4 Improving Students' Scientific Thinking

David Klahr, Corinne Zimmerman, and Bryan J. Matlen

The production of a scientifically literate population is a fundamental goal of our educational system. The justifications for that goal, and descriptions of paths toward it, have been reiterated many times in recent decades, as exemplified by major policy statements and specific recommendations from prestigious organizations ranging from "Benchmarks of Scientific Literacy" (AAAS, 1993) to the recent "Framework for K-12 Science Education" (NRC, 2012). Consequently, substantial effort has been devoted to determining how to increase the likelihood that, as students progress through school, they will acquire at least a rudimentary understanding of fundamental domain-general scientific concepts and procedures, as well as a nontrivial amount of domain-specific concepts. However, given the vast number of those procedures and concepts, it is not surprising that the full science curriculum presented to students from pre-school through high school has often been characterized as "a mile wide and an inch deep" (Li, Klahr, & Siler, 2006; Santau et al., 2014).*

Thus, the challenge facing researchers interested in improving science education is to enhance the quality and generality of the answers to two related questions: *What is scientific thinking?* and *How can it be taught?* In this chapter, we attempt to answer the first question by presenting a brief summary of a broad framework that characterizes the essential aspects of scientific thinking and reviewing the developmental origins of scientific thinking. We answer the second question by describing a few representative examples of research on teaching science in specific domains, such as physics, biology, and earth sciences – organized according to the framework – and selected from the extensive literature on different ways to improve children's basic ability to think scientifically.

What Is Scientific Thinking?

Scientific thinking is a particular form of human problem-solving that involves mental representations of (1) hypotheses about the structure and processes of the natural world and (2) various methods of inquiry used to determine the extent

* Thanks to Audrey Russo for her painstaking proofreading and editing of the references. Preparation of this chapter was supported in part by grants from the Institute of Education Sciences to Carnegie Mellon University (R305A100404 and R305A170176), and from the National Science Foundation to the University of Chicago (1548292). Opinions expressed do not represent the views of the US Department of Education or the National Science Foundation.

to which those hypotheses are consistent with phenomena. Scientific discovery is a type of problem-solving that involves a high-level search through two complementary problem spaces (Newell & Simon, 1972). One such space is the *Hypothesis Space*, and the other is the *Experiment Space*. The two spaces are linked through the process of *Evaluating Evidence*. This dual-space framework, dubbed SDDS (for Scientific Discovery as Dual Search) by Klahr and Dunbar (1988), is used to frame the literature reviewed in this chapter.

The three cognitive processes from the SDDS model are listed in Table 4.1, along with corresponding science practices that have been elucidated in recent policy statements about science education (NRC, 2012). Hypothesis Space Search involves the formulation and refinement of hypotheses. The specific practices that support this search are (1) asking questions and (2) developing and using analogies and models. Experiment Space Search involves planning and carrying out investigations. Finally, Evaluating Evidence requires (1) analyzing and interpreting data and then (2) constructing explanations. Moreover, each of these science practices can vary along a dimension from being highly domain-specific to domain-general. In Table 4.1, we have listed some representative publications that focus on specific cells in the overall taxonomy of scientific thinking and we will describe some of them in more detail below.

The particular studies summarized in this chapter were chosen because they exemplify specific *psychological processes* subsumed in the aspects of scientific thinking that have been used to organize the rows in Table 4.1. We have further divided the studies into those focusing on children's domain-specific knowledge and those focusing on domain-general knowledge, even though many of those we have classified as domain-general are – of necessity – situated in specific domains. For example, we do not focus on children's developing knowledge about such *concepts* as the periodic table of the elements, or friction, or the definition of absolute zero. Nor do we focus on how children learn about particular domain-specific *processes* (e.g., alternating current, photosynthesis, Newton's Laws, or chemical equilibrium). Instead, we approach the topic of scientific thinking from the perspective of cognitive and developmental psychology. Using this perspective, we describe some highly general aspects of what it means to "think like a scientist" or to exhibit "scientific thinking" about some domain or problem, and we comment on how different instructional approaches and pedagogical strategies can facilitate children's acquisition and mastery of this kind of knowledge.

What Are the Developmental Origins of Scientific Thinking?

Beginning in infancy, children learn about the natural world in ways that will influence their later scientific thinking. A notable achievement in the early years of life is the ability to think representationally – that is, being able to think of an object both as an object in and of itself and as a representation of something else. Even three-year-old children are able to use scale models to guide their search for hidden objects in novel locations by applying the spatial relations in the model to the corresponding

Table 4.1 *A taxonomy for categorizing psychological investigations of aspects of science education, with representative examples of each type*

		Type of knowledge	
Cognitive processes	Science practices	Domain-specific	Domain-general
Forming and Refining Hypotheses (Hypothesis Space Search)	**Asking questions**	**A1** Samarapungavan, Mantzicopoulos, and Patrick (2008) King (1991)	**D1** Chouinard (2007) Jirout and Klahr (2012) Kuhn and Dean (2005)
	Developing and using analogies and models	**A2** Christie and Gentner (2010) Matlen et al. (2011) Clement (1993, 1982, 2000) Vendetti et al. (2015) Lehrer and Schauble (2004)	**D2** Raghavan and Glaser (1995)
Investigation Skills (Experiment Space Search)	**Planning and carrying out investigations**	**B** Metz (1997) Schwichow, Zimmerman et al. (2016)	**E** Chen and Klahr (1999) Sodian, Zaitchik, and Carey (1991) Siler and Klahr (2012) Zimmerman and Croker (2014)
Evaluating Evidence	**Analyzing and interpreting data/evidence**	**C1** Amsel and Brock (1996) Penner and Klahr (1996a, b) Kuhn (2011) Masnick, Klahr, and Knowles (2016)	**F1** Masnick and Morris (2008)
	Constructing explanations	**C2** Inagaki and Hatano (2008) Lehrer and Schauble (2004)	**F2** ynatt, Doherty, and Tweney (1977, 1978) Kelemen et al. (2014)

Note: The structure of this table is taken from Zimmerman and Klahr (2018). The cognitive processes categories (row headings) and knowledge types (column headings) are adapted from Klahr and Dunbar (1988), Klahr (1994), and Klahr and Carver (1995). The science practice row subheadings are from the *Framework for K-12 Science Education* (NRC, 2012). Cell contents cite a few studies that exemplify the row and column headings for each cell. The corresponding text in the chapter cites additional examples.

relations in the real world (DeLoache, 1987), and four- to five-year-olds develop the ability to engage in pretend play, where one (usually mundane) object can represent another (often exotic) object (Hopkins, Dore, & Lillard, 2015; Sutherland & Friedman, 2013), such as when a stick becomes a magic wand or a sword. This fundamental ability lies at the heart of important scientific skills such as analogical reasoning and lays the groundwork for later scientific tasks such as interpreting and reasoning with models, maps, and diagrams (Uttal, Fisher, & Taylor, 2006).

The nature of knowledge acquisition in young children is a source of intense debate in developmental science. Some developmental theories espouse that children's acquisition of knowledge is similar, in many respects, to scientists' – that is, guided by top-down processes and theory-level explanations (e.g., Gelman & Coley, 1990; Gopnik & Sobel, 2000; Keil et al., 1998). Other theories espouse that knowledge is acquired through lower-level perceptual, attentional, and memory-based processes (e.g., Rakison, Lupyan, & Oakes, 2008; Sloutsky & Fisher, 2004; Smith, Jones, & Landau, 1996). Regardless of the way in which children acquire knowledge, it is clear that by the time they enter formal instruction they have strong ideas about the way in which the world works, that is, about what causes what. However, many of those ideas are partially or entirely incorrect (Vosniadou, 2013). In the science education literature, these mistaken, distorted, or partially correct notions, ideas, and beliefs about the world that children often bring to the science classroom are sometimes referred to as "preconceptions" and other times as "misconceptions." As Horton (2007) put it:

> "Misconceptions" seems excessively judgmental in view of the tentative nature of science and the fact that many of these conceptions have been useful to the students in the past. "Preconceptions" glosses over the fact that many of these conceptions arise *during* the course of instruction. (Horton, 2007, p. 4, emphasis added)

In this chapter we use both terms, roughly according to the context as described by Horton.

Children's preconceptions have been shown to influence learning in a variety of scientific domains including physics (Clement, 1982), thermodynamics (Lewis & Linn, 1994), astronomy (Vosniadou & Brewer, 1994), biology (Inagaki & Hatano, 2008; Opfer & Seigler, 2004), geoscience (Gobert & Clement, 1999), and chemistry (Wiser & Smith, 2008). Moreover, preconceptions can exist not only for domain-specific knowledge but also for domain-general science concepts, such as learning the principles of experimental design (Chen & Klahr, 1999; Lorch et al., 2010; Siler, Klahr, & Matlen, 2013) or understanding the purpose of scientific models (Grosslight, Unger, & Jay, 1991; Treagust, Chittleborough, & Mamiala, 2002).

One particularly interesting and widespread preconception was first reported in Vosniadou and Brewer's (1992, 1994) classic studies of children's mechanistic explanations regarding the day and night cycle. Prior to formal schooling, many children have ideas about how and why day turns to night, ideas that are deeply rooted in children's everyday concrete experiences (and which parallel the history of early prescientific notions). For instance, many children initially believe that the Earth is a flat plane and the sun travels from above to below the Earth to create day and night. Such preconceptions often continue to exist even after children receive formal instruction. For example, on being instructed that the Earth is in fact round, many children refine their conception of the Earth to being a flat disk-shaped object in order to account for what they are told (i.e., the Earth is round) and what they observe (i.e., the Earth is flat). Vosniadou and Brewer (1992) identified several similar morphed or synthetic understandings, such as the belief that the Earth is a hollow sphere with a flat bottom on which we live. One challenge to science education is the fact that such preconceptions are particularly resistant to formal instruction. Several studies have identified preconceptions that conflict with scientific conceptions even after semester-long or university-level courses (Clement, 1982; Lewis & Linn, 1994; Wiser & Smith, 2008; Zimmerman & Cuddington, 2007).

