Three-Dimensional Science Learning and Assessment in Biology

Mickey MacDonald
P. K. Yonge Developmental Research School

Abstract

The elimination of academic tracking alongside a move to institute standards-based assessment and proficiency-based learning led a biology teacher to adopt the national Next Generation Science Standards along with the three-dimensional learning framework for instruction and assessment which led to two questions of practice: (a) What is three-dimensional learning? and (b) How can three-dimensional learning be implemented within a high school biology course? Pre- and post-test data, and student scientific arguments were examined during the implementation of a pilot three-dimensional learning unit on evolution. Results from analyses of these data led the biology teacher to seek professional learning to deepen her understanding of three-dimensional learning and to develop scaffolds to support students in transitioning to this change in assessment practice.

Keywords: three-dimensional learning, biology, assessment, argument-driven inquiry, NGSS

Establishing the Problem of Practice

I currently teach honors biology to all incoming ninth-grade students at a suburban, public, K–12 school in Florida. The demographic breakdown of students within my five honors biology classes during the 2017–2018 school year was 43% White, 28% Black, 17% Hispanic, 7% Asian, and 6% Multiracial.

Over the past five years my ninth-grade colleagues and I have eliminated academic tracking in three of four ninth-grade core classes: English language arts, geometry, and biology. Academic tracking is the practice of sorting students into different leveled classes based on standardized test scores, perceived ability, work ethic, prior classroom achievement, teacher recommendations, IQ scores, and/or motivation (Burris & Garrity, 2008; NASSP, 2006; Oakes, 2005; Tyson, 2013). In many high schools, students are separated into either an honors level course or a general level course based on their historic performance on standardized tests and classroom performance. Additionally, placement into course tracks is confounded by previous teacher recommendations based on perceptions of students' work ethic and perceived academic ability. Instead of tracking, all ninth graders are enrolled in the one academic track we offer, the honors track. We, like Burris and Garrity (2008) and Oakes (2005) who oppose academic tracking as a pervasive practice that characterizes most secondary schools across America, have found that the elimination of academic tracking increases educational equity and reduces social stratification along racial and socioeconomic divisions.

As I approached how to meet the learning needs of my more diverse learners within my detracked biology honors classes, I incorporated differentiated instruction and blended learning. Differentiated instruction is a collection of best practices used to adjust content, process, and product in response to student learning needs. Through differentiating the resources students use to engage with content, the ways in which students access resources, and how students demonstrate mastery of the learning objectives, a teacher can match content and assessment with student readiness levels, their preferred modes of learning, and their interests (Anderson, 2007; Rock, Gregg, Ellis, & Gable, 2008; Tomlinson, 2004; Watts-Taffe et al., 2012; Wormeli, 2005). Riel, Lawless, and Brown (2016) describe blended learning pedagogies as the combination of face-to-face instruction involving interactions among students and teacher with online activities in which students interact with content outside of the typical scheduled school day. Such pedagogical approaches (face-to-face interactions and online applications) not only cultivate more student-centered learning activities but these pedagogies enhance the student school experience (Riel, Lawless, & Brown, 2016).

To ensure that I maintained high expectations for all learners I turned to standards-based grading using the Florida Next Generation Sunshine State Standards (FL NGSSS) for science within the course of biology honors as a guide to assess student understanding. McMillan (2009) defines standards-based grading (SBG) as a method of assessment that "compare[s] student performance to established levels of proficiency in knowledge, understanding, and skills" (p. 108). Although my students consistently performed much higher than the state average on the End of Course Biology exam developed by the Florida Department of Education, I continued to search for ways to make the learning of science more effective for all of my students.

At my school, we have been studying SBG through multiple cycles of practitioner research and simultaneously working toward implementation of proficiency-based learning. The Great Schools Partnership (2014) defines proficiency-based learning as "[S]ystems of instruction, assessment, grading, and academic reporting that are based on students demonstrating that they have learned the knowledge and skills they are expected to learn as they progress through their education" (para. 1). As I continued to delve into standards-based assessment and proficiency-based learning, I have begun assessing students using the national Next Generation Science Standards (NGSS), specifically the science and engineering practices. Eight science and engineering practices have been identified by the National Academies of Science, Engineering, and Medicine (NSTA, 2014). They include:

1. asking questions and defining problems,
2. developing and using models,
3. planning and carrying out investigations,
4. analyzing and interpreting data,
5. using mathematics and computational thinking,
6. constructing explanations and designing solutions,
7. engaging in argument from evidence, and
8. obtaining, evaluating, and communicating information.

