Note: This is the fourth of four papers in a "related paper set" presented at the National Association for Research in Science Teaching, April 2013. The other papers are:

- Measurement of Analogical Reasoning around Earth Science Models (Rivet, Schmalstig & Kastens)
- Investigating Students' Use of Models to Support Claims about Earth Phenomena (Lyons, Rivet, Kastens & Miller)
- Emergence of Science Practices around Physical Models in Earth Science Classrooms (Miller, Rivet, Kastens & Lyons)

Students' Use of Physical Models to Experience Key Aspects of Scientists' Knowledge-Creation Process

K.A. Kastens, A. Rivet, C. Lyons, and A. R. Miller
(correspondence to kkastens@edc.org)

Purpose:

Our project concerns how students learn with physical models. This paper situates students' use of physical models within a theoretical framework of how scientists use physical and computational models to create new knowledge, and suggests that instructional use of physical models informed by this framework could deepen students' understanding of the nature of science.

Theoretical Framework:

This paper focuses on models that are devices (Nersessian & Patton, 2009) with an existence in the world outside of any individual's brain, in other words, "external" or "expressed" models rather than mental or conceptual models (e.g. Buckley & Boulter, 2000; Justi & Gilbert, 2000). Moreover, the models of the current paper exhibit observable behaviors in response to actions taken by the researcher, an attribute that we will refer to as "runnable" models. At the frontiers of science, runnable models are becoming an increasingly powerful component of the scientific arsenal. Global climate models (Weart, 2011) and the recently announced software simulation of an entire bacterium (Freddolino & Tavazoie, 2012) are two prominent examples.

How scientists use expressed runnable models:

Figure 1 depicts a set of processes that scientists may engage with when creating new knowledge with an external runnable model. Scientists make observations (a) of the referent (the Earth system, in our case) and interpret those observations as being the result of processes (b). They represent those processes in a physical or computational device (c), then activate the device (i.e. "run" the model), and compare the behavior of the model with the behavior of the Earth as captured in data (d). Where the model behavior and the behavior of the Earth do not correspond, they improve the model (e). They then use the model to make predictions of how the Earth would be expected to behave under a set of not-yet-observed conditions (f). They collect more data (g), and once again compare the behavior of the model with the behavior of the Earth as captured in the new data (d, again). For a complex system, there can be many trips around this circle (d-e-f-g-d), involving many research groups and many data types. Only after a model has been refined by repeated trips around the improvement circle, yielding fewer and smaller areas of non-correspondence between model behavior and Earth behavior, does the model earn the credibility to be used to make inferences about times and places for which no observations are available (h).