Chi (2013) has distinguished between two types of learning based on preconceptions. The first is knowledge acquisition, or *gap-filling*, where students' prior knowledge is incomplete and they need to acquire new knowledge. The second is where students have prior knowledge that is in conflict with to-be-learned knowledge (e.g., scientific concepts). For example, preschool-age children's tendency to believe that plants are inanimate objects conflicts with the scientific perspective that plants are living organisms. When children begin to conceive of plants as living organisms, it is said that they have undergone a recategorization of plants into a new ontological category – a *conceptual change*.

In sum, children come to school with the necessary prerequisites for scientific thinking, but they also bring with them strong beliefs about how the natural world works and these "theories" are often deeply flawed. As a result, science educators face three important challenges. First, although all educators are responsible for teaching that can be described as "gap-filling" (because of incomplete knowledge), they also have to deal with the many documented scientific misconceptions that are robust and continue into adulthood (Zimmerman & Cuddington, 2007). Second, children's understanding of the procedures used by scientists to investigate the world, such as the design of unconfounded experiments, is subject to conceptual difficulties and misconceptions (Chen & Klahr, 2008; Matlen & Klahr, 2013). Finally, learning to conduct scientific investigations involves the coordination of many science practices (see Table 4.1), which require extended instruction and refinement over the school years.

How Can Children's Scientific Thinking Be Taught?

Not only does scientific thinking require the development of several fundamental precursors, such as the ability to think representationally and to develop and revise causal theories about the world, but it is also highly culturally and educationally mediated (Zimmerman & Croker, 2014). Thus, the essential goal of science education can be characterized as having three primary aspects.

i. Fostering conceptual thinking in science: The aim here is to teach children something about scientific knowledge, that is, what has been learned from tens of thousands of years of human efforts to better understand the natural world.

This is the *product* of science and it includes a vast knowledge base. As noted at the beginning of the chapter, one of the challenges of science education is the fact that rather than coming to school with *no* knowledge of specific science concepts, children come with deep and often robust misconceptions (e.g., children – and many adults, alas! –often fail to distinguish between mass and weight, or they believe that the sun circles the Earth, or that energy and force are the same thing). These misconceptions must be detected and remediated based on an understanding of how misconceptions can be changed in the face of new information.

ii. Fostering procedural thinking in science: To teach children some of the ways that humans have devised to *acquire* that knowledge. That is, to teach them some of the fundamental *processes* of science. Here, too, children often have deep misconceptions. For example, with respect to experimental procedures, young children, even as late as their middle school years, do not know how to distinguish confounded experiments from unconfounded experiments (Chen & Klahr, 1999; Kuhn et al., 1995).

iii. Fostering the ability to apply "School Science Knowledge" to "Everyday Scientific Thinking": That is, to encourage and facilitate children's ability to use what they have learned about scientific products and scientific processes to extend and enrich both of them during their day-to-day engagement with the world around them – i.e., to teach children how to use what they know about science to do even more science. Of course this last goal has a wide range of objectives, from simply being able to approach everyday problems by proposing simple hypotheses and being able to identify causal factors ("Why is my basement wet?: Is the downspout at the front of the house clogged?"; "Why am I so jumpy this evening?: Is it because I ate a lot of candy at lunch?") to creating a research project for submission to the International Science and Engineering Fair.[1]

Fostering Conceptual Thinking

Science educators are responsible for ensuring that students understand both the *concepts* (i.e., the "Disciplinary Core Ideas"; NRC, 2012) and the *processes* of science. With respect to understanding the vast number of domain-specific science concepts, the literature is difficult to adequately summarize. As of 2009, Reinders Duit's bibliography of research studies on conceptual change in science had more than 8,400 entries.[2] To illustrate, consider a single chapter on children's understanding of physical science concepts. Hadzigeorgiou (2015) reviews studies of children's ideas about matter, heat, temperature, evaporation, condensation, the water cycle, forces, motion, floating, sinking, electricity, and light. Each of these topics can be further unpacked to constituent subcomponents (e.g., electricity concepts include current, voltage, charge, electrons, resistance, and circuits, to name a

1 https://student.societyforscience.org/intel-isef-2017
2 For the most recent update, see http://archiv.ipn.uni-kiel.de/stcse/

few). Given this daunting proliferation of specific topics, in this section, we briefly focus on key educational strategies that have been used *across* different conceptual domains to support scientific thinking.

As mentioned above, preconceptions significantly influence how children think about and learn scientific concepts. Therefore, facilitating scientific knowledge acquisition requires an understanding of both the preconceptions children bring and how likely that prior knowledge is to change in the face of new information. Work by Kelemen and colleagues (2014) has demonstrated that children's deep scientific misconceptions (in this case, about natural selection) can be remediated by relatively brief engagement in well-designed explicit instruction that directly targets the misconception. Other research has shown that sharpening children's observational skills is important because prior belief may influence what is "observed" (Chinn & Malhotra, 2002; Echevarria, 2003). For example, many children (and adults) assume that a heavy object falls faster than a light object; because it is difficult to observe both objects simultaneously, this ambiguous situation results in expectations of what will be observed, which in turn influences interpretation, generalization, and retention. Chinn and Malhotra (2002) concluded that belief change based on unexpected evidence is possible but making the correct observations is key. Instructional interventions with scaffolding were successful in promoting conceptual change such that children learned how to make observations unbiased by their initial conceptions.

One of the more effective strategies for effecting conceptual change is to (1) present information that conflicts with children's preconceptions and then (2) gradually support the adoption of scientifically accurate knowledge. For instance, a correct understanding of physical forces is a common source of difficulty for students. Many students believe that objects resting on top of a surface (e.g., books resting on a table) exhibit only a downward force (gravity) but not an upward force (the table). To confront this preconception, Minstrell (1992) developed an ingenious intervention to induce cognitive conflict. Specifically, Minstrell prompted students to hold a book in the outstretched palm of one hand, and then asked them to explain the forces that were acting on the object. Most students initially asserted that there was only a downward force (i.e., gravity). Then Minstrell started piling an increasing number of books onto the students' hands until they acknowledged that their hands were exerting an upward force in order to counteract the gravitational force. Using this revised conceptual model of compensating forces, students were able to generalize this new knowledge to new related instances (Minstrell, 1992). A similar strategy was used by Clement (1993), who attempted to anchor students' understanding of the same concept with a series of close analogies. Clement's first example consisted of a spring: Few students disagree that the spring exhibits an upward force if weight is placed on it. Before moving to the example of books on a table, Clement used an intermediate example – books resting on a foam base. The intermediate example acted as a bridge between the spring and the table examples, inducing students to recognize the similarity between the cases. Cognitive conflict has also been used to address procedural misconceptions that students have about how to conduct controlled experiments (Schwichow, Zimmerman, et al., 2016). This study will be described in more detail in the section below on "Early Experimentation Skills." These cases illustrate that students are most likely to undergo conceptual change when new information is progressively sequenced, starting from students' preconceptions and moving in well-designed instructional steps toward the resolution of the misconception. We return to related strategies in later sections.

Other strategies for fostering conceptual change include interventions designed to facilitate the process of making accurate observations and measurements, including the need to teach students that observations can be biased and measurements include error (Chinn & Malhotra, 2002). Koerber, Osterhaus, and Sodian (2017) demonstrated that scaffolds in the form of diagrams and explanations were effective in remediating misconceptions in kindergarteners and 2nd graders. Schuster and colleagues (2017) compared two different epistemic approaches to teaching Disciplinary Core Ideas. They found that both the "guided" and the "direct" instructional approaches, which both involved active student engagement, were effective at promoting conceptual learning gains.

Conceptual change is clearly essential for scientific thinking. Here, we focused on describing just a few educational strategies that may be used across scientific concepts. Shtulman (2017) argues that replacing the intuitive ideas that children have about the world is a key challenge of science education. Shtulman describes the conceptual challenges for a variety of physical and biological science topics (e.g., energy, motion, inheritance, illness). Intuitive theories, whether about growth or the cosmos, have many properties in common (e.g., they tend to be rooted in perceptual features). However, Shtulman (2017) argues that helping students develop more scientifically accurate theories requires educators to analyze individual concepts in depth to ascertain the most effective way to challenge intuitive theories: "Instruction that neglects the domain-specific nature of intuitive theories and their scientific counterparts has about as much chance of working as the chance of a nuclear physicist making an important discovery in immunology" (p. 249).

Fostering Procedural Thinking

As noted previously, the three key cognitive processes involved in scientific thinking that correspond with various science practices (see Table 4.1) can be applied to any domain of science. As children learn to engage in inquiry, these investigation skills can be fostered individually and in concert with other skills. For these scientific practices, we review some of what is known about children's developing abilities along with illustrative research on fostering these types of procedural thinking in science.

Forming and Refining Hypotheses

Of the three cognitive processes of SDDS, search in the hypothesis space has the most in common with conceptual thinking in science, as it typically involves a search of relevant domain-specific knowledge as represented in the hypothesis space. When one is engaged in inquiry or investigation activities, however, hypothesis-space search is instantiated in the service of the scientific practices of asking questions and developing or using models (NRC, 2012).

