Building Science Mastery with BrainPOP®:

Three Dimensional Research-Based Learning Unlocked

Co-authors:
Yigal Rosen, Ph.D.
Laurel Ozersky, M.A.
Melissa Whitehead, Ed.D.
Ilia Rushkin, Ph.D.
Maneeza Dawood, Ph.D.
Elizabeth O'Riordan, B.A.

Certified by Digital Promised as Research-Based Designed Product[1]

BrainPOP Science

---

[1] BrainPOP® has earned the Research-Based Design product certification from Digital Promise in August 2020. The new product certification is intended to serve as a rigorous, reliable signal for consumers, including school administrators, educators, and families, looking for evidence of research-based educational technology products. BrainPOP® submitted evidence to Digital Promise confirming a link between research on how students learn and the design of our products (Hubert, & Rosen, 2020).

Introduction

Advances in science, technology, and engineering, as well as changes in the science education landscape, create opportunities to engage the new generation of scientific thinkers in more powerful real-world learning experiences. Becoming scientifically competent embodies the idea that the objectives of science education should be both broad and applied. Thus, within this framework, BrainPOP Science provides an opportunity for educators to build student mastery through three-dimensional learning that refers both to a knowledge of science, cross-cutting concepts, and scientific and engineering practices. Research-based scientific inquiry introduced by BrainPOP involves scaffolded interactive investigations, embedded assessments, and actionable diagnostic tools designed to inform differentiated learning and instruction.

In this paper, we provide an overview of the key learning and assessment design principles that guided the development of BrainPOP Science and supporting evidence from efficacy studies conducted over the years in collaboration with researchers, educators, and students.

BrainPOP is currently one of the most widely used science resources in US schools

(usage between August 2020 to March 31, 2021)

3,108,916 students

have completed one or more Science activity (median activities = 4)

149,117 teachers

have completed one or more Science activity

77,703 teachers

have created one or more assignment with Science activity

Five most popular topics in Science: Scientific Method, States of Matter, Ecosystems, Water Cycle, and Forces

Changing Landscape of Science Education

The landscape of education has gone through a major transformation for K-12 science and has fundamentally changed its focus. Historically, K-12 science standards and instruction were focused on students memorizing facts about science and the physical act of doing science was minimal. A growing body of evidence in science learning has shown that engaging students in scientific inquiry is critical for the development of scientific competency (National Academy of Science, 1995; 2018; National Research Council, 2006; OECD, 2017). Scientific competency requires not just a knowledge of the concepts and theories of science, but also a knowledge of the common processes and practices associated with scientific inquiry. Scientifically competent students have a deep understanding of the major conceptions and ideas that form the foundation of scientific and engineering thought and the degree to which such knowledge is supported by evidence. The role of science education is to prepare students with core knowledge that they can build upon, on their own, in the future. Preparing our students with a set of core ideas and practices allows them to continue their development as scientific learners, users of scientific knowledge, and possibly as producers of new knowledge and new scientific practices.

This major transformation in science has shifted the role of both teachers and students. Instead of learning through teacher lectures, students are engaging in sense-making and the cycle of inquiry through exploring and discussing phenomena in the natural world. The separation of science content and the physical act of doing science is in the past, and students are now engaging simultaneously with the three dimensions outlined in the Framework: Science and Engineering Practices, Disciplinary Core Ideas and Crosscutting Concepts (National Research Council, 2012). The integration of the three dimensions into science curriculum, instruction and assessment are crucial to this necessary shift to science education. Next generation science requires students to have a sense of context understanding with regard to scientific knowledge, how it is acquired and applied, and how science is connected through a series of concepts that further the understanding of the natural world. The introduction of a Framework for K-12 Science Education, the Next Generation Science Standards and state standards consistent with the Framework has provided a structure for rethinking how students engage in science and engineering and how they can use investigations/design to engage deeply with phenomena, gather and analyze data to support explanations of the causes of phenomena and to design solutions. It has moved teaching science away from covering many isolated facts toward helping students build a rich network of connected ideas. This network serves as a conceptual tool for explaining phenomena, solving problems, making decisions, and acquiring new ideas. In addition to content knowledge, both procedural and epistemic knowledge are necessary to identify questions that are critical to scientific inquiry, to judge whether appropriate procedures have been used to ensure that the claims are justified.

Investigations provide the evidence that students need to construct explanations for the causes of phenomena. Constructing understanding by actively engaging in investigation and design also creates meaningful and memorable learning experiences for all students. These experiences pique students' curiosity and lead to greater interest and identity in science (Baldeh, Dawood & Purdie-Greenaway, 2018). When investigation and design are at the center of learning, students can gather evidence and take ownership of the evidence they have gathered (Rosen et al., 2020). This process contributes to student agency as they make sense of phenomena and designs and extend their understanding of the natural and designed world.

Scientific argumentation is an important competency for scientifically literate laypeople as well as scientists. This competency requires not only understanding how data or evidence could support a specific claim or conclusion but also evaluating multiple possible interpretations to devise a compelling argument for one explanation over another. Engaging in scientific argumentation involves combining content, procedural, and epistemic knowledge, and supports both learning new scientific knowledge and participating in scientific debates (Osborne et al., 2016). Mechanistic reasoning is an important component of scientific argumentation (Odden, & Russ, 2019), requiring thinking about the potential cause of a phenomenon. Rather than stating a description of the events that took place, mechanistic thinking involves following the causal chain of events that give rise to the observations.

Incorporating modeling into science education is an effective method to teach students about the nature of science (Berland et al., 2016). Students who understand scientific models know that models are a representation or an abstraction of a phenomenon that details the hypothesized reason for why a phenomenon occurs (Namdar & Shen, 2015). Models engage students' mechanistic reasoning by encouraging them to articulate explicitly their hypothesized causal chain that leads to the phenomenon (Krist et al., 2019). Additionally, by being simplified representations of the phenomenon at hand, students need to use systems thinking to identify the most relevant factors among a myriad of potential variables (Yoon et al., 2018). The ubiquity of scientific models in all fields of science also makes them an important cross-cutting skill, as it might help students integrate different topics in science (Park et al., 2017).

BrainPOP Science

BrainPOP Science is designed to help 6-8th grade science teachers navigate changes to standards, assessments, and learning environments. Scaffolded investigations, embedded interactive tools, and actionable diagnostics are designed for three-dimensional learning, drive student mastery, and inspire the next generation of scientific thinkers. Each investigation covers up to one week of instruction and includes key activities such as examining phenomena, collecting evidence, articulating claims and reasoning, and checking for understanding along the way. The investigations provide opportunities for students to interact with data manipulatives, simulations, 3D Worlds, exclusive science movies, and checks-for-understanding. When used in tandem, every investigation aligns with the 5Es of Science Instruction: Engage, Explore, Explain, Elaborate and Evaluate (Carin, Bass & Contant, 2003).