Simply using these science and engineering practices as the basis for instruction and assessment was not enough to meet the state of Florida requirements for biology honors. However, teaching using the national NGSS requires that teachers use three-dimensional learning. Three-dimensional learning "allows students to actively engage with the [scientific and engineering] practices and apply the crosscutting concepts to deepen their understanding of [disciplinary] core ideas across science disciplines" (NGSS Lead States, 2013). By including the national NGSS Grades 9–12 disciplinary core idea standards for life science to both my instruction and assessment practice I was provided with a framework to meet the state of Florida NGSSS within the biology honors course.

The dilemma that I currently faced, however, was in both the design of the curriculum and the assessment of the national NGSS, which were written such that students demonstrate their content knowledge in the life science disciplinary core ideas through the science and engineering practices. First, in order for students to meet the expectations of the standards, they needed exposure to multiple opportunities to practice at the performance levels of these standards. My current curriculum was not designed in a way that provided students with multiple opportunities to practice meeting the high-level performance expectations to show mastery of these standards. Furtak (2017) describes how traditional instruction in science is designed and carried out:

> Traditionally, instruction chunked and sequenced content from simpler to more abstract; emphasis on vocabulary was a precursor to understanding more complex ideas and processes. In courses of study such as biology, traditionally students begin with simple macromolecules, cells, and cellular functions before learning about tissues, organs, organ systems, and finally interactions among organisms (ecology) and the formation of species (evolution). Vocabulary is front-loaded, decontextualizing science learning from any sense of purpose. The associated assessment privileges what students “take away” from instruction as stored-up knowledge, versus valuing engagement in scientific practice. (p. 857)

Furtak’s (2017) description mirrored my own teaching practice in both the sequence of topics (micro-level to macro-level) and in the frontloading of vocabulary, as well as my assessment practice, which had traditionally valued memorization of content knowledge that was then “churned-out” on a summative assessment rather than valuing engagement in the science and engineering practices.

Second, the assessments that could measure student progress toward meeting these three-dimensional learning standards do not currently exist and are not only difficult to create but once created, are very time intensive to assess. Furtak (2017) asserts that “assessments aligned with the NGSS will require multicomponent tasks that weave together different elements of performance expectations” (p. 859). Although I had a surface level understanding of NGSS three-dimensional learning, I did not feel comfortable enough in my understanding of three-dimensional learning to articulate how the Science and Engineering Practices, the Life Science Disciplinary Core Ideas, and the Crosscutting Concepts could be effectively integrated within my curriculum and I was not adept enough in this understanding to develop appropriate assessments to evaluate student progress toward meeting these three-dimensional learning standards. Therefore, the purpose of my research was to better understand what three-dimensional learning in science is and how it fit within my continued transformation of the biology honors course within a proficiency-based learning and assessment framework. My wonderings which guided my practitioner research study were two-fold:

1. What is three-dimensional learning?
2. How do I begin to implement three-dimensional learning and assessment within my biology course?

Inquiry Design

Dana and Yendol-Hoppey (2014) describe the intersection of the complex nature of teachers’ work within the classroom and the felt-difficulties within that work as the root of questions of practice or wonderings around which practitioner research takes place. As I continued to transform my Biology curriculum, the complex nature of three-dimensional learning and assessing three-dimensional learning bridged four of the eight passions that Dana and Yendol-Hoppey (2014) describe as triggers to the development of practitioner research questions: curriculum, content knowledge, teaching strategies/techniques, and social justice.

At the beginning of the 2017–2018 school year I elected to pilot standards-based grading on SKYWARD, our school’s student information system. Instead of using traditional grading categories (i.e., tests, quizzes, labs, homework, classwork) I selected the national NGSS Science and Engineering Practices as the overarching categories to guide student assessment. Within these practices, I broke apart the 9–12 life science disciplinary core ideas and added skills related to identifying and testing variables as I knew that students would be tested on these skills during the Biology End of Course exam. This task was my first foray in my understanding of three-dimensional learning.