Asking questions and curiosity. Asking questions is one of the foundational process skills of scientific practice (NRC, 2012). However, rather than viewing science as a process of posing and then finding answers to questions, students often believe that the goal of science is to demonstrate what is already known (Kuhn, 2005), or to see if something "works" or to invent things (Carey et al., 1989). However, asking questions for which the answer is *not* yet known is a crucial element of inquiry that students must learn (Kuhn & Dean, 2005). Students must learn that *question-asking* is a defining feature of science. An essential precursor to asking good questions is *curiosity* (Klahr, Zimmerman, & Jirout, 2011). The fundamental importance of curiosity in science education is indicated by its nearly universal inclusion as a desired "habit of mind" across a variety of influential science curricula, educational standards, and assessment goals (AAAS, 1993; NEGP, 1993, 1995; NAEYC, 2012; NRC, 2000). For example, the National Science Teachers Association's official position statement on early childhood science education[3] recommends that teachers

> recognize the value and importance of nurturing young children's curiosity and provide experiences in the early years that focus on the content and practices of science with an understanding of how these experiences connect to the science content defined in the *Next Generation Science Standards.* (NSTA, 2014, p. 3)

Nevertheless, although "curiosity" is acknowledged to be an essential part of science at all ages and levels of sophistication, it remains a notoriously elusive psychological construct in both the adult (Lowenstein, 1994) and the child (Jirout & Klahr, 2012) literature.

Simple problem-solving tasks that require question-asking have been used to investigate children's ability to recognize specific instances of uncertainty and to evaluate information. Chouinard (2007) and others have demonstrated not only that young children can determine which questions to ask to address uncertainty but also that they can use information yielded by the answers to their questions to resolve it. Other research has examined children's abilities to ask questions in particular domains in order to investigate their understanding of various phenomena. For example, Greif and colleagues (2006) investigated young children's ability to ask domain-specific questions on a structured task. Children were instructed to ask questions about unfamiliar objects and animals, which they were able to do – averaging twenty-six questions asked across twelve pictures. Many questions were quite general, such as "What is it?" Other questions, however, showed that children recognized and understood that different questions should be asked of the different categories (i.e., objects and animals).

Several studies have gone beyond simply demonstrating young children's nascent ability to ask "good" questions about the natural world and have created procedures aimed at increasing the efficacy of such questions. For example, King (1991)

[3] www.nsta.org/about/positions/earlychildhood.aspx

demonstrated that 5th graders could be trained to use strategic questions to guide their cognitive and metacognitive activity during problem-solving with partners, and that when they did use such "good" questions, they learned more about the system they were investigating.

Analogical Thinking and Use of Scientific Models. Analogical reasoning is a very powerful way to form and refine hypotheses and to scaffold scientific understanding. It involves aligning representations based on their shared relations (Gentner, 1983, 2010). When one of the representations is better understood than the other, information from the familiar case (i.e., by convention, termed the "source") can be used to inform the scientist's understanding of the unfamiliar case (i.e., by convention, termed the "target"). For example, an important concept in molecular biology is how enzymes and substrates interact. This concept is easier to understand when it is compared to a lock and key – the key acts as an unlatching mechanism, fitting into the lock to open it, just as the enzyme fits into the substrate to break it apart. By putting these domains into correspondence based on their shared relations, further inferences might be drawn: for example, a specific key only fits a specific lock, therefore, enzymes may only react with specific substrates. This example illustrates the inferential power that can be derived from analogies, even where there is limited knowledge of the target domain.

Scientists frequently use analogies to generate hypotheses and explain scientific phenomena and to interpret and construct scientific models (Dunbar, 1997). Thus, it is not surprising that analogies have played a role in many scientific discoveries. A fitting example is that of Johannes Kepler (Gentner et al., 1997), who observed that the planets farther away from the sun moved in an elliptical path that was slower than that of the planets closer to the sun. To explain this phenomenon, Kepler drew an analogy to light – he reasoned that, just as light is weakened when viewed from a distance, the sun might dispense a moving power (*vis motrix* – an early predecessor to gravity) that becomes weakened when objects are farther away (Gentner et al., 1997). This analogy in turn contributed to his discovery of the laws of planetary motion.

In addition to the "classical" analogies underlying some of the great scientific discoveries, contemporary scientists use analogical reasoning in their everyday practice, such as the design and interpretation of experiments. Dunbar (1995), in his groundbreaking investigations of the thinking processes involved by scientists in several contemporary molecular biology laboratories, discovered that there are two general forms of analogies used by scientists in discussing and explaining their work. Within lab group discussions, the analogies tend to be fairly "local," as when one particularly well-known specific process in the lab is used as the base for analogically interpreting a recent empirical result. In contrast, when describing their work to outsiders (e.g., science reporters in the media), the scientists use "distant" analogies, in which the base domain is fairly familiar to the public and it is used to describe some important features of a new discovery in the lab.

Even early school-age children have at least rudimentary analogical reasoning abilities. However, owing in part to limited domain knowledge (Bulloch & Opfer, 2009; Goswami, 1991, 2001) and a less than fully developed prefrontal cortex

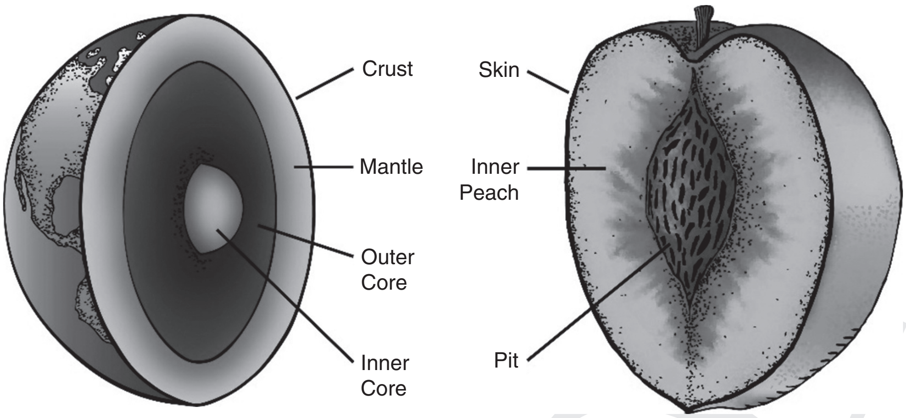


Figure 4.1 *Visual representation of the "Earth is a like a peach" analogy (after Matlen et al., 2011)*

In this analogy, the peach, whose structure is known to the learner, serves as the source and the to-be-learned inner structure of the Earth serves as the target.

(Vendetti et al., 2015; Wright et al., 2008), young children are easily distracted by perceptual features of analogies at the expense of overlooking their relational structure, making them more prone to irrelevant encodings and conceptual misalignments. Despite young children's tendency to overlook relational information, researchers have identified strategies that children can be trained to use to support their scientific reasoning. One strategy is to prompt comparison-making between source and target domains (Christie & Gentner, 2010; Gick & Holyoak, 1983). Having been directly instructed to compare two domains, children are more likely to look beyond superficial features of problems and identify their common relational structure.

The process of comparison-making is facilitated when children's attention is directed toward relevant relationships (Vendetti et al., 2015). Visually representing both the source and the target of the analogy is one way in which comparison-making can be augmented (Gadgil, Chi, & Nokes, 2013; Richland & McDonough, 2010). As an illustration, Matlen and colleagues (2011) aimed to teach elementary-age students geological concepts, such as how mountains and volcanoes are formed. Matlen and colleagues (2011) presented children with analogy-enhanced text passages that were accompanied by either a visualization of both the source and the target (see Figure 4.1) of the analogy or a visualization of just the target (the latter of which is the most common form of presentation in elementary science texts). Matlen and colleagues (2011) found that children were more likely to learn and retain geoscience concepts when text passages were accompanied by visualizations of both the source and the target. Simultaneous visual presentation prompted students to compare the two domains and reduced the cognitive effort of having to recall each representation. Further, the impact of visual aids is enhanced when the to-be-aligned components are made perceptually similar (Jee et al., 2013) or when they are spatially aligned

(Matlen, Gentner, & Franconeri, 2014; Kurtz & Gentner, 2013), such that attention is directed to the corresponding components.

Scientific models also rely on the process of analogical reasoning and are frequently used in science. The source of the analogy constitutes a familiar case, which helps the scientist understand the phenomenon under study. Developing and using models constitute a "signature practice of the sciences" (Quellmalz et al., 2012, p. 366) that is related to the search for hypotheses or explanations. Models are commonly used in science and engineering to support theory-building, argumentation, and explanation (Nersessian, 2008). For example, Watson and Crick's physical model of the structure of DNA, drawings, and schematic representations (e.g., Faraday's sketches of electromagnetic tori or Darwin's tree of life) are cases of models that not only are used by practicing scientists but also provide powerful pedagogical value in teaching these concepts to students.