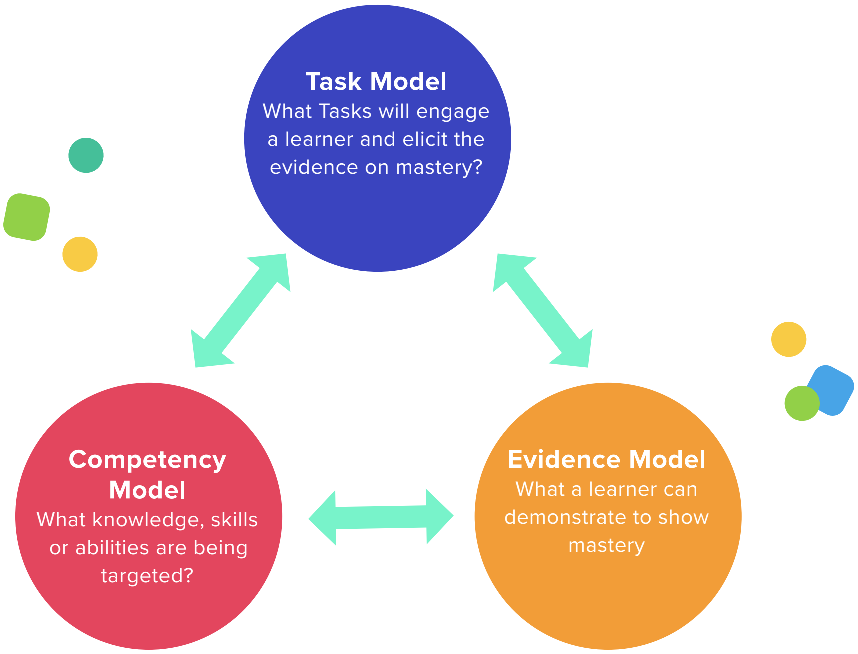
Design for Deeper Learning and Formative Assessment

BrainPOP Science content, tools, and educator support resources are guided by learning science research on how to effectively meet the needs of a diversity of learners. Online learning platform design must also center learning science research on cognitive and affective learning outcomes, like the development of transferable thinking skills.

As our guiding design framework for the development of BrainPOP Science we used a combined Evidence Centered Design (ECD) and Universal Design (UD) approach. The integration and application of ECD+UD increases the likelihood that learning and assessment activities will be well-aligned with the competency model while also being accessible to all students, including students with disabilities. The advantage of following an integrated ECD+UD-principled design is particularly evident when the goal is to develop and assess complex competencies such as scientific argumentation and modeling by using complex performance tasks (e.g., multidimensional tasks such as simulations or games). It is important to explicitly identify how the relevant competencies and behaviors are connected because the complexity of the focal competencies and/or the rich data the tasks provide might pose difficulties in making inferences from behaviors to competencies. ECD formulates the process of development to ensure consideration and collection of validity evidence from the onset of the learning and assessment design (Arieli-Attali et al., 2019; Kim, Almond, & Shute, 2016; Mislevy et al., 2006). ECD is built on the premise that a learning and assessment system with specific claims about the associated evidence and that a good learning and assessment system is a good match of the activities and students' knowledge and skills. The ECD Framework defines several interconnected models, including competency or student model, evidence model, task model, and assembly model. Using ECD as an organizing framework for learning and assessment can help address important design questions: What are the relationships between learning and assessment within the system? Which constructs or processes does each task reveal? Do the proposed scoring methods effectively recognize and interpret the evidence generated by students' responses and interactions? How is all of the evidence that is generated by students' choices synthesized across multiple tasks? Is all the evidence for a particular construct comparable when different students attempt different tasks? How do learning and assessment relate within an integrated system?

Measurement models, combined with the learning sciences perspective in the development of assessment tasks and items, provide the essential foundation for the BrainPOP Science approach to assessment design. The latent competencies that are articulated and defined in the NGSS and state standards establish the conceptual basis of the assessments; based on a theory or previous empirical research related to assessment goals (e.g., valid and reliable measure of students' competencies. Since we cannot tap directly into the latent competencies, we designed tasks and items to elicit behaviors that can reflect on or provide indications about the latent competencies. The task/item model specifies the task features that are supposed to elicit the observables, and only them, in order to allow inferences about the latent competencies.

Students' responses to the assessment tasks and items provide the evidence for this reasoning process, and psychometric analyses establish the sufficiency of the evidence for evaluating each claim as illustrated in the following figure (Rosen, 2013; Rosen, Ferrara, & Mosharraf, 2015; Rosen et al., 2020).



Adopting the ECD process for the Science learning and assessment requires the following sequence of steps:

1. Domain analysis: reviewing NGSS and Science state standards and engaging with experts to define measurable knowledge, skills and attributes in the educational context of grades 6-8 students. This foundational work clarifies the critical competences policy makers and educators wish to promote, and the types of performance indicators that students in grade 6-8 contexts can achieve and that can be most meaningfully and feasibly assessed.

2. Competency model: reviewing the precise measurable definitions in the NGSS and state standards and specifying the claims that can be made about the relevant attributes of students based on the assessment.

3. Evidence model: describing the evidence that needs to be generated in the assessment to support the subsequent claims made about students (i.e., the behaviors or performances that demonstrate the skills being assessed).

4. Task/item design: identifying, conceptualizing, and prototyping tasks and items that provide the desired evidence within the constraints of test specifications.

5. Assessment development: assembling the tasks into a range of formats that support all the stated assessment claims with sufficient evidence.

6. Validation: ensuring all assessment instruments provide valid and reliable evidence within and across selected groups. This step is typically implemented in the intended assessment setting and provides opportunities to refine the assessment tasks. Validation studies data provides initial evidence on specific tasks and items that yield the most information on the targeted knowledge and skills. This evidence is typically used for optimization of the tasks and is addressed through assessment design.

7. Analysis and reporting: illustrating appropriate, meaningful and easy-to-communicate representations of the assessment results and recommendations for personalized learning and instruction.

Investigation Design

BrainPOP Science uses multi-day investigations designed to support three-dimensional learning and enhance traditional curricula. Each investigation includes all the elements of a complete lesson cycle, including: relevant and engaging hooks, hands-on scientific explorations, primary sources, related readings, and varied forms of assessment. These resources encourage students to make sense of natural phenomena and develop explanations for those phenomena by investigating high-interest Guiding Questions using BrainPOP resources. Importantly, these lessons address the four aspects of relevancy that are crucial for contextual learning: personal, professional, social, and societal (Van Aalsvoort, 2004; Dawood & Purdie-Greenaway, 2017; 2018). Students explain phenomena by developing and applying the Disciplinary Core Ideas (DCIs) and Crosscutting Concepts (CCCs) through use of the Science and Engineering Practices (SEPs). The embedded formative assessments and end-of-unit summative assessments combine to provide students, teachers, and administrators with robust mastery tracking, recommendations, and supporting resources.