Although somewhat unconventional, I enacted a pilot curriculum developed by the University of Utah's Learn Genetics Science Center during the first semester of 2017–2018 before I had a clear understanding of what three-dimensional learning was. This pilot curriculum, *Evolution: DNA and the Unity of Life*, was designed as a stand-alone three dimensional learning unit that integrated three science and engineering practices (engaging in argumentation, analyzing and interpreting data, and developing and/or using models), four crosscutting concepts (patterns, cause and effect, systems and system models, and structure and function), and five 9–12 life science disciplinary core ideas (shared biochemistry, common ancestry, heredity, natural selection, and speciation). Prior to enacting this curriculum, I had a limited understanding of three-dimensional learning. I was most familiar with the science and engineering practices and 9–12 life science disciplinary core ideas due to my limited work with setting up the standards to assess for standards-based grading. I had little background knowledge regarding the crosscutting concepts and their role in three-dimensional learning. During this pilot unit, I collected students' pre- and post-test scores to examine student learning growth. I examined a summative assessment on a written scientific argument on natural selection to look at trends in students' argumentative writing in science.

Due to its complexities, I wanted to learn more about three-dimensional learning and was motivated to find and/or design other curricular units devised to enact such learning. Likewise, I was eager to learn what kind of assessments existed for three-dimensional learning within a high school biology course. In order to develop a deeper understanding of three-dimensional learning, I attended a two-day pre-conference workshop, *Making Sense of Three-Dimensional Learning*, presented by the National Science Teachers Association. During this workshop, I kept a reflection journal in which I took notes on my evolving understanding of three-dimensional learning, made a list of ongoing questions that my deepening understanding elicited, and considered my plans for next steps in transforming biology honors using the three-dimensional learning framework.

My Learning

Three-dimensional learning in science is complex. To more effectively implement this instructional shift, I needed to deepen my understanding of three-dimensional learning. Initially, I was unaware of my limited knowledge about this pedagogical approach to learning science when I was selected to pilot the curriculum, *Evolution: DNA and the Unity of Life*. I became aware that I needed to learn more about three-dimensional learning during the implementation of the pilot curriculum. I quickly realized that this curriculum was like no other curriculum that I had used. First, the level of cohesiveness among the learning activities integrating the science and engineering practices with the life science disciplinary core ideas and the crosscutting concepts that run across all science disciplines was extremely high. Second, the authentic science data that ran throughout this curriculum not only pushed students to deepen their understanding of evolution, it caused them to think and act like scientists.

Making this instructional shift with students was and continues to be difficult as three-dimensional instruction and assessment are very different from what students have traditionally experienced in school. Rather than learning and assessing science vocabulary and discrete facts, the "stored-up knowledge" that Furtak (2017) refers to and what I had been assessing for nearly two decades, I was asking my students to, instead, construct scientific arguments using McNeill and Krajcik's (2012) Claim-Evidence-Reasoning (CER) instructional framework. Grooms, Enderle, and Sampson (2015) emphasize how success in the scientific practice of argumentation depends upon the ability of students to master many of the other scientific practices:

> …arguing from evidence (essential practice #7) also entails analyzing and interpreting data (essential practice #4)—i.e. constructing the evidence from which to argue—those data were likely gathered as a result of planning and carrying out an investigation (essential practice #3), and engaging in argument from evidence is itself a form of obtaining, evaluating, and communicating information (essential practice #8). (p. 46)

What appears as a simple instructional practice, i.e. constructing scientific arguments, is actually a significant shift in what we traditionally expect students to know and be able to do. Instead of memorizing vocabulary and biology facts and conducting a "cookbook" lab exercise in which students answer low-level questions, they are expected to assimilate multiple pieces of evidence from data they analyzed to support a scientific claim and subsequently use scientific principles to justify why their evidence supports their claim. Having students attend to the components of a scientific argument is a first step toward scaffolding argumentative writing in science.

As described by Knights-Bardsley and McNeill (2016), scientific argumentation includes three distinct components: claim, evidence, and scientific reasoning:

> The claim is an assertion that answers the question. Evidence is data, either student collected or from a secondary source, that is appropriate and sufficient to support the claim. Reasoning articulates how or why each piece of evidence supports the claim using appropriate scientific principles. (p. 650)

For less than five percent of my students, constructing a scientific argument with a claim, evidence, and scientific principles that link such evidences with the claim was a skill that they were able to achieve. This group of students already had a solid foundation in the other practices that underscore constructing scientific arguments. A written response illustrating one example of such a scientific argument is shown in Figure 1. Note that the claim, evidences, and reasoning are each highlighted.

This student was able to articulate a complete scientific argument and illustrated a clear understanding of how the scientific principles of variability, heritability, and reproductive advantage led to natural selection within this fish population. They intertwined scientific reasoning throughout the argument and provided concrete examples of evidence that supported their claim.