Because models are of central importance in scientific practice, they are now widely emphasized in science education and science assessment (Clement, 2000; Lehrer & Schauble, 2000, 2012; NRC, 2012). In both science and science education, the ability to develop and use models is becoming increasingly sophisticated due to the scaffolding provided by computers and computer simulations. Simulation models can be used to learn about and investigate phenomena that are "too large, too small, too fast, or too dangerous to study in classrooms" (Quellmalz et al., 2012, p. 367). Science education includes numerous domain-general and domain-specific examples of the instantiation of such model-based practices. Particular curricula are designed around the importance of models, such as the Model-Based Reasoning in Science (MARS) curriculum (Ragavan & Glaser, 1995; Zimmerman, Raghavan, & Sartoris, 2003). Domain-general examples include learning about variability (Lehrer & Schauble, 2004) and decomposition (Ero-Tolliver, Lucas, & Schauble, 2013) and domain-specific examples include evolution in elementary school (Keleman et al., 2014; Lehrer & Schauble, 2012), ecosystems in 6th grade (Lehrer, Schauble, & Lucas, 2008), and biomechanics of the human elbow (Penner et al., 1997; Penner, Lehrer, & Schauble, 1998).

It is important to recognize that although analogies and models have the potential to be powerful teaching tools, teachers must help students to differentiate between perceptual and conceptual similarities between the model and what is being modeled. In learning about multicomponent dynamic systems, young children tend to focus initially on perceptual features rather than on relations among components. However, with appropriate instruction, in which a model's relational structure rather than its perceptual features is reinforced, children can learn how to engage in relatively sophisticated reasoning. For example, Penner and colleagues (1997) asked 1st grade children to construct models of their elbow. Although children's models retained many superficial similarities to the arm (e.g., children insisted that their models include a hand with five fingers, represented by a foam ball and popsicle sticks), children were eventually able to construct models of their elbow that retained functional characteristics (e.g., incorporating the constraint that the elbow is unable to rotate 360 degrees), and were also more likely than a nonmodeling peer group to ignore superficial distractors when identifying functional models. With sustained practice and scaffolding, children can overcome the tendency to attend to superficial similarities and can begin to reason with more abstract models that retain mostly relational structure, such as graphing the relationship between plant growth and time (Lehrer & Schauble, 2004) or modeling variability in nature through a coin flip (Lehrer & Schauble, 2000). Early on, perceptual features serve as an invitation for children to compare the cases at a relational level (Gentner, 2010). Gradually weaning away these irrelevant perceptual features has proven to be an effective way to scaffold understanding with analogies and models (e.g., Clement 1993; Kotovsky & Gentner, 1996).

Investigation Skills: Searching the Experiment Space

Science and engineering practices, such as designing fair tests and interpreting evidence generated from controlled experiments, are included at every grade level from kindergarten through Grade 12 in the Next Generation Science Standards (NGSS Lead States, 2013). The design of an experiment to answer a question or to test a hypothesis can be construed as a problem to be solved, via search in a space of experiments (Klahr, 2000; Newell & Simon, 1972). Of course, experimentation is just one of several types of legitimate scientific inquiry processes that are involved in planning and carrying out investigations (see Lehrer, Schauble, & Petrosino, 2001), but here we focus on the substantial body of literature on ways to improve children's experimentation skills. In the following sections, we first describe research on the developmental precursors of experimentation skills, followed by studies in which participants are engaged in the full cycle of experimentation.

Early experimentation skills. Science education for young children tends to focus on investigation skills such as observing, describing, comparing, and exploring (NAEYC, 2012; NSTA, 2007). Until fairly recently, Piaget's stage theory (e.g., Inhelder & Piaget 1958; Piaget, 1970) was used to justify waiting until adolescence before attempting to teach science process skills (French & Woodring, 2013; Metz, 1995, 1997). However, an accumulation of evidence about human learning (e.g., NRC, 2000) has resulted in a more nuanced story about the developmental course of experimentation and investigation skills and the extent to which well-designed instruction can accelerate that development (NRC, 2007). Learning to conduct experiments involves the coordination of several component processes such as identifying and manipulating variables and observing and measuring outcomes. Not until the later school years, after extended instruction, scaffolding, and practice can children successfully coordinate all of these steps (e.g., Kuhn et al., 2000). Several studies have examined the precursors of the later ensemble of experimentation skills.

One of the fundamental skills in experimental design is the ability to construct a situation in which causal factors can be unambiguously identified. In her classic study, Tschirgi (1980) presented children and adults with a variety of everyday problem-solving situations (e.g., baking cakes, making paper airplanes) that involved a positive or negative outcome and several potential causal variables such as "John baked a cake using honey, white flour, and butter, and it turned out terrible" or "Susan made a paper airplane and it turned out great." The character would propose a hypothesis about a variable that may have caused the outcome (e.g., "John thinks that the honey made it taste bad" in the cake story). The participant in the study would then be asked to select one of three options in order to help the character (John) test the hypothesis. In the vary-one-thing-at-a-time (VOTAT) option, the proposed variable was changed, but the others were kept the same (e.g., bake another cake with everything the same except the sweetener: use sugar instead of honey). This strategy would produce an unconfounded experiment. In the hold-one-thing-at-a-time (HOTAT) option, the hypothesized variable was kept the same but the other variables were changed (e.g., bake another cake with the same sweetener but change the type of flour and shortening). The change-all (CA) option consisted of changing all of the variables (bake a cake with sugar, wholewheat flour, and margarine). All participants were more likely to select the HOTAT strategy when the outcome was *positive*. That is, the presumed causal variable was not changed, while all the other variables *were* changed (thus producing a confounded experiment) in the hope of maintaining the positive outcome. For a *negative* outcome, the logically correct VOTAT strategy (consistent with a controlled experiment) was chosen more frequently than HOTAT or CA, suggesting that participants were searching for the one variable to change in order to eliminate the negative outcome. Although 2nd and 4th graders were more likely to select the CA strategy for the negative outcomes (hoping to eliminate all possible offending variables at once), all participants were influenced by the desire to reproduce good effects and eliminate bad effects by choosing a strategy based on pragmatic outcomes (rather than logical grounds).

Croker and Buchanan (2011) used a task similar to Tschirgi's, but included contexts for which three-and-a-half–year-olds to eleven-year-olds held strong prior beliefs (e.g., the effect of cola vs. milk on dental health). For all age groups, there was an interaction of prior belief and outcome type. The logically correct VOTAT strategy was more likely to be selected under two conditions: (1) when the outcome was positive (i.e., healthy teeth) and consistent with prior belief or (2) when the outcome was negative (i.e., unhealthy teeth) and inconsistent with prior belief. Even the youngest children were influenced by the context and the plausibility of the domain-specific content of the situations that they were reasoning about.

In what has become a classic example of an ingenious study of young children's scientific reasoning, Sodian, Zaitchik, and Carey (1991) presented 1st and 2nd grade children with the challenge of designing a simple experiment to distinguish between two possible causal factors. Children were told that they had to figure out whether their home contained a large mouse or a small mouse. Children were shown "mouse houses" in which they could put some food that mice like. One house had a door through which either a large or a small mouse could pass. The other house had a door that only a small mouse could traverse. In the "find out" condition, the children were asked to decide which house should be used to determine the size of the mouse (i.e., to test a hypothesis). Of course, if the house with the small door is used, and the food is gone in the morning, then only a small mouse could have taken the food. If the food remains, they have a large (and now hungry!) mouse. Importantly, Sodian and colleagues had a second condition, the "feed" condition, in which children were asked what house to use if they wanted to make sure that the mouse would get fed no matter what his size. If a child can distinguish between the goals of testing a hypothesis with an experiment versus generating an effect (i.e., feeding the mouse), then he or she should select the small house in the *find out* condition and the large house in the *feed* condition. Sodian and colleagues found that children as young as six could distinguish between a conclusive and inconclusive experimental test of a simple hypothesis when provided with the two mutually exclusive and exhaustive hypotheses or experiments. Piekny and Maehler (2013) used the mouse house task with preschoolers (four- and five-year-olds) and school children (seven-, nine-, and eleven-year-olds). It was not until age nine that children scored significantly above chance, and not until age seven (a year later than in the Sodian et al. study) that children showed a recognition of, and justification for, conclusive or inconclusive tests of a hypothesis.

Klahr, Fay, and Dunbar (1993) investigated developmental differences in adults' and 3rd and 6th grade children's experimentation skills by presenting them with a programmable toy robot, in which participants first mastered most of the basic commands (see Figure 4.2). They were then challenged to find out how a "mystery key" worked by writing and then running programs that included the mystery key. In order to constrain the "hypothesis space" participants were provided with various hypotheses about the mystery key (only one of which was correct). Some examples of what the mystery key might do include (1) repeat the whole program *N* times, (2) repeat the last step *N* times, and (3) repeat the last *N* steps once. Some of these hypotheses were deemed highly plausible (i.e., likely to be correct) and others were deemed implausible. When presented with a hypothesis that was plausible, all participants set up experiments to *demonstrate the correctness* of the hypothesis (e.g., Experiment 1 in Figure 4.2).

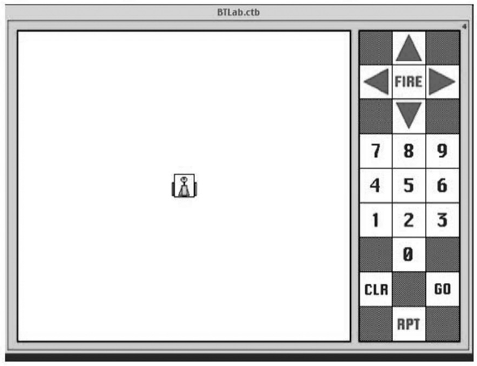
When given an implausible hypothesis to test, adults and some 6th graders proposed a plausible *rival hypothesis* and set up an experiment that would discriminate between the two. The 3rd graders also proposed a plausible rival hypothesis but got sidetracked in the attempt to demonstrate that the rival plausible hypothesis was correct. Klahr, Fay, and Dunbar (1993) identified two useful heuristics that participants used: (1) design experiments that produce informative and interpretable results and (2) attend to one feature at a time. The 3rd and 6th grade children were far less likely than adult participants to restrict the search of possible experiments to those that were informative.