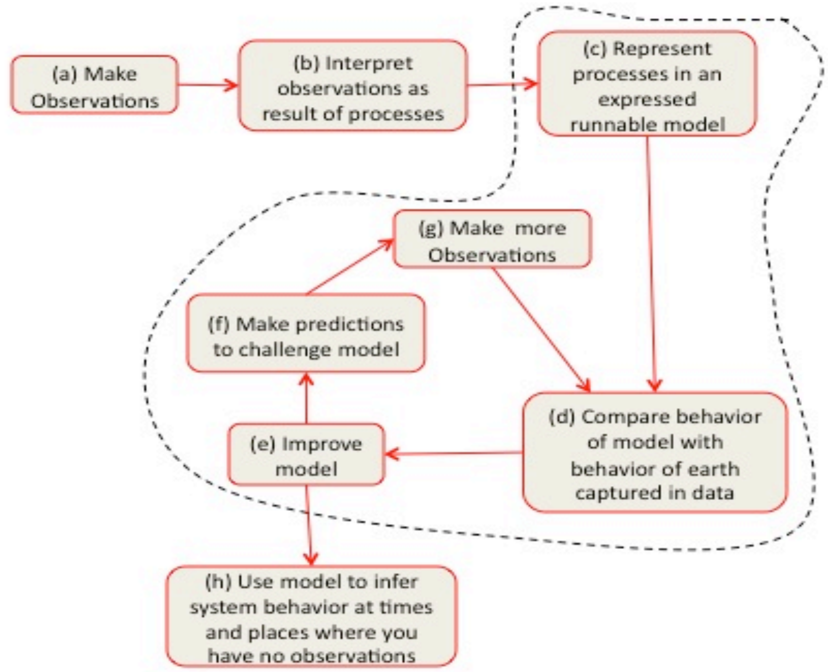


Figure 1: Schematic of how scientists develop, test, and obtain new knowledge from external runnable models. The steps inside the dashed line are almost completely invisible to students and the public.

Two examples from modeling practice at the frontiers of science:

We now provide two examples of how this sequence plays out at the frontiers of science. The first example concerns climate modeling and the development of the theory of global climate change, as recounted by Weart (2011). Climate modeling has passed through many cycles of model improvement and testing against data, spanning more than a century. The process began with observations (a) that air temperature varies from day to night, from winter to summer, and with latitude, and that there had been cycles of ice ages and interglacials. Explaining ice ages was a particular challenge. In the 1920's Milankovitch interpreted the glacial/interglacial cycles in terms of processes of subtle shifts in Earth's orbital parameters (b) and expressed his ideas in a mathematical model (c). Nearly fifty years passed before suitable data were available to continue with step (d): paleotemperature data from ice cores showed that the glacial/interglacial cycles did indeed coincide in time with Milankovitch's orbital cycles. However, the model was still flawed, because the amplitude of the temperature fluctuations produced by the model was much smaller than those inferred from the Earth data. The model had to be improved (e) by adding positive feedback loops to amplify the subtle energy shifts triggered by the orbital changes. Later models incorporated many types of feedbacks, both positive and negative, and gave a central role to greenhouse gases such as water vapor and carbon dioxide. Based on these more complete models, scientists made a series of predictions (f) about how the the spatial and temporal distribution of temperature would change as anthropogenic carbon dioxide was added to the atmosphere: that nights would warm faster than daytime, that the poles would warm faster than mid-latitudes, and that the stratosphere would cool even as the lower atmosphere warmed. With the passage of time, sufficient new data (g) accumulated to test these predictions by comparison of model behavior against data (d). The success of these and other predictions is what gives climatologists the confidence use the models to explore "what if" scenarios far in to the future (h).

The second example comes from laboratory science: software simulation of an entire bacterium (Karr, et al, 2012; Freddolino & Tavazoie, 2012). This modeling effort built on 900 prior publications of 1,900 experimentally observed parameters, a massive distributed effort comprising steps (a) and (b) of figure 1. Karr's team built a computer model (c) of the organism, by combining 28 modules into a unified runnable model. Next, they "validated the model against a broad range of independent data sets that were not used to construct the model and which encompass multiple biological functions" (p. 391): figure 1, step (d) and (e). They then used their model to "predict molecular interactions that are difficult or prohibitive to investigate experimentally," exploiting their ability to make predictions in the context of the entire cell (f). The behavior of the model led the investigators to hypothesize an emergent control of cell-cycle duration independent of genetic regulation, make quantitative predictions about the energy budget within the cell, and make inferences about which genes are essential to sustain cellular growth and division (f). They located laboratory experimental data (g) (obtained by different investigators and not incorporated into the building of the model) and compared the model behavior with the experimental data (d), and analyzed the correspondences and non-correspondences. The paper reports multiple cycles of predict-from-model (f), generate or find data to test prediction (g), and compare-data-with-model-behavior (d). To reconcile some non-correspondences between model behavior and data, they had to go back into the lab, in a process they call "model-driven biological discovery." The lab-results/model-output comparison led to some tweaks to parameters in the model (e), and also to a deeper understanding of the nuances of the lab data. Karr et al's paper concludes with suggestions for other questions that could be answered by whole-cell models (h).