Conversely, a majority of my learners, even with what I perceived to be explicit guidance, were not able to write a complete scientific argument that synthesized multiple pieces of evidence to support a claim while simultaneously integrating scientific principles to illustrate why such evidences supported the claim. An example of such an argument is shown in Figure 2. Although the student did make a claim and provided some evidence to support the claim, the evidence was not sufficient to completely support the claim. Additionally, the student was not able to link the evidence with the claim using scientific reasoning as they did not include scientific principles in their argument although they did allude to both variability and reproductive advantage.

This student, like many of my students, was able to make a claim to answer the guiding question, "Is natural selection causing the lateral plate number in the population of sticklebacks in Loberg Lake to change over time?" However, this claim was not complete in that it did not indicate how natural selection was affecting the number of lateral plates in the freshwater stickleback fish population, only that the number was changing: "…natural selection in the lake is making the number of lateral plates on the stickleback to change" (Figure 2, para. 1). In the first argument, the student was very explicit in their claim when they asserted, "Natural selection caused this decrease in the lateral plate number of individuals in the stickleback population" (Figure 1, para. 1).

When analyzing the traits of organisms, it is often noticed that natural selection occurs between generations of the organisms. Natural selection is the process of a trait becoming more or less apparent within a population and uses the ingredients of variability, heritability, and reproductive advantage to do so. Natural selection has appeared in the Threespine Stickleback fish population of Loberg Lake. In 1990, many of the sticklebacks had a number of plates over 30. In 1996, the population was split, and some individuals had over 30 while others had less than 10. In 2002, most of the individuals in the population had less than 10 lateral plates. Natural selection caused this decrease in the lateral plate number of individuals in the stickleback population.

When determining variability within a population, scientists search for how a trait varies within a population. In the Threespine Stickleback fish population, the number of lateral plates varies between individuals. Such findings are presented in the amount of lateral plates within a sample size of 10 individuals, in which some individuals had less than others over the years. In 1990, eight out of ten fish had over 30 lateral plates, while two had 30 or less. In 1996, seven out of ten fish had less than 20 lateral plates, while three had 20 or more. In 2002, six out of ten fish had less than 10 lateral plates, while three out of ten had more than 20. Not only do these results prove how the number of lateral plates have changed over time, they also present how the number of lateral plates vary within the population within one occurrence of sampling. Because the lateral plates vary within the population within one occurrence of sampling, it can be proven that variability occurs within the Threespine Stickleback fish population of Loberg Lake.

Genes are formed through the combination of one allele from each parent, which presents offspring with a total of two alleles. Heritability is defined as how a trait is transferred between generations through these combinations. One way of proving that heritability can occur is through the comparison of the trait from parents to offspring, which in the stickleback population would be done through comparing the number of lateral plates. The mean number of plates was calculated between two parents and ten offspring separately. The mean number of plates for the parents was 34.5, while the mean number of plates for offspring was 33.7. Since these two numbers only have a difference of 0.8, it can be determined that there is a correlation between the trait from parents to offspring.

Following this investigation, a comparison of how the allele combinations of the parents affected the offspring's number of lateral plates. When there were insertions or deletions within both of the parents' sequences, the offspring had a low amount of lateral plates and was known as low plated. When only one parent had an insertion or deletion within their sequence, the offspring had a medium amount of plates and was known as partially plated. When there were no insertions or deletions within both of the parents' sequences, the offspring had a high amount of lateral plates and were known as completely plated. This data shows that the parents' DNA sequence affected the sequence of the offspring, thus providing evidence for heritability within the Threespine Stickleback population of Loberg Lake.

An individual is given a reproductive advantage when a variation of a trait provides the affected individuals with a more likely chance of producing offspring than those without that trait. When collecting evidence for whether or not any of the variations had a reproductive advantage, it was found that the fish who had a low amount of plates were better able to reproduce. The most compelling evidence for the low plated fish receiving a reproductive advantage came from how low plated sticklebacks grow larger more quickly than smaller fish do, thus making it harder for predators to catch them. Lateral plates are made for bone, and the creation less bone provides them with the ability to grow more quickly.

It has been proven that the number of lateral plates varies within the population and that the number of lateral plates is transferred from parents to offspring. It has also been proven that the fish with a low amount of lateral plates have been more successful at reproducing than those with a higher amount of plates. All of these findings present evidence for variability, heritability, and reproductive advantage to support that natural selection is occurring within the population. Because variability, heritability, and reproductive advantage were found within the population, it is safe to say that natural selection is causing the decrease in the number of lateral plates in the Threespine Stickleback population of Loberg Lake.