Bullock and Ziegler (1999) collected longitudinal data on participants, starting when they were age eight and following them through to age twelve. They examined the process skills required for experimentation, using separate assessments to tease apart an *understanding* of experimentation from the ability to *produce* controlled experiments. When the children were eight-year-olds they were able to *recognize* a controlled experimental test. The ability to *produce* a controlled experiment at levels comparable to adults did not occur until the children were in the 6th grade. This study

Hx: RPT N repeats the entire program N times.
Hy: RPT N runs program, then repeats the Nth step once.

E1: ↑1 RPT 1

Hx prediction:	↑ 2
Hy prediction:	↑ 2
BT's behavior:	↑2



E2: ↑1 FIRE 2 ↓1 RPT 2

Hx prediction:	↑ 1 FIRE 2 ↓ 1	↑1 FIRE 2 ↓ 1
Hy prediction:	↑ 1 FIRE 2 ↓ 1	FIRE 2
BT's behavior:	↑ 1 FIRE 2 ↓ 1	FIRE 2 ↓ 1

Figure 4.2 *The control panel and two sample programs for discovering how the "mystery key" (labeled "RPT") works on a simulated robot (shown in its "home" position in the center of the screenshot) (after Klahr, Fay, & Dunbar, 1993)*

Hypothesis X (Hx) is that if a number N is appended to the RPT key, then the robot will repeat the entire preceding program N times. Hypothesis Y (Hy) is that the robot will repeat only the Nth step once.

For the first experiment (E1), the participant runs a very short program with only one step preceding the RPT key, that instructs the robot to go forward one unit, and then to do whatever RPT does, one time. Because the program is so short, both hypotheses make the same prediction about the robot's behavior and the robot's behavior is consistent with both.
For the second experiment (E2), a longer program with more steps is used: move forward one unit, fire laser cannon twice, backup 1 unit. Then the RPT 2 command is encountered. Hypothesis X predicts that the entire three-step program preceding RPT will be executed two times. Hypothesis Y predicts that the full program will be executed, followed by one repeat of the 2nd step (the FIRE2 command). The robot's behavior does not match either of these two predictions. Instead, it repeats the last two steps in the program once, thus refuting both hypotheses. In reality, RPT repeats the last N steps, preceding the RPT command, one time.

provides additional support for the idea that young children are able to understand the "logic" of experiments long before they are able to produce them.

When task demands are reduced – such as in simple story problems or when one can select (rather than produce) an experimental test from a set of a few alternatives – even young children show competence with rudimentary science process skills. Children, like adults, are sensitive to the context and the content of what is being reasoned about. Such precursors are important for understanding the challenges of teaching students how to conduct scientific investigations and the types of factors that can be used to facilitate or scaffold developing skills.

Planning and carrying out investigations. Much of the research on the development of investigation skills in older children and adults involves presenting participants with a multivariable causal system, such as physical apparatus or a computer simulation. The participants' goal is to investigate the system so as to identify the causal and noncausal variables in the system; they propose hypotheses, make predictions, plan and conduct experiments, collect and evaluate evidence, make inferences, and draw conclusions in the form of either new or updated knowledge (i.e., although the focus is on cells B or E in Table 4.1, other scientific practices come into play). For example, Schauble's (1996) participants conducted experiments in hydrodynamics, where the goal was to determine which variables have an effect on boat speed.

One foundational, and domain-general, science process skill is the control-of-variables strategy (CVS). The fundamental goal of an experiment is to unambiguously identify causal factors and their effects, and the essential procedure for doing this is to contrast conditions that differ only with the respect to the variable whose causal status is under investigation. *Procedurally*, CVS includes the ability to create experiments in which conditions differ with respect to only a single contrasting variable as well as the ability to recognize confounded and unconfounded experiments. *Conceptually*, CVS involves the ability to make appropriate inferences from the results of unconfounded experiments (e.g., that only inferences about the causal status of the variable being tested are warranted) as well as an awareness of "the inherent indeterminacy of confounded experiments" (Chen & Klahr, 1999, p. 1098). The conceptual aspects of CVS are relevant for argumentation and reasoning about causality in science and everyday life, as CVS includes an understanding of the invalidity of evidence from confounded experiments (or observations) and the importance of comparing controlled conditions (Kuhn, 2005). Thus, CVS is relevant to broader educational and societal goals, such as inquiry, reasoning skills, and critical thinking.

Mastery of CVS is required for successful inquiry learning as it enables students to conduct their own informative investigations. However, without instruction, students – and even adults – have poor inquiry skills (e.g., Kuhn, 2007; for review, see Zimmerman & Croker, 2014). Siler and Klahr (2012) identified the various "misconceptions" that students have about controlling variables. Typical mistakes include (1) designing experiments that vary the wrong (or "nontarget") variable, (2) varying more than one variable, and (3) not varying anything between the contrasted experimental conditions (i.e., overextending the "fairness" idea so both conditions are identical). Methods for determining children's mastery of CVS have varied from the kind of typical high-stakes test item shown in Figure 4.3 to computer-based interactive assessments of the kind shown in Figure 4.4.

A recent meta-analysis of CVS instructional interventions (Schwichow, Croker, et al., 2016) summarized the results of seventy-two studies. Possible moderators of the overall effect size included design features (e.g., quasi-experimental vs.

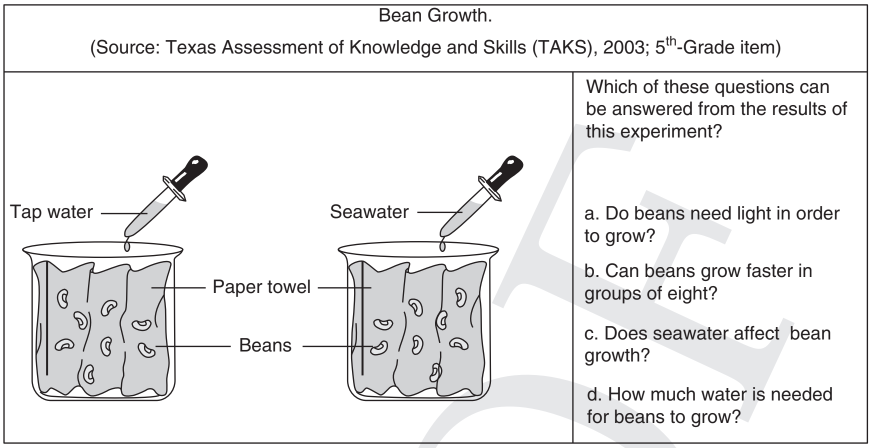


Figure 4.3 *A typical item from a "high-stakes" state assessment of domain-general experimentation skills and knowledge*

experimental studies), instructional features (e.g., use of demonstrations), training features (e.g., use of hands-on experiences), and assessment features (e.g., test format). Of the various instructional features coded for, only two were found to be effective: (1) interventions that induced a *cognitive conflict* and (2) teacher *demonstrations* of good experimental design. In this context, a teacher draws attention to a particular (confounded) comparison and asks what conclusions can be drawn about the effect of a particular variable. For example, to return to the cake baking example described earlier in the section, a teacher might note that although the cake made with butter, wholewheat flour, and sugar tasted much better than the cake made with margarine, white flour, and sugar, one could not tell for sure if the effect was due to the type of flour or the type of sweetener. Because the comparison was confounded, with two possible causal factors, either one of these potential causes might have determined the outcome.

Cognitive conflict is induced in students by drawing attention to a current experimental procedure or interpretation of data; the teacher attempts to get the student to notice that the comparison is confounded or that the conclusion is invalid or indeterminate (Adey & Shayer, 1990). Interestingly, the cognitive conflict technique is often presented *via* a demonstration by the teacher and so additional research is necessary to disentangle the unique effects of these two instructional techniques (Schwichow, Croker et al., 2016). Other instructional techniques that are often presumed to be important, such as the need for "hands-on" engagement with experimental materials, did not have an impact on student learning of CVS. And when hands-on procedures are used, at least one study demonstrated that it does not matter whether students' hands are on physical or virtual materials (Triona & Klahr, 2008). In a follow-up to the meta-analysis, Schwichow, Zimmerman, and colleagues (2016) determined that it is important for there to be a match between the way students learn CVS and the test format used to assess the extent to which they have learned it.

