revision could benefit students. If iteration has demonstrable benefits for learning and affective traits such as confidence, perhaps iteration should be explicitly built into the curriculum as a means for countering the instability of mental constructs. Additionally, iteration as a means of estimating learning gains can enable better data-driven revision of curricula and teaching strategies. Consequently, guided repetition and analysis could benefit both the student and teacher. Using iterative assignments and adapting the 'mist-shrouded lake' analogy with NMDS visualization can effectively reveal student thinking and provide a basis for supporting and challenging their thinking.

4.3 The Value and Challenge of Visualizing Student Thinking

Drawing is a process skill that is integral to the practice of science. It is used in the generation of hypotheses, the design of experiments, the visualization and interpretation of data, and the communication of results (e.g., Schwarz et al. 2009; Ainsworth and Scheiter 2021; Quillin and Thomas 2015). In education settings, drawings, or more generally, student-constructed visualizations, are "...models composed of multiple elements of abstractions of the real world…" (Quillin and Thomas 2015: 9). Quillin and Thomas (2015) noted visual literacy, namely the ability to represent thinking on a blank page, is necessary for successfully translating what students read and think about into symbolic, illustrative models. However, asking students to construct visual representations of their thinking may introduce cognitive dissonance and impede gains in understanding if the act of abstraction and illustration causes anxiety and excess cognitive load when the drawing does not accurately represent their thinking. The dissonance that may happen as a consequence of students' poor rendering of their imagination may be countered by scaffolded training in developing effective illustrations of phenomena. The "expert" illustration in figure 3 does not require artistry and should not result in the expected dissonance that may arise from the lack of realism in drawings due to poorly developed artistic capacity. Yet, it does require being comfortable constructing representations of abstract phenomena.

Challenging students to represent complex phenomena—like evolution by natural selection—and to translate their visualization into words engages students with multiple cognitive modalities that should improve learning and their capacity to communicate knowledge to others (Quillin and Thomas 2015, Jonassen et al. 2005). However, in this study, it was not evident that solely asking students to draw visual representations of their thinking about the process of evolution increased students' understanding of evolution or their ability for abstract thinking. Even though students were exposed to many examples designed to emphasize tree thinking, simply showing students Darwinian-like illustrations in class, during lectures, or with each change in context and information that was presented in class appeared insufficient for achieving robust and stable learning gains. Like most learned skills, effective representation of knowledge through drawings requires practice and should involve translation of verbal to visual information (Stern et al. 2003; Van Meter et al. 2006; Schwamborn et al. 2010), visual to visual information (Hegarty 2011) and visual to verbal (Schönborn and Anderson 2010).

4.4 Caveats and Limitations

This study happened during the spring semester of 2021 when instruction was fully remote using Zoom. There was considerable attrition in participation: the pre-assessment included 102 submissions, whereas the last (fourth) assessment near the end of class was completed by only 39 students. Furthermore, only 29 students completed all four assessments. Additionally, drop-out rates were higher than usual for this course. The decline in students' participation meant our estimates of gain in understanding about evolution were biased. We do not know how much of student work reflected a general Covid-exaggerated lack of motivation or accurately reflects understanding. One particularly noteworthy result was the attrition of students during the study, and the decline in participation differed markedly between the two separate sections of the course. These data suggest student-motivation differed markedly between the two sections presumably due to differences in the actions of the instructors.

We asked students to "Construct a unique and personalized visualization (not a Google image search result!) of your understanding of evolution and use the visualization to explain how evolution works. Try to depict evolution as thoroughly and completely as possible." Ideally, this assignment should have been completed in-person; however, limitations on interpersonal interaction due to COVID-19 meant the assessment was implemented online. Despite students having access to the internet, the majority of all collected images in this study were unique student creations, albeit there were some images copied directly from published resources. Similarly, we used a validated concept inventory (CI) for estimating student understanding and misconceptions about evolution (DeSaix et al. 2011). The pre-and-post-instruction CI scores revealed relatively large, positive gains. However, because the correct answers to the CI questions can be retrieved from internet searches, it is unclear whether the students' CI scores accurately represent their thinking or their ability to search and find answers on the internet. This remains a challenge for all published CIs when students can access the internet during their implementation.