Figure 1: Complete scientific argument with the claim, evidences, and reasoning highlighted.

Is natural selection causing the lateral plate number in the population of sticklebacks in Loberg Lake to change over time? Yes i do think that natural selection in the lake is making the number of lateral plates on the stickleback to change. The population of sticklebacks that we are analyzing is in Loberg Lake it's a freshwater lake. So in freshwater lakes low-plated sticklebacks grow larger more quickly than completely plated sticklebacks. But in salt water, there is no difference in growth rate which gives the low plated sticklebacks a advantage because the can mature faster so they can get bigger and faster more quickly to avoid predators or get food and reproduce faster. Also the bigger fish are more likely to survive their first winter than the smaller fish so the low plated sticklebacks are more likely to survive their first winter than the full plated sticklebacks so they can survive and reproduce more. Plus Young sticklebacks with fewer lateral plates are faster and more nimble than their completely plated peer so they can get away from predators and survive so they can reproduce. In 1990 most of the sticklebacks in the lake where full plated or almost full plated sticklebacks in 1996 the full plated sticklebacks went to medium to high then in 2002 they went to low.

Figure 2: Incomplete scientific argument with the claim and evidences highlighted; many students, like this one, failed to include reasoning.

Although many students had not yet met the standard for constructing a complete scientific argument, most students scored approaching proficiency using the standards-based grading rubric for writing a scientific argument. When evaluating students' arguments, I used a Claims-Evidence-Reasoning rubric developed by Kevin Anderson that was adapted from the work of McNeill and Krajcik (2012). The two components of scientific argumentation in which most students needed to improve were evidence and reasoning. Although students described the evidences, these descriptions were often vague and did not include enough detail from the data they had collected. For example, the student in Figure 2 states:

> In 1990 most of the sticklebacks in the lake where full plated or almost full plated sticklebacks in 1996 the full plated sticklebacks went to medium to high then in 2002 they went to low. (para.1)

This student did not specify the number of lateral plates in each of the generations of fish and was not clear in the language they used to describe how the number of lateral plates changed over these generations. The student in Figure 1, on the other hand, provided an evidence statement that was very specific and much clearer in the language they chose:

> In 1990, eight out of ten fish had over 30 lateral plates, while two had 30 or less. In 1996, seven out of ten fish had less than 20 lateral plates, while three had 20 or more. In 2002, six out of ten fish had less than 10 lateral plates, while three out of ten had more than 20. (para. 2)

Scientific reasoning, the ability to explain why the evidences described in the argument support the claim using appropriate scientific principles, was the area in which students most struggled, even students who were able to provide complete evidences and claims. The literature reveals that even students who are able to provide a solid description of the evidences that support the claim are often not able to provide the "warrants" or scientific reasoning that can then justify their choice of the evidences used (Bell & Linn, 2000; Erduran, Simon, & Osborne, 2004; Jimenez-Aleixandre, Rodriguez, & Duschl, 2000). This finding was confirmed by Sampson, Grooms, and Walker (2011). In their development of the Argument Driven Inquiry model which provides an eight-step framework for students to engage in scientific argumentation, Sampson et al. (2011) found that when students engaged in argumentative discourse across an entire semester, they began to understand what counted as scientific reasoning and more consistently demonstrated such justification in their scientific arguments.

Not only was I interested in trends that I saw emerge by examining students' written scientific arguments, I also wanted to see how much growth in student understanding of evolution was made following the implementation of this three-dimensional learning pilot curriculum on evolution. A test developed by the National Science Foundation grant was administered both as a pre-test and a post-test. This test included multiple-choice questions in which students read and/or interpreted data as well as extended-response questions performing the same types of tasks in addition to constructing components of the scientific argument. The students scored a mean pre-test score of 45% and a mean post-test score of 62% on the multiple-choice portion of the assessment. Of the 118 students with both a paired pre- and post-test score, 80% of the students increased their post-test score while 20% of the students scored the same or less on the post-test. Table 1 summarizes the pre- and post-test mean, standard deviation, and standard error of the mean.

Table 1: *Mean, Standard Deviation, and Standard Error Mean for Pre- and Post-test Scores*

	N	***M***	***SD***	***SEM***
Pre-test	118	0.448	0.170	0.057
Post-test	118	0.616	0.200	0.018

A paired-samples *t*-test was conducted to compare the pre-test scores with the post-test scores to indicate if the difference in the mean score was significant. As can be seen in Table 2 there was a significant difference in the scores for the pre-test (M = 0.45, SD = 0.616) and post-test (M = 0.62, SD = 0.20); t (117) = 11.02, p = 0.05.