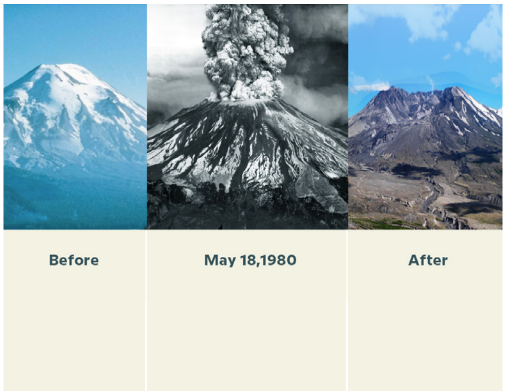
Phenomena and Guiding Questions

Investigations support learning by starting with a phenomenon and having students elicit prior knowledge as they respond to questions about the phenomenon. Phenomena-driven instruction helps students engage in practices that develop the knowledge necessary to explain or predict the observable, real-world phenomena. Per the NGSS, using phenomena shifts the focus of learning from learning about a topic to figuring out why or how something happens (2016). It is important that all students—including English language learners and students from cultural groups underrepresented in STEM—are supported in working with phenomena that are engaging and meaningful to them. Not all students will have the same background or relate to a particular phenomenon in the same way. BrainPOP Science considers many different student perspectives when choosing phenomena.

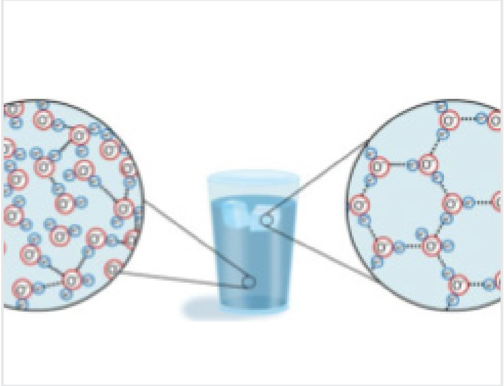
Each phenomena is paired with an image, video, or other resource. Investigation phenomena include:

What do you think the Hoover Dam is used for?

How do sneezes spread diseases?



How do volcanoes change over time?



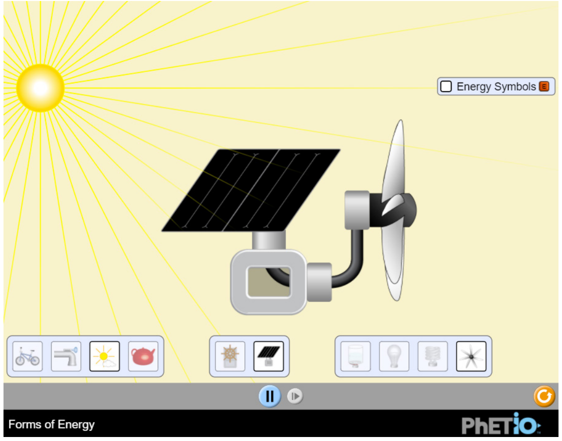
What do you notice about the difference between the ice and liquid water?

Some investigations also use a custom BrainPOP movie to introduce and contextualize the guiding question. For example, an investigation about natural versus artificial selection opens with a movie showing two dog owners walking their vastly different dogs (one very big, one very small) and wondering “how are they related?”

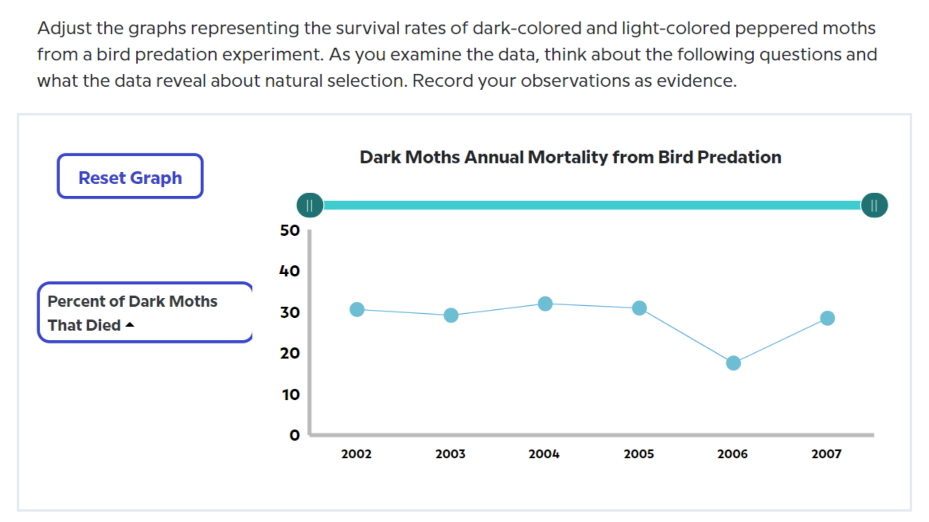
Collecting Evidence

Scientific reasoning uses thinking skills (inductive and deductive reasoning) to develop scientific knowledge. Some of these thinking skills include reading and evaluating scientific evidence, deducing and inferring conclusions from evidence, creating claims backed by evidence, being open to new evidence, seeing both sides of an issue, and communicating and justifying proposed explanations. The bulk of a BrainPOP Science investigation provides students the opportunity to learn by collecting evidence from a variety of resource types.

Most investigations include a **simulation** to support hands-on, interactive learning. Simulations allow students to observe science principles in action, manipulate models to test hypotheses, and collect evidence all while strengthening students’ conceptual understanding. Experimentation with the simulation enacts constructivist principles of learning by doing while building new or enhanced understanding



of the scientific concept. This strategy also elicits intrinsic motivation, as the student explores the concept and learns autonomously (Rosen, et al., 2020). In the “Forms of Energy” simulation shown above, students manipulate the source, mechanism, and output to see how different forms of energy are related.



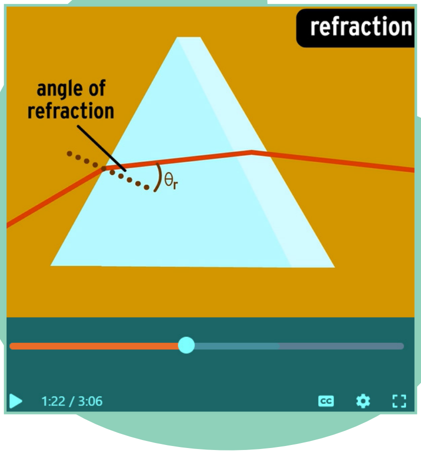
Adjust the graphs representing the survival rates of dark-colored and light-colored peppered moths from a bird predation experiment. As you examine the data, think about the following questions and what the data reveal about natural selection. Record your observations as evidence.