These limitations notwithstanding, our results were not surprising because many introductory courses are designed to improve student performance on traditional multiple-choice format tests and focus less on alternative, verbal, or visual communication and modeling (Laverty et al. 2016). This is true despite increasing recognition that multiple-choice-question formats hinder critical thinking (Stanger-Hall 2017) and "...overestimate the ability of students to use their knowledge…" (Laverty et al. 2016: 5; see also Linn 1993). It is important to keep in mind that all assessments, including CIs, are designed "...to test the effectiveness of a particular pedagogy..." (Sands et al. 2018:180). Our cursory comparison of CI scores with students' representations of their understanding of evolution using illustration and text suggest existing evolution CIs are not designed to evaluate students' ability to integrate multiple concepts and engage in abstraction. Nehm and Schonfeld (2009: 360) wrote "Teachers should be less interested in tests that can only reveal isolated fragments of student thinking and be more interested in tests that can reveal how students choose to assemble and employ these elements in explanatory models." Our approach was a direct response to Nehm and Schonfield's recommendation: we allowed students to use illustration and text to convey their understanding of how evolution happens rather than relying on selected-response questions. We discovered that our free-response assessment was designed "...to test the effectiveness of a particular pedagogy…", one that emphasized higher-order thinking well above recognizing a single correct answer among familiar distractors. While we will continue to use selected-response questions to monitor gain in specific knowledge, our comparison between the CI scores and the richness of students' free responses to an open-ended question underscores the importance of developing a robust curriculum for revealing gain in explanatory models and abstract thinking and implementing assessments matched with the pedagogy.

References

Ainsworth, S. E., & Scheiter, K. (2021). Learning by drawing visual representations: Potential, purposes, and practical implications. *Current Directions in Psychological Science, 30*, 61-67. https://doi.org/10.1177/0963721420979582

Allmon, W. D. (2011). Why don't people think evolution is true? Implications for teaching, in and out of the classroom. *Evolution and Outreach, 4*, 648-665. https://doi.org/10.1007/s12052-011-0371-0

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: a call to action*. Derwood: American Association for the Advancement of Science. http://www.visionandchange.org/

Baum, D. A., Smith, S. D., & Donovan, S. S. (2005). The tree-thinking challenge. *Science, 310*, 979-980. https://doi.org/10.1126/science.1117727

BSCS. (2018). *Strategies for effective science teaching*. BSCS, Colorado Springs, CO.

Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. *Bioscience, 58*, 976-987. https://doi.org/10.1641/B581011

Coley, J. D., & Tanner, K. D. (2012). Common origins of diverse misconceptions: cognitive principles and the development of biology thinking. *CBE Life Sci Educ, 11*, 209-215. https://doi.org/10.1187/cbe.12-06-0074

Darwin, C. (1859). *On the origin of species by means of natural selection: Or the preservation of favoured races in the struggle for life 1st ed*. London: John Murray. http://darwin-online.org.uk/converted/pdf/1859_Origin_F373.pdf

DeSaix, J., Katcher, J., Urry, L., Young, C., Bridges, C., & Frost, J. (2011). Natural selection misconceptions diagnostic. https://evolution.berkeley.edu/evolibrary/teach/evo_misconceps_diagnostic.pdf. https://evolution.berkeley.edu/evolibrary/search/lessonsummary.php?&thisaudience=13-16&resource_id=426

DiSessa, A. A. (1988). *Knowledge in pieces*. In G. Forman & P. B. Pufall (Eds.), Constructivism in the computer age. Lawrence Erlbaum Associates, Inc;1988. p. 49-70. https://doi.org/10.4324/9780203771242

Elliot, K. C. (2012). Epistemic and methodological iteration in scientific research. *Stud Hist Philosophy Sci, 43*, 376-382. https://doi.org/10.1016/j.shpsa.2011.12.034

Evans, E. M. (2001). Cognitive and contextual factors in the emergence of diverse belief systems: creation versus evolution. *Cognitive Psychol, 42*, 217-266. https://doi.org/10.1006/cogp.2001.0749

Hegarty, M. (2011). The cognitive science of visual - spatial displays: Implications for design. *Topics in Cognitive Science, 3*, 446-474. https://doi.org/10.1111/j.1756-8765.2011.01150.x

Gould, S. J. (1989). *Wonderful life*. The Burgess shale and the nature of history. W. W. Norton & Co., New York and London.