Steps that students and the public rarely see:

In general, science students and the public see only the beginning and the end of this process. They observe scientists executing the beginning steps (a) and (b) in such contexts as *Nova* videos. Typical laboratory and field-based student activities (e.g. making and interpreting observations of weather, a stream, or an outcrop) also involve steps (a) and (b). Students and the public see the final step (h) reported in the newspaper, as for example when air temperature or sea level is forecast for twenty years into the future, a time for which we have no empirical observations. But everything inside the dashed line of figure 1 is typically invisible to individuals outside of the research community. This is a serious problem, because without understanding the testing/confirming/calibrating steps that occur inside the dashed line, there is no compelling reason for students or the public to believe the inferences or predictions that emerge in the final step.

Application to Earth Science education

Physical models can also be runnable models. In this paper we will use as an example a lunar phases model comprising a globe or large ball for the Earth with a doll attached for the observer, a smaller ball for the moon, and a lamp for the Sun. This is one of three physical models employed in our project "Bridging the Gap between Tabletop Models and the Earth System" (Rivet & Kastens, 2012; Rivet, et al, this paper set). The impressive educational affordances of simple physical models of the sun/moon/earth system have been analyzed by

Lehrer & Schauble (2006) and Stewart, et al. (2005), who described how over the course of instruction students were able to modify their models to account for an increasing range of empirical data.

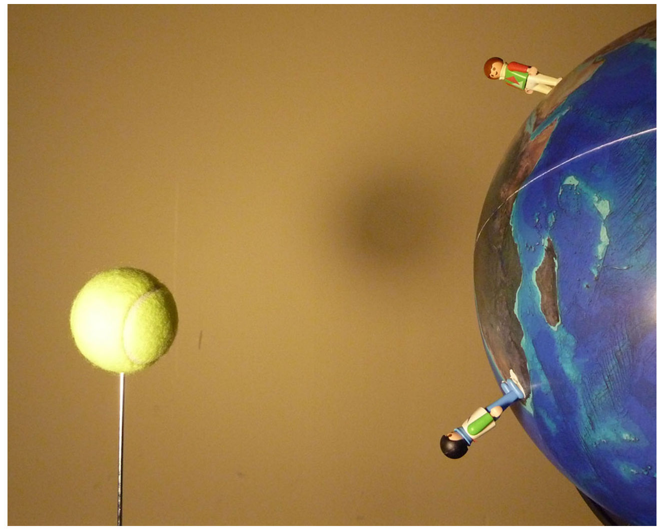
Three instructional strategies recommended by our project offer students opportunities to experience and practice aspects of the knowledge-creation process of figure 1, including the steps inside the dashed line, using simple physical models. In our first instructional strategy, the teacher sets up purposeful opportunities for students to identify and articulate correspondences and non-correspondences between the physical model and the Earth system. Ability to identify such correspondences and non-correspondences underlies model-using steps (d) and (e), because correspondences indicate aspects of the model that are successfully reproducing the behavior of the referent system, while non-correspondences indicate aspects of the model that are candidates for improvement in step (e). In our second instructional strategy, students use the model to answer questions or solve problems. Students thereby experience the power of distributed cognition (Hutchins, 1995; Nersessian, et al, 2003), in which part of the cognitive load is offloaded onto the runnable model. In our third instructional strategy, students reason about the relationships among concepts, the physical model, and empirical observations or data from the real Earth system. This ability underlies model-using steps (g)→(d), in which the model is challenged by exposure to empirical data.