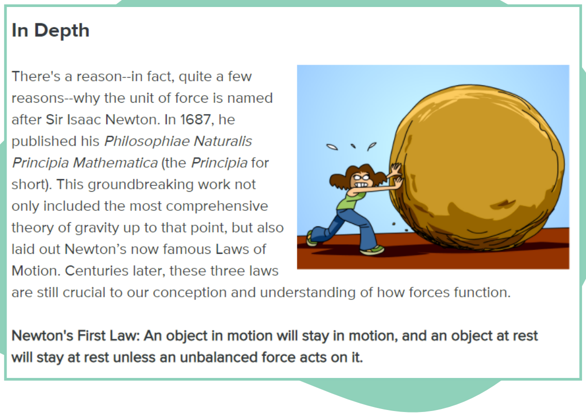
Data Manipulatives utilize a simple user interface that allows students to observe the effect of changing one variable at a time. In the data manipulative shown above, students can quickly toggle between the graphs of “Percent of Dark Moths That Died” and “Percent of Dark Moths That Lived” over time. These manipulatives use real data distilled to a form that is easily interpreted by young scientists.

The **3D Worlds** feature, shown below, enables students to explore virtual worlds that reflect real-world phenomena and encourage scientific practices. Through their self-driven investigation of the world, students are prompted to wonder about various questions related to the world as posed by in-world characters. The broad questions posed by these characters further empower students to explore the world through the lens of the investigation’s guiding question, but without narrowing student curiosity within the topic. This open-endedness drives a more authentic and immersive experience for students.

RITA: Look around and listen in to gather useful clues to answer the question. Now, where are those bats?

BrainPOP Science leverages the power of BrainPOP’s signature animations to explain some of the hardest-to-teach concepts with new **videos**, available in English and Spanish, embedded every lesson. These videos included important images and diagrams through a character-driven narrative. In doing so, abstract concepts are clearly represented in easily digestible, student-friendly visuals. Additionally, videos cue to important keywords and concepts which makes each video an effective tool throughout the entire lesson sequence. Videos are typically introduced earlier in the investigation,but are a valuable resource later when reviewing the investigation as a whole.





In Depth

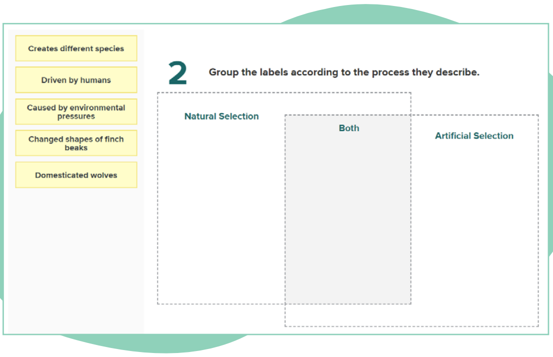
There's a reason--in fact, quite a few reasons--why the unit of force is named after Sir Isaac Newton. In 1687, he published his *Philosophiae Naturalis Principia Mathematica* (the *Principia* for short). This groundbreaking work not only included the most comprehensive theory of gravity up to that point, but also laid out Newton's now famous Laws of Motion. Centuries later, these three laws are still crucial to our conception and understanding of how forces function.

Newton's First Law: An object in motion will stay in motion, and an object at rest will stay at rest unless an unbalanced force acts on it.

Each investigation contains between one and three **related readings**. Readings are grade-level appropriate and target the key concepts within the investigation. These resources also provide opportunities to practice important reading comprehension skills such as making inferences, identifying main ideas and details, and comparing & contrasting. Some readings are required for the investigation, while others can be additionally assigned by the instructor in order to scaffold or differentiate the investigations for each students' needs.

Finally, in addition to the available content resources, BrainPOP Science provides students with short, informative movies that support the science and engineering practices. For example, accompanying every resource in an investigation is a link to watch the "How do I Collect Evidence" movie. In this three-minute movie, Moby learns that evidence is facts and information, and the best evidence can be used to support a claim. These movies also cover how to use data (e.g. recognizing patterns and organizing/displaying data based on the research question). These movies remind students how to collect relevant evidence as they progress through the investigation.

Formative Assessment


2 Group the labels according to the process they describe.

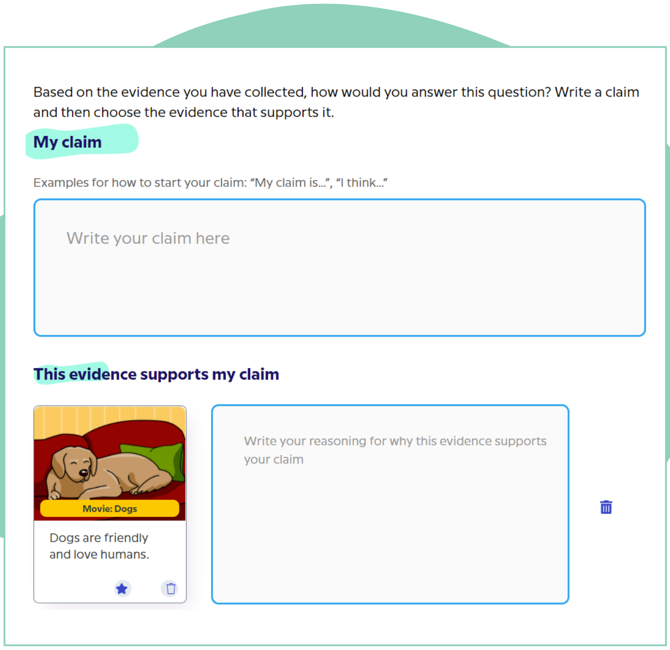
BrainPOP Science offers students two types of formative assessment in each investigation: an auto-scored Check-For-Understanding (CFU) and a teacher-graded Claim-Evidence-Reasoning (CER) paragraph.

The **Check-For-Understanding** is an integral part of the BrainPOP Science investigation. These quick, low-stakes checks serve as both a learning and assessment tool. Through CFUs students, teachers, and administrators have multiple opportunities to evaluate students' learning as they progress from resource to resource within the investigation. Each CFU is NGSS-aligned with an emphasis on the DCIs and CCCs, and incorporating SEPs when possible. Additionally, these varied data points contribute to the students' estimated mastery level, which will inform automated resource recommendations in future releases of BrainPOP Science.

CFUs utilize a multitude of autoscored item types. In addition to the standard multiple response and matching types, CFUs use a variety of drag-and-drop items to assess more advanced depth of knowledge (DOK) levels. These include Venn diagrams (pictured below), cloze reading texts, and concepting mapping. Each investigation exposes students to a mixture of item types in order to prepare them for: the investigation CER, the culminating unit assessment, and relevant state assessments.