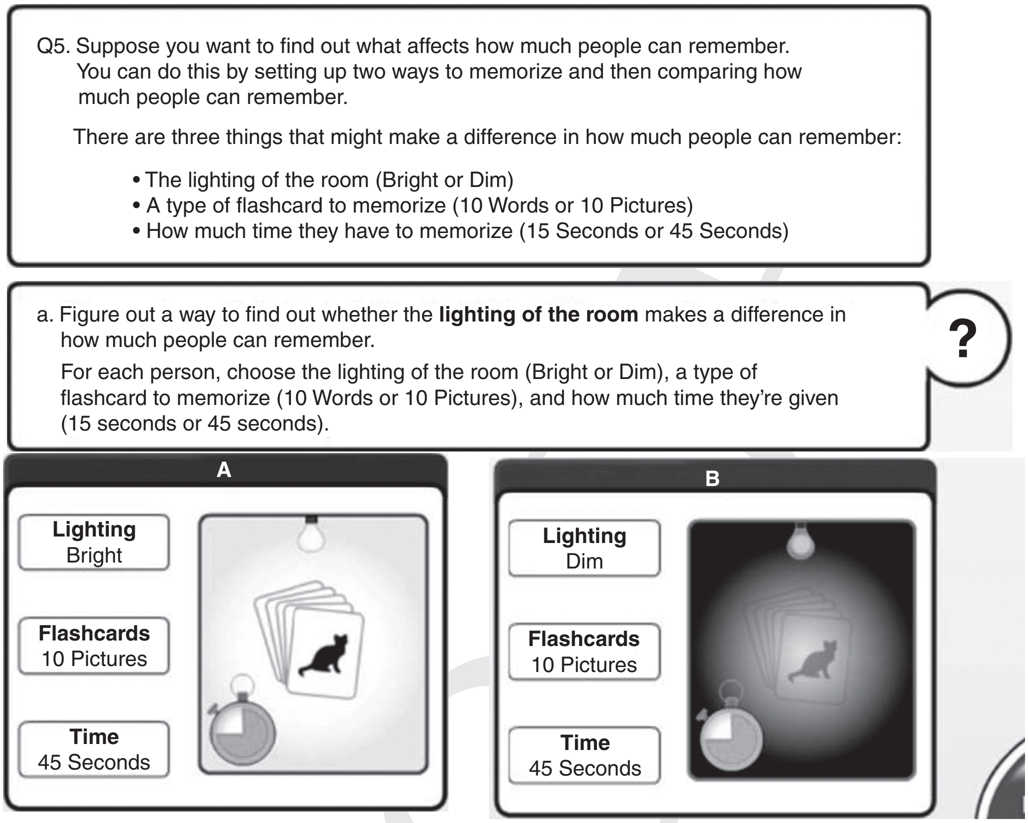


Figure 4.4 *Typical computer-interface item for assessing children's ability to design unconfounded experiments as part of their CVS training*

Evaluating Evidence

The goal of most experiments is to produce evidence that bears on a hypothesis, and once that evidence is generated it must be interpreted. (We say "most" here because, in some cases, scientists may perform experiments in the absence of any clearly articulated hypothesis, just to get a "feel" for the nature of the phenomenon.) The final cognitive process and scientific practices we will discuss are those that enable people to *evaluate, analyze*, and *explain* how evidence relates to the hypothesis that inspired it (i.e., cells C and F in Table 4.1). Evidence evaluation is the part of the cycle of inquiry aimed at determining whether the result of an experiment (or set of experiments) is sufficient to reject or accept a hypothesis under consideration (or whether the evidence is inconclusive), and to construct possible explanations for how the hypothesis and evidence are related.

Evaluating patterns of evidence. One method of examining the developmental precursors of skilled evidence evaluation with children involves presenting them with pictorial representations of potential causes and effects. These are often simple representations such as those between types of food and health (e.g., Kuhn, Amsel, & O'Loughlin, 1988) or plant treatment (e.g., sun, water) and plant health (Amsel & Brock, 1996). The pictures may represent perfect covariation between cause and effect, partial covariation, or no covariation. This cognitive skill is facilitated by the meta-cognitive ability to make a distinction between a hypothesis and the evidence to support a hypothesis (Kuhn, 2005, 2011).

In their classic study, Ruffman and colleagues (1993) presented four- to seven-year-old children with simple story problems involving one potential cause (e.g., type of food: red or green) and an outcome (tooth loss). A "faked evidence task" was used to determine whether children could form different hypotheses based on varying patterns of evidence. For example, children would be shown that green food perfectly covaries with tooth loss: this situation represents the "real evidence." Next, the evidence was tampered with; anyone who was unaware of the original pattern would be led to believe that red food causes tooth loss (i.e., the "faked evidence"). Children were asked to interpret which hypothesis the faked evidence supported. The key advantage of this type of task is that it is diagnostic with respect to whether a child can make a distinction between a hypothesis and a pattern of evidence to support a hypothesis. This task requires children to understand that their own hypothesis would be different from that of a story character who only saw the faked evidence. When considering the responses to *both* the initial veridical-evidence task and the faked-evidence task, only the five- to seven-year-olds performed above chance level. Partial covariation evidence was used to determine if five- to seven-year-olds could form hypotheses based on *patterns* of evidence. When considering both hypothesis-evidence and faked-evidence questions, only the performance of the six- and seven-year-olds was above chance level. Most children understood that veridical versus faked evidence would lead to different beliefs and that a newly formed hypothesis can be used to generalize to future cases.

Ruffman and colleagues showed that some of the very basic prerequisite evidence evaluation skills required for scientific thinking are present as early as six years of age. In follow-up research, Koerber and colleagues (2005) examined the performance of four- to six-year-olds on a variety of evidence evaluation tasks to examine whether *existing causal beliefs* influence evidence evaluation in the preschool years. In situations where there are no strong prior beliefs and the outcomes are equally plausible, preschoolers correctly interpreted perfect and partial covariation evidence. Preschoolers had difficulty, however, with evidence that contradicts prior plausible beliefs; this finding is consistent with the performance of both older children and adults on scientific thinking tasks (Zimmerman & Croker, 2014). Although young children demonstrate some of the precursors to more advanced evidence-evaluation skills, they too are susceptible to the influences of prior beliefs and considerations of the plausibility of what is being evaluated.

From the foundational work of Kuhn and colleagues (1988), we know that the process of revising and acquiring knowledge on the basis of evidence is highly influenced by the prior knowledge that a participant brings to the task. Evaluating evidence is guided by an assessment of the *plausibility* of a hypothesized cause; we make judgments about the world in ways that "make sense" or are consistent with what we already know about how things work. Plausibility is a known constraint in belief formation and revision (Holland et al., 1986) and is a domain-general heuristic that is used to guide the choice of which hypotheses to test and which experiments to run (Klahr et al., 1993). Because the strength of existing beliefs and assessments of plausibility are considered when evaluating evidence, children and adults often choose to maintain their prior beliefs rather than changing them to be in line with newly acquired evidence (e.g., Chinn & Brewer, 1998; Chinn & Malhotra, 2002). A common finding is that it is generally more difficult to integrate evidence that disconfirms a prior causal belief (which involves restructuring one's belief system) than it is to integrate evidence that disconfirms a prior noncausal belief (which involves incorporating a newly discovered causal relation, or "gap-filling"; Chi, 2013). For example, children and adults have robust physics misconceptions about weight, mass, and density, and these misconceptions influence the evaluation of evidence in tasks that involve the motion (e.g., falling, sinking, rolling) of objects. In the case of sinking objects, it is difficult to give up the belief that weight matters but it is easy to add the belief that shape (sphere vs. cube) speeds up or slows down an object based on first-hand evidence (Penner & Klahr, 1996b). Other research shows pervasive difficulties with revising knowledge on the basis of evidence, even when that evidence is generated and observed directly (rather than being provided by researchers; e.g., Chinn & Malhotra, 2002; Renken & Nunez, 2010).

Analyzing and interpreting data. The NRC (2012) science standards include the scientific practice of analyzing and interpreting *data*, noting that "scientific investigations produce data that must be analyzed in order to derive meaning . . . data do not speak for themselves" (p. 51). One important aspect of data interpretation is the capacity to distinguish data patterns that are sufficient to reach a conclusion from data patterns that are ambiguous with respect to a conclusion. Fay and Klahr (1996) described the skill of differentiating determinate and indeterminate situations in terms of children's ability to distinguish "knowing" from "guessing." They found that in a simple logical puzzle context, four- and five-year-old children tend to prematurely terminate search for a cause rather than continuing to search for more information that might render their judgment as incorrect.

The Fay and Klahr research was based on puzzles designed to preclude any empirical "noise." However, an inescapable aspect of real empirical research is that all measurements in the physical world include some degree of error, and children must learn how to deal with this. Masnick, Klahr, and Morris (2007) describe the challenge for the young scientist:

> A young child eagerly awaits the day when she will pass the 100 cm minimum height requirement for riding on the "thriller" roller coaster at her local amusement park. She regularly measures her height on the large-scale ruler tacked to her closet door. As summer approaches, she asks her parents to measure her every week. A few weeks ago she measured 98 cm, last week 99.5 cm, but today only 99.0 cm. Disappointed and confused, when she gets to school she asks the school nurse to measure her, and is delighted to discover that her height is 100.1 cm. Success at last! But as she anticipates the upcoming annual class excursion to the amusement park, she begins to wonder: what is her real height? And more importantly, what will the

> measurement at the entrance to the roller coaster reveal? Why are all the measurements different, rather than the same? Because she is a really thoughtful child, she begins to speculate about whether the differences are in the thing being measured (i.e., maybe her height really doesn't increase monotonically from day to day) or the way it was measured (different people may use different techniques and measurement instruments when determining her height). (p. 3)

Although the processes associated with understanding and interpreting error and data variability draw heavily on mathematical reasoning, and therefore are beyond the scope of this chapter, there are a few studies that capture the intersection of analyzing quantitative data and identifying sources of error. Masnick and Morris (2008) examined how the characteristics of measurement data, such as sample size and variability within the data set (e.g., magnitude of differences, presence of outliers), influenced conclusions drawn by 3rd and 6th graders and adults. Participants were shown data sets with plausible cover stories (e.g., testing new sports equipment) and asked to indicate what conclusions could be drawn and their reasons. The 3rd and 6th graders had rudimentary skills in detecting trends, overlapping data points, and the magnitude of differences. The 6th graders had developing ideas about the importance of variability and the presence of outliers for drawing conclusions from data. At all ages, participants were more confident of conclusions based on larger samples of observations.