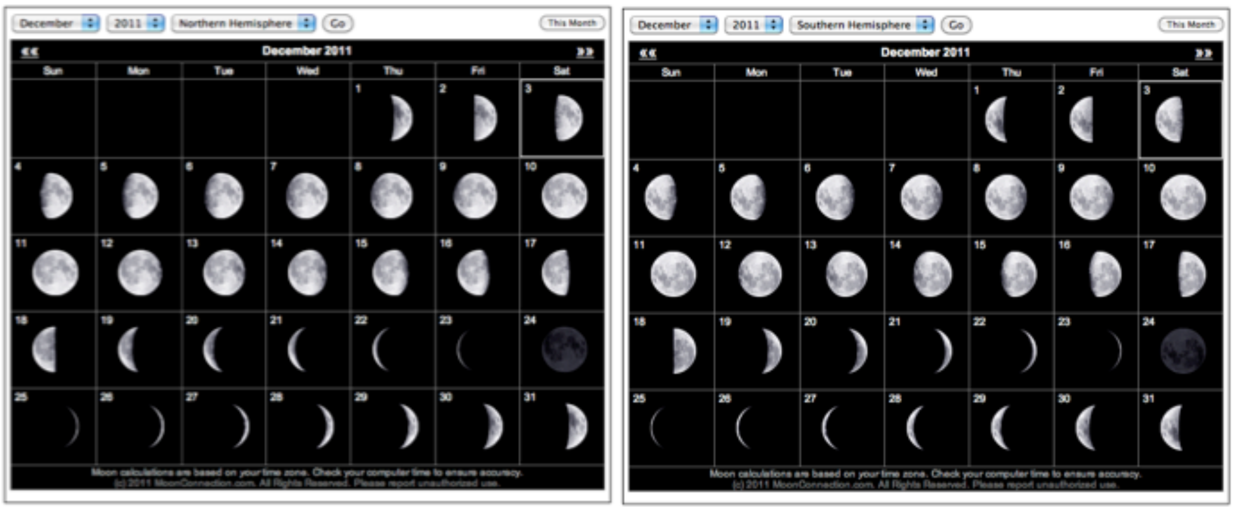
We suggest that students can experience the entirety of the knowledge-creation process of figure 1, including a complete cycle through the steps inside the dashed line, using the lunar phases physical model. First, students make observations of the changing appearance of the moon over time (a), discuss how these observations could be interpreted in terms of the moon's motion and position (b), and then create a model in which the illuminated portion of the small ball as seen from the doll's vantage point changes methodically as the model is run (c). When we have done steps (b) and (c) with students and teachers, their initial models typically contain an ambiguity: some groups have the small ball going one way around the large ball and other groups have the small ball going the other way around. At this point, students are asked to figure out which way is correct by comparing the behavior of the model with the Earth's behavior as captured in their data (d). The data permit only one answer: only if the small ball/moon goes clockwise around the large ball/Earth does the moon fill from right to left as seen by the northern hemisphere doll/observer. Students then improve their model (e) by removing the direction ambiguity. Next students are asked to use their model to predict (f) what the moon phases would look like from mid-latitudes in the southern hemisphere. After some struggle, students may think to enhance the model's capabilities by adding a second doll to the southern hemisphere. From the southern hemisphere doll's perspective, the right side of the moon is illuminated when the northern hemisphere doll "sees" the left side illuminated (Figure 2, upper). By running the model and recording the appearance of the moon from both vantage points, students can can then predict that new moon would fill from left to right from the perspective of a mid-latitude southern hemisphere observer, rather than the right to left progression familiar to northern hemisphere observers (Figure 2, lower). They can't go to the southern hemisphere to make more measurements, but they can access an online southern hemisphere lunar calendar to compare (d) their model-based prediction with observation. At this point they have completed a circuit of the figure 1 loop.

Figure 2: Use of a physical model to make a prediction about a place for which students don't have observations, step (f) of figure 1.

(right) The northern hemisphere doll "sees" the tennis-ball moon with its illuminated side on her left, while the southern hemisphere doll "sees" the same moon with its illuminated side on her right.

(below) Using the physical model and perspective-taking, students can predict what the moon phases would look like in the southern hemisphere. The answer supported by data is that new moon fills from right to left in the northern hemisphere and from the left to right in the southern hemisphere (compare calendars carefully).