Table 2: *Pre- and Post-test Paired t-test Statistics*

	t	***df***	***p* (2-tailed)**	**Difference**
Difference Post – Pre	11.31	117	< 0.00001	0.17

These results suggest that the three-dimensional learning curricula, *Evolution: DNA and the Unity of Life*, did affect students' post-test scores. Specifically, these results suggest that when students engaged in this three-dimensional learning curriculum, they increased their understanding of evolution. To ensure that this method and its results are valid, the difference between the pre-test and post-test scores should be normally distributed. I used the Shapiro-Wilk's measure to test for normality. As the results in Figure 3 illustrate, the distribution of scores are normally distributed and thus, engagement in the three-dimensional learning pilot curricula on evolution contributed to an increase in student post-test scores. [i]

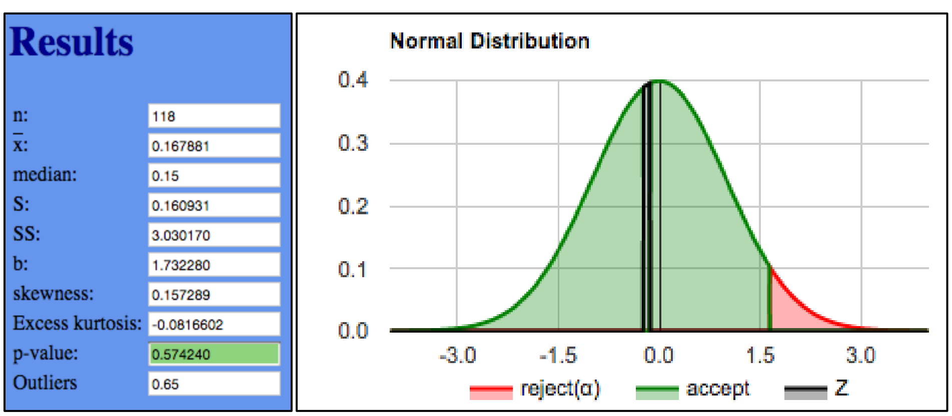

Figure 3: Shapiro-Wilk test for normality in pre- and post-test scores.

Increased scores are not surprising, however, for students who were provided instruction between the pre- and post-assessment. To evaluate if this difference in post-test scores could be attributed to engagement in the three-dimensional learning unit, I contacted the researchers conducting this National Science Foundation study. The researchers disclosed that a 17% mean increase in the post-test scores of my students who engaged in the pilot curriculum was significant compared to the increase in students' post-test scores for control classrooms whose teachers implemented their traditional evolution curriculum with preliminary data analysis indicating a large effect size. Thus, the increase in average post-test scores for my students could likely be attributed to the three-dimensional learning curriculum.

After having completed the three-dimensional learning pilot unit with my students, I was much more aware of where I lacked understanding with this new approach to science learning and assessment. To address my need to better understand three-dimensional learning, I attended a two-day workshop, Making Sense of Three-Dimensional Teaching and Learning. During this workshop, I captured how I was making sense of three-dimensional learning, my unanswered questions regarding this instructional approach, and ideas that I had as I began to think about how I could continue to transform my biology honors course using this framework.

When first introduced to three-dimensional learning during this workshop, the presenters had the participants engage in an actual unit. Initially, we were provided with a guiding question, "How can we sense so many different sounds from a distance?" and we observed a perplexing phenomenon, in this case, a homemade vinyl record player using a sewing needle and a paper cone as seen in Figure 4.

Figure 4: Homemade record player—anchor phenomenon.

As we observed the record spinning, we were asked to write down observations only. After sharing our observations and questioning one another about our observations, we generated questions for further investigation. During each investigation, after debriefing our observations and challenging each other's ideas, we would begin to piece together and modify a model that represented our learning up to that point in the investigation. Each model we developed needed to provide insight into our guiding question. This repeated process is what is referred to as a "storyline." According to the Next Generation Science Storyline design team, a storyline is:

> …a coherent sequence of lessons, in which each step is driven by students' questions that arise from their interactions with phenomena. A student's goal should always be to explain a phenomenon or solve a problem. At each step, students make progress on the classroom's questions through science and engineering practices, to figure out a piece of