Some CFUs combine item types in multi-step questions to better understand students' thought processes and misunderstandings. For instance, a CFU provides students with a claim to answer a research question and a related text. In part one, the student must highlight a sentence or two from the text which best supports the given claim. In part two, they need to choose a line or two of reasoning which best connect their chosen evidence to the given claim. The choices in part two include distractors that highlight misunderstandings when answered incorrectly - did the student correctly parse the text passage? Did the student understand the relevant vocabulary? Is the student struggling with the science concept or with the writing/CER structure? Here the CFU provides valuable feedback that a teacher or recommendation engine can use to adjust instruction as necessary.

When a student finishes collecting evidence from the investigation resources, they are ready to answer the Guiding Question in the form of a Claim-Evidence-Reasoning paragraph. Students construct a claim that answers the question and generate reasoning to connect each piece of collected evidence to the claim. CERs must be manually scored by teachers in the current release of BrainPOP Science; however auto-scored CERs are on the project roadmap for future versions. Current investigations come with sample CER responses and a grading rubric in the Teaching Guide to support instructors.


Based on the evidence you have collected, how would you answer this question? Write a claim and then choose the evidence that supports it.

My claim

Examples for how to start your claim: "My claim is...", "I think..."

Write your claim here

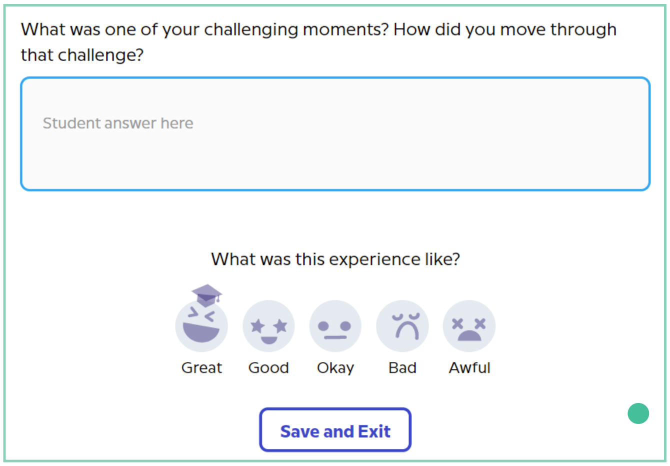
This evidence supports my claim

Dogs are friendly and love humans.

Write your reasoning for why this evidence supports your claim

Reflection

Each investigation concludes with a brief reflection question and rating for students. This step requires students to pause and reflect on their experience and understanding throughout the investigation. This also provides instructors with additional insight into the student experience and supplements the CFU and CER by providing informal data on the student's understanding of the investigation's material.



What was one of your challenging moments? How did you move through that challenge?

What was this experience like?

Unit-Level Assessments

In addition to the formative assessments, BrainPOP Science also includes summative assessments at the end of each instructional unit. Unit-Level Assessments are intentionally designed to provide administrators and teachers with actionable diagnostic insight on students' mastery of knowledge and skills aligned to the NGSS and state science standards. Utilizing assessment best practices, these high-quality assessments were developed to provide quality data that can be used to make sound educational decisions for instruction.

The Unit-Level Assessments target all DCIs, CCCs, and SEPs covered by the unit and provide a comprehensive three-dimensional view on student mastery, including strengths and weaknesses. These summative assessments are driven by state-level assessment rigor and design so that measurements are highly correlated with summative outcomes. Unit-Level assessments are intended to be completed in one class period (40 – 45 minutes) and include a diverse set of auto-scored item types. Each Unit-Level Assessment item was designed to measure the breadth and depth of NGSS and State-specific standards while also integrating all 3-dimensions.

Our intentional design and use of assessment best practice for the Unit-Level Assessments will result in highly actionable classroom data that can greatly impact instruction and learning.

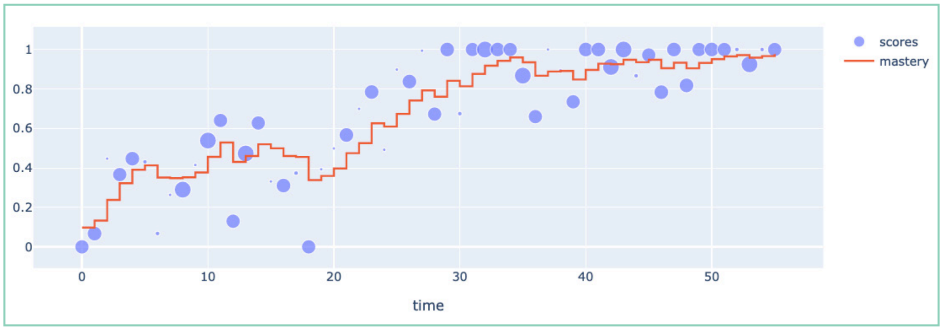
Actionable Diagnostics

In diagnostics, we estimate the learner's levels of mastery on a number of different skills. Action then can be taken based on these estimates: a high mastery on a skill suggests that the learner does not need any more content for that skill, and a low mastery on a skill suggests remediation. The word "skill" here is an umbrella term for any component of a standard (any of the Disciplinary Core Ideas, or Cross-cutting Concepts, or Science and Engineering Practices, or Performance Expectations in the NGSS, and likewise for any other system of standards).

The mastery level estimation is done via a proprietary machine-learning algorithm from the family of knowledge-tracing models. Such models have been used for decades in intelligent tutoring systems and similar learning environments, especially where personalization of instruction is the goal (Corbett & Anderson, 1995; Hawkins, et al., 2014; Rosen et al., 2018; Rushkin et al., 2017). They track each learner's actions in the learning environment and revise the individual mastery levels after each new interaction of the learner with relevant content. The degree of relevance of content elements to this or that skill is provided by content-tagging. Initially, we rely on the tagging done by Subject Matter Experts (SMEs) manually, or in an automated but SME-supervised way. Once a substantial amount of data about learner-content interactions is collected, the content tagging can be refined based on that data, using machine-learning methods. The user model, underlying the knowledge-tracing algorithm, contains a number of parameters, which likewise can be regularly updated based on the incoming data. As a result, the mastery estimation adapts to the changing learning conditions, and its accuracy improves.

The diagram below illustrates the mastery tracing for a single skill and a single simulated learner. The learner responded to a number of assessment questions over a period of more than 50 days and received a score on the 0-to-1 scale for each

each response, shown by the circles. Since the questions were of varying difficulty and targeted this particular skill to a varying degree, their contributions to the mastery estimates are not equal, as shown by the circle sizes. The red line shows the outcome of the algorithm: the skill mastery level evolving in time. You can see how the mastery level jumps every time a new assessment score has been received.