Relationship with Modeling as a Practice of Science in the NGSS

The Framework for K-12 Science Education (NRC, 2011) establishes "Practice Two: Developing and Using Models" as one of eight central and essential "Practices of Science and Engineering." The meaning of each practice is further explicated in Appendix F of the Next Generations Science Standards (NGSS) (Achieve, 2013). In the preamble to the section on modeling, NGSS includes the excellent statement that: "Students can be expected to evaluate and refine models through an iterative cycle of comparing their predictions with the real world and then adjusting them to gain insights into the phenomenon being modeled." This statement can be read as a narrative description of the cycle depicted in figure 1, and Figure 1 can be viewed as an unpacking of this narrative into a series of steps. Importantly, the preamble wording presents models as something to be improved over time through successive refinement rather than as an authoritative final answer, a stance that is further explicated by Windschitl, et al. (2008) and Schwartz, et al. (2009).

NGSS Appendix F also presents a table or "practice matrix" which specifies the components of each practice that students are expected to master at the end of each grade band. By grade 2, students should be able to "Develop … models (e.g. … physical replicas….) that represent relationships, … and/or patterns in the natural and designed world." By grade 5, they should be able to "Develop and revise models collaboratively to… explain frequent and regular events." These components correspond with our step (c) of figure 1.

By grade 8, students should be able to "Modify models—based on their limitations—to increase detail or clarity, or to explore what will happen if a component is changed." This component superimposes three of the steps of figure 1, in a way that could be confusing to students and teachers. Identifying the "limitations" of a model is related to our step (d), but only if students understand that the appropriate measure of the quality of a model is how well it reproduces the attributes and behavior of the referent system as captured in data or observations from the real world. We can't take this for granted, insofar as Pluta et al (2011) tell us that only 28% of middle school students say that agreement with evidence is an attribute of a "good" model. "Modify a model …to increase detail or clarity" is our step (e): "improve the model." "Modify a model … to explore what will happen if a component is changed" corresponds to our step (h). The potential problem here is that a model cannot appropriately be used in step (h) fashion to explore "what if" scenarios unless it has already been validated by passage around the loop of figure 1. This point is obscured by combining step (e) reasoning with step (h) reasoning into the same standards statement.

By grade 12, students should be able to "Use models… to predict phenomena…."; this is our step (f). And they should be able to "Design a test of a model to ascertain its reliability"; this could be an expression of our sequence of steps (f), (g), (d). And finally, they should be able to "evaluate the merits and limitations of two different models of the same … system in order to select or revise a model that best fits the evidence"; comparing model with evidence is the essence of our step (d).

In summary, the components of modeling practice spelled out the NGSS practice matrix capture all of the elements of the iterative modeling process depicted in our figure 1 and described in the Practice 2 preamble statement. However, the connective links that shape these components into an integrated, multi-step process, a process that can result in new knowledge of a previously not-understood aspect of the referent system, are not in the practices matrix. Duncan & Rivet (2013) stress that the trio of concepts, practices, and epistemology is at the heart of the efforts to revise the K-12 science standards. Here is a case where there is still one step missing to get all the way to an understanding of how scientists know what they know, all the way to scientists' epistemology. The standards writers have performed a great service to science education by unpacking the intertwined components of a complex practice of science. However, they have not told teachers or curriculum writers how to put the components back together again into an integrated whole. We hope that figure 1 can help.

Significance

The means by which scientists use runnable models to create new knowledge about a referent system is rarely explored in science education and is almost completely opaque to the general public. It is not just that the internal workings of scientists' computer models are inaccessible to non-scientists; the intellectual steps have also been largely invisible. Our work suggests that it is possible for novices to experience some of the crucial knowledge-creating affordances of scientists' models while working with simple, comprehensible physical models. We envision a teaching sequence that would build on our current work to guide students through all of the steps of figure 1, including the usually-hidden steps inside the dashed line. Activities of this sort will be needed to reach the high bar set by the NGSS when they direct that "Students can be expected to evaluate and refine models through an iterative cycle of comparing their predictions with the real world and then adjusting them to gain insights into the phenomenon being modeled" (Achieve, 2013, Appendix F). With appropriate metacognitive reflection, such instruction could help students better understand not only an aspect of how the Earth system works, but also an important aspect of how modern science works.