Masnick and colleagues (2016) explored how adults and children (aged 9–11) responded to (1) variability in the data collected from a series of simple experiments and (2) the extent to which the data were consistent with their prior hypotheses. Participants conducted experiments in which they generated, recorded, and interpreted data to identify factors that affect the period of a pendulum. In Study 1, several children and most adults used observed evidence to revise their initial understanding but participants were more likely to change incorrect noncausal beliefs to causal beliefs than the reverse. In Study 2, participants were oriented toward either an "engineering" goal (to produce an effect) or a "science" goal (to discover the causal structure of the domain) and presented with variable data about potential causal factors. Science goals produced more belief revision than engineering goals. Numerical data, when presented in context and with appropriate structure, can help children and adults reexamine their beliefs and initiate and support the process of conceptual change and robust scientific thinking.

Constructing explanations. The National Research Council's Framework for Science Education (NRC, 2012) emphasizes the importance of scientific theories and explanations: "The goal for students is to construct logically coherent explanations of phenomena that incorporate their current understanding of science, or a model that represents it, and are consistent with the available evidence" (p. 52). Scientific explanations are typically constructed after investigations that produce evidence that is to be evaluated and, ultimately, explained.

Much has been written in the scientific thinking literature about the ability to differentiate between evidence (i.e., data, observation, patterns) and the explanation or theory that purports to account for that evidence. In particular, Kuhn's

(1989, 2005, 2011) research has emphasized that mature scientific thinking requires the cognitive and metacognitive skills to differentiate between *evidence* and the theory or *explanation* for that evidence. Kuhn argues that effective coordination of evidence and theory depends on three metacognitive abilities: (1) the ability to encode and represent evidence and theory separately, so that relations between them can be recognized; (2) the ability to treat theories or explanations as independent objects of thought (i.e., rather than a representation of "the way things are"); and (3) the ability to recognize that theories can be false and explanations flawed and that, having recognized that possibility, to assess the evidence in order to determine whether the theory is true or false. These metacognitive abilities are necessary precursors to sophisticated scientific thinking and representative of the ways in which children, adults, and professional scientists differ.

As noted previously, children are inclined to notice and respond to causal events in the environment; even infants and young children have been shown to have a rudimentary understanding of cause and effect (Bullock & Gelman, 1979; Piaget 1929). Keil's (2006; Keil & Wilson, 2000) work on the nature of explanation in general indicates that children and adults alike have a propensity to generate explanations. We often privilege causal explanations, which are arguably quite important in scientific thinking. Koslowski's (1996, 2012, 2013) research shows that people are good at noticing evidence for the covariation between events in the world but there is a tendency to only make causal inferences when the link can be explained with a causal mechanism. Participants consider or generate plausible causal mechanisms to explain the relationship between potential causes and their effects. Similarly, if a plausible causal mechanism exists to explain why a cause and effect *should* be linked, it is difficult to let go of that belief. Therefore, we see across many types of scientific thinking tasks that both children and adults have a strong tendency to maintain beliefs rather than change them based on evidence (e.g., Chinn & Brewer, 1998; Chinn & Malhotra, 2002) because the strength of existing beliefs, assessments of plausibility, causal mechanisms, and alternative causal mechanisms are all potentially salient and brought to bear when reasoning (Koslowski et al., 2008).

Conclusions: How Can Research Inform Practice in Science Education?

The studies reviewed in this chapter reveal important insights into the ways in which science is learned. Although there is substantial evidence that children enter the period of school-based instruction with cognitive abilities that are foundational for engaging in scientific reasoning (e.g., Gopnik, Meltzoff, & Kuhl, 2000), these developmental precursors require careful guidance and sustained effort before this nascent knowledge can be transformed into robust scientific practices. Science is a highly educationally and culturally mediated activity that is supported by cultural tools (Morris et al., 2012; Zimmerman & Croker, 2014). Prior research has suggested several ways in which effective instructional strategies can scaffold scientific learning, and it suggests some ways in which these strategies can be applied to classroom teaching and produce meaningful improvements in students' understanding of science. However, even though this body of research has advanced our knowledge about the development of scientific thinking and about how to enable students to understand the body of knowledge about the natural world that has been accumulated, as well as to begin to advance that knowledge beyond their intuitive theories, many challenges remain before a clear-cut reformulation for science teaching can emerge. Scientists and STEM professionals are immersed in years of disciplinary training to overcome the cognitive heuristics and biases that may work well enough in everyday problem-solving but which can lead to incorrect responses or strategies in scientific contexts (Lilienfeld, Ammirati, & David, 2012). Science educators recognize that, insofar as possible, students should be exposed to learning experiences that correspond to the ways in which real science is conducted and communicated (NRC, 2007, 2012). The problem is how best to engage students in inquiry activities that are simultaneously (1) developmentally appropriate (i.e., that do not exceed their cognitive capacity) and (2) scientifically authentic (i.e., that do not distort or oversimplify the fundamental scientific ideas being taught).

There have been several attempts to summarize findings on science education research in such a way that they can be used by classroom teachers to refine their instructional practices based on solid scientific evidence. For example, in 2007, the National Research Council published the report of a blue-ribbon panel of science education experts, entitled *Taking Science to School* (NRC, 2007), that has since been cited thousands of times in the research literature. However, they also published a companion volume, *Ready, Set, SCIENCE!* (NRC, 2008) that was carefully crafted to be accessible to, and used by, practicing science teachers from K-8. Similar NRC reports summarize the current state of the art with respect to translating research on teaching science at the college level into actionable suggestions for college instructors (NRC, 2015; Singer, Nielsen, & Schweingruber, 2012). In addition, the US Office of Education, through its research arm, the National Institute of Education, has produced dozens of "Practice Guides" that distill the basic research in a wide variety of education topics – a few of which include math and science education – into practical, actionable suggestions for classroom teachers.[4] Nevertheless, there remain substantial gaps between findings from scholarly research on how children learn science and how curricula and instruction are typically enacted in real-world settings. We outline some possible reasons for these gaps and some ideas for how to bridge them.

Findings from research need to be more broadly accessible. Currently, researchers are incentivized to publish in peer-reviewed academic journals. However, the audiences for such journals are other highly trained researchers with deep content

[4] The Practice Guides are available at: https://ies.ed.gov/ncee/wwc/PracticeGuides

knowledge and methodological and statistical backgrounds in cognitive and developmental psychology. Researchers often speak a language that is not easily comprehended by the general public, including teachers. For research to make a broader impact on educational practice, findings from research should be accessible to curriculum developers, educators, and policy makers. The need for this type of communication between researchers and teachers has been acknowledged by initiatives such as that by the National Science Teachers Association to identify "Research Worth Reading." The Publications Advisory Committee for the National Association of Research in Science Teaching selects a set of published articles for a teacher audience that have the potential to inform the practice of science teaching.

One way to advance this goal would be for research to be conducted in collaboration with science educators as true partners, rather than as entrée points to students for research projects. Although researchers and teachers often share similar goals, those goals are not entirely overlapping and, in some cases, are at variance with one another. For example, researchers are often interested in advancing theory and in producing generalizable findings, whereas teachers are mainly interested in what is feasible for their classroom application (i.e., "what works"). Nevertheless, their common ground is that both are interested in finding strategies that enhance student learning. Incentives should be in place to support research collaborations between both educators and researchers, so that the goals of both are aligned. We believe these kinds of collaborations will be the most fruitful in fostering work that is conducted at the most appropriate grain size to be impactful for educational practice and over the most educationally meaningful time periods.

References

AAAS (American Association for the Advancement of Science). (1993). *Benchmarks for scientific literacy: A project 2061 report*. New York: Oxford University Press.

Adey, P. & Shayer, M. (1990). Accelerating the development of formal thinking in middle and high school students. *Journal of Research in Science Teaching*, *27*, 267–285.

Amsel, E. & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development*, *11*, 523–550.

Bulloch, M. J. & Opfer, J. E. (2009). What makes relational reasoning smart? Revisiting the perceptual-to-relational shift in the development of generalization. *Developmental Science*, *12*(1), 114–122.

Bullock, M. & Gelman, R. (1979). Preschool children's assumptions about cause and effect: Temporal ordering. *Child Development*, *50*(1), 89–96.

Bullock, M. & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider (eds.), *Individual development from 3 to 12: Findings from the Munich longitudinal study* (pp. 38–54). Cambridge: Cambridge University Press.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989) "An experiment is when you try it and see if it works": A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, *11*, 514–529.

Chen, Z. & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variables strategy, *Child Development*, *70*(5), 1098–1120.

Chen, Z. & Klahr, D. (2008). Remote transfer of scientific-reasoning and problem-solving strategies in children. *Advances in Child Development and Behavior*, *36*, 420.

Chi, M. T. (2013). Two kinds and four sub-types of misconceived knowledge, ways to change it, and the learning outcomes. In S. Vosniadou (ed.), *International handbook of research on conceptual change* (pp. 49–70). New York: Routledge.

Chinn, C. A. & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, *35*, 623–654.

Chinn, C. A. & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, *94*, 327–343.

Chouinard, M. M. (2007). Children's questions: A mechanism for cognitive development. *Monographs of the Society for Research in Child Development*, *72*, 1–129.

Christie, S. & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *The American Journal of Physics*, *50*(1), 66–71.

(1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching*, *30*(10), 1241–1257.

(2000). Model based learning as a key research area for science education. *International Journal of Science Education*, *22*, 1041–1053.

Croker, S. & Buchanan, H. (2011). Scientific reasoning in a real world context: The effect of prior belief and outcome on children's hypothesis testing strategies. *British Journal of Developmental Psychology*, *29*, 409–424.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*(4833), 1556–1557.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (eds.), *The nature of insight* (pp. 365–395). Cambridge, MA: MIT Press.