Our mastery-tracing algorithm allows incorporating the evidence from content types other than assessment questions. In fact, even with the assessment questions, it treats them not merely as a measurement of mastery but also as a learning experience.

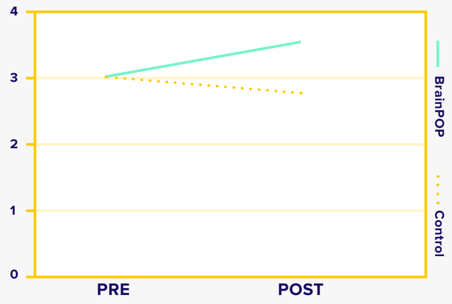
In addition to tracing the individual mastery level on every skill as illustrated in the diagram, our system also traces mastery on combinations of certain skills to account for multi-dimensional learning. For instance, in NGSS, we trace combinations of Core Ideas and Cross-cutting Concepts, or Core Ideas and Science and Engineering Practices. The mastery of such a 2-dimensional skill is not simply derived from the mastery levels of its constituent 1-dimensional skills. The reason is that it takes into account the evidence only from the learner's interaction with 2-dimensional content, whereas each constituent 1-dimensional skill also takes in evidence from 1-dimensional content. While we note in passing that our model is capable of tracing 3-dimensional mastery as well, the main challenge is rather how to present it to an administrator.

The mastery estimates can be presented to teachers and administrators, both on the level of individual learners and on a group level, to inform teaching and remediation. Another goal is to use it as one of the inputs for adaptive learning: a recommendation engine built into BrainPOP, which suggests the optimal path for a learner to go through content. The real-time "mastery profile" of a learner (their current mastery estimates on all skills) is important evidence in determining what content elements would be best for that learner to see next.

Efficacy Research Evidence

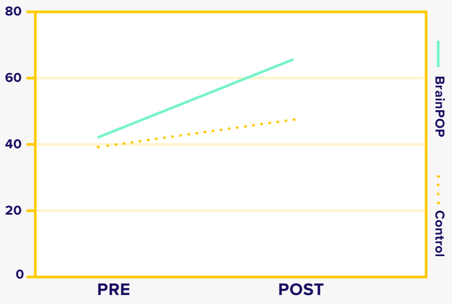
Prior findings provide evidence for the effectiveness of BrainPOP in improving learning outcomes in Science.

Study 1: Increased impact on learning outcomes and science learning motivation in the experimental (BrainPOP) group compared to the control group (no BrainPOP).



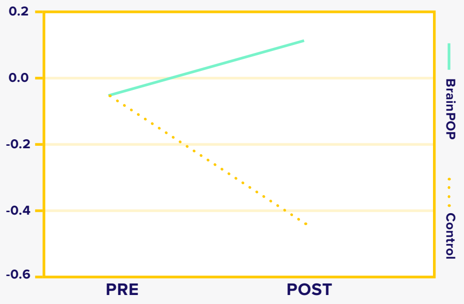
Learning outcomes: 5th grade Effect Size = 1.00; 7th grade Effect Size = 0.93
Science learning motivation: 5th grade Effect Size = 1.70; 7th grade Effect Size = 0.91

Study 2: Students in experimental group (BrainPOP movies) had higher scores on a science thinking skills assessment compared to the control group (textbooks with still images).



4th grade Effect Size = 0.72; 5th grade Effect Size = 0.56)

Study 3: BrainPOP subscribers perform better in Science compared to non-subscribers.



Science: Average Effect Size across 5 states = 0.46

Efficacy Study I: Effects on Transfer of Knowledge and Learning Motivation

The first study to assess the efficacy of BrainPOP addressed the following research questions (Rosen, 2009):

1. What is the effect of learning with BrainPOP on the transfer of knowledge, within the context of science and technology learning?

2. What is the effect of learning with BrainPOP on motivation for science and technology learning?

Participants

This study was conducted in five elementary schools and 3 secondary schools. All participating schools integrated teaching and learning based on the BrainPOP animation environment in their standard curriculum on a subject matter "Science and Technology."

418 students participated in the study: 250 from 5th grade and 168 from 7th grade. Students in the experimental group participated in science and technology lessons at least once a week and used BrainPOP for two to three months depending on the topic being taught. None of the students had participated in this type of instruction in the past. The students had full access to BrainPOP during after-school hours as well. Eight science and technology teachers volunteered to participate in the study. To eliminate or reduce the teacher effect, in most cases the same teacher taught both the experimental and the control group. All teachers had at least seven years of seniority.

Method

The study consisted of performance-based assessment tasks and a questionnaire administered using the pre-post method:

Pre-test: Before participation in classes integrating BrainPOP into the learning process.
Post-test: After the end of the learning period (2–3 months) of the relevant topics for each grade.

Results

The research revealed the following key findings:

- Learning with BrainPOP significantly increased the ability to transfer scientific and technological knowledge of elementary-school students ($ES = 1.00$, $t = 11.50$, $p < .001$) and secondary-school students ($ES = .93$, $t = 8.41$, $p < .001$). During the same period, the findings showed only a low increase in the same ability of the control group.

- Learning with BrainPOP into the learning process significantly increased science and technology learning motivation of elementary-school students ($ES = 1.70$, $t = 15.28$, $p < .001$) and secondary-school students ($ES = .91$, $t = 9.90$, $p < .001$). During the same period, the findings showed a decrease in motivation in the control group.

- Pre-post qualitative analysis of students' drawings showed that most of the elementary-school students showed the learner at the center of classroom interactions (58.1% compared with 20.7 before the experiment and 19.6% in the control group during the same period). The drawings illustrated the use of technology (63.6%) and showed greater emphasis on scientific equipment (38.8%). Most of the students' figures that appear in the drawings showed interest in learning (64.5% compared with 32.4% before the experiment and with 28.3% in the control group during the same time). Only small differences were found in the control group between pre- and post-test drawings. Similarly to elementary-school students, after the integration of BrainPOP animations into the learning experience, most of the secondary-student drawings placed the students in the center of the classroom interaction (51.2% compared with 6% before the experiment and with 13.6% in the control group during the same period). The drawings illustrated technology (45.8%) and emphasized scientific equipment (52.1%). Most of the students' figures in the drawings showed happiness and interest in learning (52.5%). Only small differences were found in the control group between pre- and post-test drawings.