References

Adams, R. J., &Khoo, S. T. (1996). Quest. Melbourne, Australia: Australian Council for Educational Research.

Achieve. (2013). Next Generation Science Standards (January 2013 version). Downloaded from http://www.nextgenscience.org/next-generation-science-standards February 2, 2013.

Buckley, B. C., & Boulter, C. J. (2000). Investigating the role of representations and expressed models in building mental models. In J. K. Gilbert & C. J. Boulter (Eds.), Developing Models in Science Education (pp. 119-135). Dordecht: Kluwer Academic Publishers.

Duncan, R. G., & Rivet, A. E. (2013). Science learning progressions. *Science, 339*(6118), 396-397.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Science Education, 32*(1), 268-291.

Freddolino, P. L., & Tavazoie, S. (2012). The dawn of virtual cell biology. *Cell, 150*, 248-250.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. Cognitive Science, 34(5), 752-775.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39*(2), 93-104.

Justi, R., & Gilbert, J. (2000). History and philosophy of science through models: some challenges in the case of 'the atom.' *International Journal of Science Education, 22*(9), 993-1009.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: The MIT Press.

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Jr., . . . Covert, M. W. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell, 150*, 389-401.

Kennedy, C. A., Wilson, M., Draney, K., Tutunciyan, S., &Vorp, R. (2008). ConstructMap Version 4 (computer program). University of California, Berkeley, CA: BEAR Center.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 371-387). Cambridge: Cambridge University Press.

National Research Council. (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.

Nersessian, N. J., Kurz-Milche, E., Newsletter, W. C., & Davies, J. (2003). Research laboratories as evolving distributed cognitive systems. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society* (pp. 857-862). Mahway, NJ: Lawrence Erlbaum.

Nersessian, N. J., & Patton, C. (2009). Model-based reasoning in interdisciplinary engineering. In A. Meijers (Ed.), *Handbook of the Philosophy of Technology and Engineering Sciences* (pp. 687-718). Amsterdam: Elsevier.

Pluta, W. J., Chinn, C. A., & Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. *Journal of Research in Science Teaching, 48*(5), 486-511.

Rivet, A. E., Schmalstig, M., and Kastens, K A.,2013, Measurement of analogical reasoning around Earth Science models, this paper set.

Rivet, A. E., & Kastens, K. A. (2012). Developing a construct-based assessment to examine students' analogical reasoning around physical models in Earth Science. *Journal of Research in Science Education, 49*(6), 713-743.

Sadler, T. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. Journal of Research in Science Teaching, 41(5), 513-536.

Schwarz, C., Reiser, B., Davis, B., Kenyon, L., Acher, A., Fortus, D., . . . Krajcik, J. (2009). Designing a learning progression for scientific modeling: making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632-368.

Stewart, S., Passmore, C., Cartier, J., Rudolph, J., & Donovan, S. (2005). Modeling for understanding in science education. In T. Romberg, T. Carpenter & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 159-184). Mahwah, NJ: Lawrence Erlbaum Associates.

Weart, S. (2011). General Circulation Models of the Atmosphere, in The Discovery of Global Warming. Retrieved January 4, 2012, from http://www.aip.org/history/climate/GCM.htm

Windschitl, M., Thomson, J., & Braaten, M. (2008). Beyond the scientific method: model-based inquiry as a new paradigm of preference for school science investigations. *Science Education, 92*(5), 941-967.

Wilson, M. (2005). Constructing Measures: An Item Response Modeling Approach. Mahwah, NJ: Erlbaum.