Jonassen, D., Strobel, J., & Gottdenker, J. (2005). Model building for conceptual change. *Interact Learn Environ, 13*, 15-37. https://doi.org/10.1080/10494820500173292

Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., & Cooper, M. M. (2016). Characterizing college science assessments: The three-dimensional learning assessment protocol. *PloS One, 11*, e0162333. https://doi.org/10.1371/journal.pone.0162333

Linn, R. L. (1993). *Educational Measurement*. American Council on Education Series on Higher Education. Oryx Press, 4041 North Central at Indian School Road, Phoenix, AZ 85012-3397.

Lira, M., & Gardner, S. M. (2020). Leveraging multiple analytic frameworks to assess the stability of students' knowledge in physiology. *CBE Life Sci Educ, 19*, ar3. http://dx.doi.org/10.1187/cbe.18-08-0160

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *Irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84.1. https://CRAN.R-project.org/package=irr

Moharreri, K., Ha, M., & Nehm, R. H. (2014). EvoGrader: an online formative assessment tool for automatically evaluating written evolutionary explanations. *Education: Evolution and Outreach, 7*. https://doi.org/10.1186/s12052-014-0015-2

National Research Council. (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*. Washington DC: National Academies Press.

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *Bioscience, 57*, 263-272. http://dx.doi.org/10.1641/b570311

Nehm, R. H., & Schonfeld, I. S. (2009). *The future of natural selection knowledge measurement: A reply to Anderson et al. (2009)*. J Res Sci Teach. http://dx.doi.org/10.1002/tea.20330

Oksanen, J, Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara. R. B., Simpson, G. L., Solymos. P., Stevens. M. H., & Wagner H. (2013). *Community ecology package*. R package version. 2.0. http://CRAN.Rproject.org/package=vegan

Paulson, E. L., Chaudoin, A., & Martin, A. P. (2021). Ecological divergence of a habitat constructed to harbor an endangered species. *Conservation Science and Practice, 27,* e471. https://doi.org/10.1111/csp2.471

Pelaez, N. J., Boyd, D. D., Rojas, J. B., & Hoover ,M. A. (2005). Prevalence of blood circulation misconceptions among prospective elementary teachers. *Adv Physiol Educ, 29*, 172-181. https://doi.org/10.1152/advan.00022.2004

Quillin, K., & Thomas, S. (2015). Drawing-to-learn: a framework for using drawings to promote model-based reasoning in biology. *CBE Life Sci Educ, 14*(1), es2. https://doi.org/10.1187/cbe.14-08-0128

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Sands, D., Parker, M., Hedgeland, H., Jordan, S., & Galloway, R. (2018). Using concept inventories to measure understanding. *Higher Education Pedagogies, 3*, 173-182. https://doi.org/10.1080/23752696.2018.1433546

Schönborn, K. J., & Anderson, T. R. (2010). Bridging the educational research - teaching practice gap: Foundations for assessing and developing biochemistry students' visual literacy. *Biochemistry and molecular biology education, 38*, 347-354. https://doi.org/10.1002/bmb.20436

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology, 102*, 872-879. https://doi.org/10.1037/a0019640

Sherin, B. L., Krakowski, M., & Lee, V. R. (2012). Some assembly required: How scientific explanations are constructed during clinical interviews. *J Res Sci Teach, 49*, 166-198. https://doi.org/10.1002/tea.20455

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and outreach, 1*, 189-195. https://doi.org/10.1007/s12052-008-0037-8

Stanger-Hall, K. F. (2012). Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. *CBE-Life Sciences Education, 11*, 294-306. https://doi.org/10.1187/cbe.11-11-0100

Stanley KO, & Lehman J. (2015). *Why greatness cannot be planned: The myth of the objective*. 2015th ed. Cham, Switzerland: Springer International Publishing.

Stephens, S. (2012). From tree to map: Using cognitive learning theory to suggest alternative ways to visualize macroevolution. *Evolution: Education and Outreach, 5*, 603-618. https://doi.org/10.1007/s12052-012-0457-3

Stern, E., Aprea, C., & Ebner, H. G. (2003). Improving cross-content transfer in text processing by means of active graphical representation. *Learning and Instruction, 13*, 191-203. https://doi.org/10.1016/S0959-4752(02)00020-8

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology, 31*, 142-166. https://doi.org/10.1016/j.cedpsych.2005.04.001

Zingaro, D., & Porter, L. (2015). Tracking student learning from class to exam using isomorphic questions. *In Proceedings of the 46th ACM Technical Symposium on Computer Science Education Feb, 24*, 356-361. http://dx.doi.org/10.1145/2676723.2677239