Efficacy Study II: Integrating Animated Movies into Teaching

The goal of the second efficacy study was to examine how and when teachers integrated animated movies into their instruction and their perspectives on the role of animation in enhancing critical thinking skills (Barak, & Dori, 2011). Additionally, they explored the effect of animated movies on 4th and 5th grade students' academic performance. This study asked three primary research questions:

1. What methods do teachers use to integrate animated movies into their instruction?

2. What are teachers' views about the role of animations in enhancing students' thinking skills?

3. How does learning through animated movies affect students' conceptual understanding of science and reasoning ability?

Participants

Data were collected from 15 science teachers and 1335 students. Using stratified sampling to obtain an even distribution of ages and classes, 926 students from 5 elementary schools were included in the experimental group, and 409 students from two elementary schools were included in the control group. Experimental classes were sampled as those who utilized BrainPOP movies and supplementary activities in their science classes at least once a week. Control classes were sampled by classes that only used textbooks and still-images in their science classes.

Method

A mixed-methods approach was used, beginning with informal interviews with teachers during classroom breaks. Following this, the research was conducted in two stages: beginning with a pilot study to assess the reliability and validity of the instruments, followed by the main study. Animated movies were presented to the experimental group at least once a week, and teachers were trained to integrate web-based animations into their teaching by BrainPOP experts. In the control group, teachers used traditional teaching methods by following a textbook sequence that included colored images of science topics.

Results

Teachers' methods for integrating animated movies into their instruction: Before the study, 20% of teachers did not use computers at all in their instruction, 32% used educational technology for classroom demonstrations, 20% grouped students by teams for computer usage, 18% used education technology for home assignments, and only 10% used a combination of the three methods. After the study, 35% of teachers used a combination of methods to integrate education technology in their teaching.

Teachers' views about the role of animated movies in enhancing students' thinking skills: Qualitative responses found three primary perspectives on how teachers' view the role of animated movies in enhancing students' learning skills: enhancing scientific curiosity, acquiring scientific language, and fostering scientific thinking. Animated movies can increase motivation and incentives to study science and connect the material to their own lives. They also promote a greater understanding of science vocabulary through the movie narrator's explanations. Animated movies were also viewed by teachers as tools to help students think systematically, gather data, and solve complex problems.

Students' understanding of science concepts and reasoning ability: Science thinking skills were assessed through a pre/post 'scientific thinking skills' questionnaire. The study found that students who used animated movies as part of their science learning had improved understanding and implementation of science concepts compared to students who only used textbooks with pictures ($F = 127.50$, $p < 0.001$). Eta Squared analysis indicated that 9.3% of the growth in students' science thinking skills could be attributed to animated movies. Both 4th and 5th grade students in the experimental group had significantly higher scores on the 'science thinking skills' questionnaire compared to the control group (4th grade: $ES = .72$, $F = 81.6$, $p < 0.001$; 5th grade: $ES = .56$, $F = 53.3$, $p < 0.001$). 4th grade students also had more correct explanations to science reasoning problems than the control group ($F = 7.10$, $p < .05$), although these results were not significant for 5th graders in either the experimental or control groups.

Efficacy Study III: Effects on State Test Scores

The third efficacy study was conducted using BrainPOP subscriber usage data. It addressed research questions associated with the impact of BrainPOP as an educational platform on students' state test scores in Math, English Language Arts, and Science in grades 3-8 (see BrainPOP, 2018 for further details). This efficacy study takes a broad perspective on the use of BrainPOP in schools. Considering the non-prescriptive nature of BrainPOP's relationship to teaching practices, it's difficult to consider what implementation with fidelity may look like. For the purposes of this research, we considered the broadest use case—simply being a BrainPOP subscriber—to be the most inclusive intervention category that best accommodates the multitude of use cases that occur with BrainPOP. Research that considers which use cases of BrainPOP are most effective at fostering student achievement will be left to further studies. This study determined whether the use of BrainPOP in some form generally leads to higher student performance.

The approach to efficacy used in this paper also allows a determination of product value at a large scale. Studies of efficacy or effectiveness frequently focus on a single district or state. In contrast, this analysis extended the methodology across five states, each with a different achievement test. We chose five states for the analysis that best fit a mix of the following criteria: available and easily accessible public test score data, public test score data that used raw numbers rather than percentiles for schools, states with a significant BrainPOP subscriber base, and states with a relatively large number of schools. The last two criteria were intended to ensure relatively large sample sizes in both the intervention and control groups to best aid statistical testing. These criteria led to the selection of the following five states: California, Colorado, Florida, New York, and Texas.

For this study, we included data from the 2015-2016 school year. A school was considered a BrainPOP subscriber (intervention group) if it had an active subscription to BrainPOP for the entire school year (September 2015 to June 2016). The non-subscriber group (control group) included schools that did not subscribe to BrainPOP, as well as those that had a subscription that either started or ended mid-way through the school year. BrainPOP offers multiple products, but only a subscription to the flagship BrainPOP product was used to segment schools into the intervention and control groups. It is worth noting that many of these schools had a "Combo" subscription, which included BrainPOP Jr. (K-3), BrainPOP Español, and BrainPOP Français. This analysis was also limited to K-8 public schools due to ease of access to public school test scores; private schools (both subscribers and non-subscribers) were excluded from the analysis.

	ELA	**Math**	**Science**
California	0.27 **	0.28 **	0.26 **
Colorado	0.50 *	0.73 **	0.65 *
Florida	0.55 *	0.55 *	0.55 **
New York	0.32 *	0.39 **	0.60 ***
Texas	0.28 ***	0.29 ***	0.26 **

Table 1 *- Effect sizes for each state and subject test, based on school-wide scores (difference from mean averaged over all available grades). Significance values are denoted as * $p<0.10$, ** $p<0.05$, *** $p<0.01$*

The results qualify as Moderate ESSA Evidence, showing that schools with a BrainPOP subscription had a greater increase in standardized state test scores than a matched control group in all three subject tests: Math, ELA, and Science. The effect was always positive, always statistically significant, and verified in five states. To further validate the results, two additional correlational analyses that qualify as Promising ESSA Evidence were also performed. These analyses found generally positive results that were often statistically significant. The strongest effects were in grade 3-6 and in Math and Science.

Summary

Science education is among those showing the highest importance in the 21st century economy. Yet, many U.S. schools and students fall behind international peers on measures of science literacy, notably in scientific inquiry (OECD, 2017). In order to be competitive in an increasingly global workforce, the U.S. school system needs to prepare its students for higher levels of scientific literacy, deeper understanding of the core scientific concepts, and better mastery of the scientific process of developing and testing hypotheses (National Academies of Sciences, Engineering, and Medicine, 2019). Furthermore, incomplete or inaccurate perceptions of the practices of science and engineering can preclude students from making informed determinations about their interest and competencies in these fields. A better understanding of what scientists and engineers do—gained in part through investigative science and engineering tasks—might help middle school students to see these fields as relevant to them.

We believe that BrainPOP Science is well-positioned to provide teachers with critical tools to drive deeper learning with all students through innovation, holistic approaches, and evidence-based practices.