> a science idea. Each piece they figure out adds to the developing explanation, model, or designed solution. Each step may also generate questions that lead to the next step in the storyline. Together, what students figure out helps explain the unit's phenomena or solve the problems they have identified. A storyline provides a coherent path toward building disciplinary core ideas and crosscutting concepts, piece by piece, anchored in students' own questions. (n.d., para. 2)

When I was presented with this idea of storylines, I immediately asked myself where they existed and why I hadn't been aware of them before this workshop. When I began to look for these storylines, I was directed to the Next Generation Science Storylines website, http://www.nextgenstorylines.org/. On this site, I found two completed high school level storylines, one aligned with a life science disciplinary core idea and one aligned with two physical science disciplinary core ideas. I also found three additional storylines under construction with an anticipated release date of summer 2018. Of these three storylines, two address the same life science disciplinary core idea and the third storyline addresses a physical science and an earth science disciplinary core idea.

The number of storylines that exist are extremely limited and the development of these storylines require the collaboration of many stakeholders including research scientists, classroom teachers, and curriculum specialists confirming what I suspected based on what I knew about the educators and scientists involved with the development of the *Evolution: DNA and the Unity of Life* three-dimensional learning pilot curriculum. Teachers in isolation cannot create three-dimensional learning units that are well-aligned to the national NGSS, coherent, and scientifically accurate.

Carlson, Davis, and Buxton (2014) and Anderson et al. (2018) affirm that as a nation, we are lacking national NGSS aligned curricula. "In order to meet the goals of the NGSS, teachers need flexible and learning progression-based curricular resources aligned to the three-dimensions of the standards" (Anderson et al, 2018, p. 11).

Implications for Practice

As with any practitioner research study, the final step of the research process involves an implications' stage in which the practitioner researchers take action based on what they have learned from their research study and ultimately, take action to transform the larger teaching profession (Dana & Yendol-Hoppey, 2014). Each summer I reflect on what I have learned from my previous year's cycle of practitioner research to inform my planning as I move into the next school year.

The immediate action that I took occurred within my biology honors course. I reworked student assessment using the CER framework as the way in which students would demonstrate meeting proficiency of the NGSS science and engineering practices to convey their understanding of the life science disciplinary core ideas. I developed instructional scaffolds in each of the three components of the Claim-Evidence-Reasoning (CER) framework to support students in meeting the underlying skills necessary to construct a coherent scientific argument. Besides providing the scaffolds that learners could use as they constructed scientific arguments, I created a set of exemplars and non-exemplars for students to evaluate using the CER rubric as well as for students to reflect on what makes the exemplars "exemplary" and what could be revised with the non-exemplars so they meet proficiency of all components of the scientific argument.

In addition to taking action within my own classroom, I also used my role as a teacher leader to guide the direction of our professional learning for the 2018–2019 school year. As more secondary teachers transform their instruction and assessment practice to align with the principles of standards-based grading and begin to report student progress using the standards-based side of SKYWARD, it is imperative that we maintain high expectations for all of our students. I believe that one way to help faculty maintain high expectations for all students is to critically examine student work together. Because we already had dedicated professional learning time built into our weekly schedule, I lobbied for using some of that time to reestablish Critical Friends' Groups. Critical Friends' Groups or CFGs developed out of the work of the Annenberg Institute for School Reform at Brown University in the mid-1990s. A CFG consists of 8–12 educators who meet periodically to discuss issues of teacher practice and student learning (Moore & Carter-Hicks, 2014). In the mid-2000s, our secondary teachers participated in CFGs over a course of two years. As a member and facilitator for CFGs at the time, I found great value in sharing my student work with faculty and using protocols to drive discourse. I believed that as a secondary faculty we could use CFG time to examine our assessment practice with a focus on student work across grade levels and disciplines.

Having completed this cycle of practitioner research, I do feel more informed about what three-dimensional learning is and how to begin to implement it within my biology course. I know that using this three-dimensional learning framework and assessing the national NGSS will challenge both my students and me. However, I believe that it is the next step that I need to take as I strive to increase my students' scientific literacy. I also know that as our entire school begins to enact standards-based grading, working collaboratively will be key to our collective success. In the spirit of the "critical" portion of the CFG, "critical" does not refer to critique of work, but rather how others are critical or vital in our own learning (Moore & Carter-Hicks, 2014).

References

Corresponding Author: Mickey MacDonald
Author Contact Information: mmacdonald@pky.ufl.edu

Anderson, K. M. (2007). Differentiating instruction to include all students. *Preventing School Failure, 51*(3), 49–54.