References

Arieli-Attali, M., Ward, S., Thomas, J., Deonovic, B., & von Davier, A. A. (2019). The Expanded Evidence-Centered Design (e-ECD) for Learning and Assessment Systems: A Framework for Incorporating Learning Goals and Processes Within Assessment Design. *Frontiers in Psychology, 10.*

Baldeh, M., Dawood, M. & Purdie-Greenaway, V. J. (2018). *Striving for more: exploring social identity and academic motivation through possible selves theory.* Presented at the Society for Personality and Social Psychology Annual Convention; 2018 Mar 1-3; Atlanta, GA.

Barak, M., & Dori, Y. J. (2011). Science education in primary schools: Is an animation worth a thousand pictures?. *Journal of Science Education and Technology, 20*(5), 608-620.

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in Practice: Making Scientific Practices Meaningful for Students. *Journal of Research in Science Teaching, 53,* 1082–1112.

BrainPOP (2018). *The Impact of BrainPOP on state assessment results: A study of the effectiveness of BrainPOP in grades 3 - 8.* https://content.brainpop.com/rs/567-ANS-609/images/The%20Impact%20of%20BrainPOP%20on%20State%20Assessment%20Results%202018.pdf

Carin A. A., Bass J. E., Contant T. L. (2003). *Methods for teaching science as inquiry.* (9th Ed.,) Upper Saddle River, NJ: Prentice Hall pp 111–123.

Corbett, A. T.; Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4* (4): 253–278.

Dawood, M. & Purdie-Greenaway, V. J. (2017). *An identity and belonging based approach to educational attainment among Muslim adolescents.* Presented at the Five Colleges Conference; 2017 May 5; New York, NY.

Dawood, M. & Purdie-Greenaway, V. J. (2018). *Adolescent religious identity, school belonging and educational attainment.* Presented at the Society for Personality and Social Psychology Annual Convention; 2018 Mar 1-3; Atlanta, GA.

Hawkins, W.J., Heffernan, N.T., & Baker, R.S. (2014). *Learning Bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities.* In International Conference on Intelligent Tutoring Systems (pp. 150-155). Springer, Cham.

Hubert, B. & Rosen, Y. (2020). *Equity in Learning with BrainPOP: Fostering Access and Impact for All.* BrainPOP. https://cdn-about.brainpop.com/wp-content/uploads/2020/07/Equity-in-Learning-with-BrainPOP.pdf.

Kim, Y. J., Almond, R. G., & Shute, V. J. (2016). Applying evidence-centered design for the development of game-based assessments in physics playground. *International Journal of Testing, 16*(2), 142-163.

Krist, C., Schwarz, C. V., & Reiser, B. J. (2019). Identifying Essential Epistemic Heuristics for Guiding Mechanistic Reasoning in Science Learning. *Journal of the Learning Sciences, 28*(2), 160–205.

Mislevy, R. J., Steinberg, L. S., Almond, R. G., & Lukas, J. F. (2006). *Concepts, Terminology, and Basic Models of Evidence-Centered Design.* In D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Eds.), Automated Scoring of Complex Tasks in Computer-Based Testing (pp. 15–47). Lawrence Erlbaum Associates, Inc.

Namdar, B., & Shen, J. (2015). Modeling-Oriented Assessment in K-12 Science Education: A synthesis of research from 1980 to 2013 and new directions. *International Journal of Science Education, 37*(7), 993–1023.

National Academy of Science. (1995). *National Science Education Standards.* Washington, D.C.: National Academy Press.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures.* Washington, DC: The National Academies Press.

National Research Council. (2006). *America's Lab Report: Investigations in High School Science.* Washington, DC: The National Academies Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC.: Committee on a Conceptual Framework for New K-12 Science Education Standards.

OECD. (2017). PISA 2015 *assessment and analytical framework: Science, reading, mathematics, financial literacy, and collaborative problem solving*. Paris, France: OECD Publishing.

Odden, T. O. B., & Russ, R. S. (2019). Vexing Questions that Sustain Sensemaking. I*nternational Journal of Science Education*, 41(8), 1052–1070.

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S.-Y. (2016). The Development and Validation of a Learning Progression for Argumentation in Science. *Journal of Research in Science Teaching*, 53(6), 821–846.

Ozersky, L. & Rosen, Y. (2020). Integrate computational thinking in your classroom using free digital resources. Creation Lab to be presented at the Annual Conference of the International Society for Technology in Education (ISTE), Anaheim, CA.

Park, M., Anderson, E., & Yoon, S. (2017). *Learning Biology Coherently Through Complex Systems, Scientific Practices, and Agent-Based Simulations*. Philadelphia, PA: International Society of the Learning Sciences.

Rosen, Y. (2009). The effects of an animation-based on-line learning environment on transfer of knowledge and on motivation for science and technology learning. *Journal of Educational Computing Research*, 40(4), 451-467.

Rosen, Y. (2013). *Measuring 21st century skills in a common core era: How to address both CCSS and 21st century skills assessments.* Workshop conducted at The National Conference on Student Assessment (NCSA), National Harbor, MD.

Rosen, Y., Arieli-Attali, M., Ward, S., Seery, J., Simmering, V., & Ozersky, L. (2020). *HERA: Exploring the power of adaptive scaffolding on scientific argumentation and modelling competencies in online learning systems*. Paper presented at The International Conference of the Learning Sciences (ICLS), Nashville, TN.

Rosen, Y. Ferrara, S., & Mosharraf, M. (2015). *Handbook of research on technology tools for real-world skill development*. IGI Global.

Rosen, Y., Rushkin, I., Rubin, R., Munson, L., Ang. A., Weber, G., Lopez, G., & Tingley, D. (2018). *The effects of adaptive learning in a massive open course on learners' skill development.* Proceedings of the Fifth ACM Conference on Learning @ Scale. UK: London.

Rosen, Y., Stoeffler, K., & Simmering, V. (2020). Imagine: Design for creative thinking, learning, and assessment in schools. *Journal of Intelligence*, 8(2), 16

Rushkin, I., Rosen, Y., Ang, A. M., Fredericks, C., Tingley, D., Blink, M. J., & Lopez, G. (2017). Adaptive Assessment Experiment in a HarvardX MOOC. In *EDM*.

Using Phenomena in NGSS-Designed Lessons and Units. (2016). Retrieved April 2, 2021, from https://www.nextgenscience.org/sites/default/files/Using Phenomena in NGSS.pdf

Van Aalsvoort, J. (2004). Logical positivism as a tool to analyse the problem of chemistry's lack of relevance in secondary school chemical education. *International Journal of Science Education*, 26(9), 1151-1168.

Yoon, S. A., Goh, S.-E., & Park, M. (2018). Teaching and Learning About Complex Systems in K–12 Science Education: A Review of Empirical Studies 1995–2015. *Review of Educational Research*, 88(2), 285–325.