Anderson, C. W., de los Santos, E. X., Bodbyl Roels, S., Covitt, B., Edwards, K. D., Hancock, J. B., Lin, Q., Morrison Thomas, C., Penuel, W., Welch, M. (In press, 2018). Designing educational systems to support enactment of the *Next Generation Science Standards.* Submitted to the *Journal of Research in Science Teaching.*

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22*(8), 797–818.

Burris, C. C. & Garrity, D. T. (2008). *Detracking for excellence and equity*. Alexandria, VA: Association for Supervision and Curriculum Development.

Carlson, J., Davis, E. A., & Buxton, C. (2014). *Supporting the implementation of the Next Generation Science Standards (NGSS) through research: Curriculum materials.* Retrieved from https://narst.org/ngsspapers/curriculum.cfm

Dana, N. F. & Yendol-Hoppey, D. (2014). *The reflective educator's guide to classroom research: Learning to teach and teaching to learn through practitioner inquiry, 3rd Edition.* Thousand Oaks, CA: Corwin Press.Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*, 915–933.

Furtak, E. M. (2017). Confronting dilemmas posed by three-dimensional classroom assessment: Introduction to a virtual issue of Science Education. *Science Education, 101*, 854–867, https://doi.org/10.1002/sce.21283

Great Schools Partnership (2014). The glossary of education reform. Retrieved from https://www.edglossary.org/

Grooms, J., Enderle, P., & Sampson, V. (2015). Coordinating scientific argumentation and the Next Generation Science Standards through Argument Driven Inquiry. *Science Educator, 24*(1), 45–50.

Jimenez-Aleixandre, M., Rodriguez, M., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education, 84*(6), 757–792

Knights-Bardsley, A. & McNeill, K. L. (2016). Teachers' pedagogical design capacity for scientific argumentation. *Science Education, 100*(4), 645–672,

McMillan, J. H. (2009). Synthesis of issues and implications for practice. In T. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading*, Thousand Oaks, CA: Corwin Press.

McNeill, K. L., & Krajcik, J. (2012). *Claim, evidence and reasoning: Supporting grade 5–8 students in constructing scientific explanations*. New York, NY: Pearson Allyn & Bacon.

Moore, J. & Carter-Hicks, J. (2014). Let's talk! Facilitating a faculty learning community using a critical friends group approach. *International Journal for the Scholarship of Teaching and Learning, 8*(2), 1–17.

NASSP (2006, July 13). Tracking and ability grouping in middle level and high schools. Retrieved from http://www.nassp.org/Content.aspx?topic=53615.

Next Generation Science Storylines (n.d.) *What are storylines? Next generation science storylines*. Retrieved from http://www.nextgenstorylines.org/what-are-storylines/

NGSS Lead States (2013). *Next generation science standards: For states, by state*s. Washington, DC: The National Academies Press

NSTA (2014). *Science and Engineering Practices*. Retrieved from https://ngss.nsta.org/PracticesFull.aspx

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd Ed). New Haven, CT: Yale University Press.

Riel, K., Lawless, K.A., & Brown, S. W. (2016). Listening to the teachers: Using weekly online teacher logs for ROPD to identify teachers' persistent challenges when implementing a blended learning curriculum. *Journal of Online Learning Research*, *2*(2), 169–200.

Rock, M. L., Gregg, M., Ellis, E., & Gable, R. A. (2008). REACH: A framework for differentiating classroom instruction. *Preventing School Failure, 52*(2), 31–47.

Sampson, V., Grooms, J., & Walker, J. (2011). Argument-Driven Inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: An exploratory study. *Science Education*, *95*(2), 217–257.

Tomlinson, C. A. (2004). Sharing responsibility for differentiating instruction. *Roeper Review, 26*(4), 188–189.

Tyson, K. (2013). Tracking, segregation, and the opportunity gap: What we know and why it matters. In P. Carter & K. Welner (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 169–180). New York, NY: Oxford University Press.

Watts-Taffe, S., Laster, B. P., Broach, L., Marinak, B. A., McDonald Connor, C., & Walker-Dalhouse, D. (2012). Differentiated instruction: Making informed teacher decisions. *The Reading Teacher, 66*(4), 303–314. doi: 10.1002/TRTR.01126

Wormeli, R. (2005). Busting myths about differentiated instruction. *Principal Leadership: Middle Level Edition*, *5*(7), 28–33.

---

[i] I want to acknowledge Dr. Jose De Leon, P.K. Yonge Statistics Instructor who helped me with the quantitative analyses of the pre- and post-test scores.