(1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Viad (eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 461–493). Washington, DC: American Psychological Association.

Echevarria, M. (2003). Anomalies as a catalyst for middle school students' knowledge construction and scientific reasoning during science inquiry. *Journal of Educational Psychology*, *95*, 357–374.

Ero-Tolliver, I., Lucas, D., & Schauble, L. (2013). Young children's thinking about decomposition: Early modeling entrees to complex ideas in science. *Research in Science Education*, *43*, 2137–2152.

Fay, A. & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development*, *67*, 689–716.

French, L. A. & Woodring, S. D. (2013). Science education in the early years. In O. N. Saracho & B. Spodek (eds.), *Handbook of research on the education of young children*, 3rd edn (pp. 179–196). New York: Routledge.

Gadgil, S. M., Nokes, T. J., & Chi, M. T. H. (2013). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*, 47–61.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental P*sychology, *26*(5), 796.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

(2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*, 752–775.

Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B. B., Wolff, P., & Forbus, K. D. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of the Learning Sciences*, *6*, 3–40.

Gick, M. L. & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38.

Gobert, J. D. & Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching*, *36*(1), 39–53.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (2000). *The scientist in the crib: What early learning tells us about the mind*. New York: Harper Collins.

Gopnik, A. & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205–1222.

Goswami, U. (1991). Analogical reasoning: What develops? A review of research and theory. *Child Development*, *62*(1), 1–22.

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science*, *17*, 455.

Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, *28*(9), 799–822.

Hadzigeorgiou, Y. (2015). Young children's ideas about physical science concepts. In K. C. Trundle & M. Sackes (eds.), *Research in early childhood science education* (pp. 67–97). Dordrecht: Springer.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction*. Cambridge, MA: MIT Press.

Hopkins, E. J., Dore, R. A., & Lillard, A. S. (2015). Do children learn from pretense?, *Journal of Experimental Child Psychology*, *130*, 1–18. http://doi.org/10.1016/j.jecp.2014.09.004

Horton, C. (2007) Student alternative conceptions in chemistry. *California Journal of Science Education*, *7*(2), 23–38.

Inagaki, K. & Hatano, G. (2008). Conceptual change in naïve biology. In S. Vosniadou (ed.), *International handbook of research on conceptual change* (pp. 240–262). New York: Routledge

Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: Analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, *14*(2), 175–187.

Jirout, J. & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, *32*, 125–160.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227–254.

Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, *65*(2), 103–135.

Keil, F.C. & Wilson, R.A. (eds.) (2000). *Explanation and cognition*. Cambridge, MA: MIT Press.

Kelemen, D., Emmons, N., Seston, R., & Ganea, P. (2014). Young children can be taught basic natural selection using a picture storybook intervention. *Psychological Science*, *25*, 893–902.

King, A. (1991). Effects of training in strategic questioning on children's problem-solving performance. *Journal of Educational Psychology*, *83*, 307–317.

Klahr, D. (1994). Searching for the cognition in cognitive models of science. *Psycoloquy*, *5* (94), 7–13.

(2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Klahr, D. & Carver, S. (1995). Scientific thinking about scientific thinking. *Monographs of the Society for Research in Child Development*, *60*(4), 137–151.

Klahr, D. & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science*, *12*, 1–48.

Klahr, D., Fay, A., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *25*, 111–146.

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, *333*, 971–975.

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*, 141–152.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.

(2012) How explanation makes information evidentially relevant. In R. Proctor & J. Capaldi (eds.), *The psychology of science: Implicit and explicit processes* (pp. 112–136). New York: Oxford University Press.

(2013). Scientific reasoning: Explanation, confirmation bias, and scientific practice. In G. Feist and M. Gorman (eds.), *Handbook of the psychology of science* (pp. 151–192). New York: Springer.

Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information becomes evidence when an explanation can incorporate it into a causal framework. *Cognitive Development*, *23*, 472–487.

Kotovsky, L. & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*(6), 2797–2822. https://doi.org/10.2307/1131753

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674–689.

(2005). *Education for Thinking*. Cambridge, MA: Harvard University Press.

(2007). Reasoning about multiple variables: Control of variables is not the only challenge. *Science Education*, *91*, 710–726.

(2011). What is scientific thinking and how does it develop? In U. Goswami (ed.), *Handbook of childhood cognitive development*, 2nd edn (pp. 497–523). Oxford: Wiley-Blackwell.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*, 495–523.

Kuhn, D. & Dean, D. (2005). Is developing scientific thinking all about learning to control variables?, *Psychological Science*, *16*, 866–870.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, *60*, v–128

Kurtz, K. J. & Gentner, D. (2013). Detecting anomalous features in complex stimuli: the role of structured comparison. *Journal of Experimental Psychology: Applied*, *19*(3), 219.

Lehrer, R. & Schauble, L. (2000). Developing model-based reasoning in mathematics and science. *Journal of Applied Developmental Psychology*, *21*, 39–48.

(2004). Modeling natural variation through distribution. *American Educational Research Journal*, *41*, 635–679.

(2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education*, *96*, 701–724.

Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. *Cognitive Development*, *23*, 512–529.

Lehrer, R., Schauble, L., & Petrosino, A. J. (2001). Reconsidering the role of experiment in science education. In K. Crowley, C. Schunn, & T. Okada (Eds.). *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 251–277). Mahwah, NJ: Lawrence Erlbaum Associates.

Lewis, E. L. & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. *Journal of Research in Science Teaching*, *31*(6), 657–677.

Li, J., Klahr, D., & Siler, S. (2006). What lies beneath the science achievement gap? The challenges of aligning science instruction with standards and tests. *Science Educator*, *15*, 1–12.

Lilienfeld, S.O., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology*, *1*, 7–36.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98.

Lorch, R.F., Jr., Lorch, E.P., Calderhead, W.J. et al. (2010). Learning the control of variables strategy in higher- and lower-achieving classrooms: Contributions of explicit instruction and experimentation. *Journal of Educational Psychology*, *1*, 90–101.

Masnick, A. M., Klahr, D., & Morris, B. J. (2007). Separating signal from noise: Children's understanding of error and variability in experimental outcomes. In M. Lovett & P. Shaw (eds.), *Thinking with data* (pp. 3–26). Mahwah, NJ: Erlbaum.

Masnick, A. M., Klahr, D, & Knowles, E. R. (2016). Data-driven belief revision in children and adults. *Journal of Cognition and Development*, *18*(1), 87–109.

Masnick, A. M. & Morris, B. J. (2008). Investigating the development of data evaluation: The role of data characteristics. *Child Development*, *79*, 1032–1048.

Matlen, B.J., Vosniadou, S., Jee, B., & Ptouchkina, M. (2011). Enhancing the comprehension of science text through visual analogies. In L. Carlson, C. Holscher, and T. Shipley (eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 2910–2915).

Matlen, B.J., Gentner, D., & Franconeri, S. (2014). Structure mapping in visual comparison: Embodied correspondence lines? Poster presented at the 37th Annual Conference of the Cognitive Science Society, Pasadena, California, July 22–25.

Matlen, B & Klahr, D. (2013) Sequential effects of high and low instructional guidance on children's acquisition and transfer of experimentation skills. *Instructional Science*, *41*, 621. https://doi.org/10.1007/s11251-012-9248-z

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research*, *65*, 93–127.

(1997). On the complex relation between cognitive developmental research and children's science curricula. *Review of Educational Research*, *67*, 151–163.

Minstrell, J. (1992). Facets of students' knowledge and relevant instruction. In R. Duit, F. Goldberg, & H. Niedderer (eds.), *Research in physics learning: Theoretical issues and empirical studies, proceedings of an international workshop* (pp. 110–128). Kiel: IPN.

Morris, B. J., Croker, S., Masnick, A. M., & Zimmerman, C. (2012). The emergence of scientific reasoning. In H. Kloos, B. J. Morris, and J. L. Amaral (eds.), *Current topics in children's learning and cognition* (pp. 61–82). Rijeka: InTech.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *The Quarterly Journal of Experimental Psychology*, *29*, 85–95.

(1978). Consequences of confirmation and disconfirmation in a simulated research environment. *The Quarterly Journal of Experimental Psychology*, *30*, 395–406.

NAEYC (National Association for the Education of Young Children). (2012). All criteria document. https://www.naeyc.org/our-work/families/10-naeyc-program-standards

NEGP (National Education Goals Panel). (1993). *Promises to keep: Creating high standards for American students*. Washington, DC: National Education Goals Panel.

(1995). *The 1995 national education goals report: Building a nation of learners*. Washington, DC: National Education Goals Panel.

Nersessian, N. J. (2008). *Creating scientific concepts*. Cambridge MA: MIT Press.

Newell, A. & Simon, H. A. (1972). *Human problem solving*. New York: Prentice Hall.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press.

NRC (National Research Council). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.

(2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.

(2008). *Ready, Set, SCIENCE! Putting Research to Work in K-8 Science Classrooms*. Washington, DC: National Academies Press.

(2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

(2015) *Reaching students: What research says about effective instruction in undergraduate science and engineering.* Washington, DC: National Academies Press. https://doi .org/10.17226/18687

NSTA (National Science Teachers Association). (2007). NSTA position statement: The integral role of laboratory investigations in science instruction. www.nsta.org/ about/positions/laboratory.aspx. Accessed March 18, 2013.

Opfer, J. E. & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, *49*(4), 301–332.

Osterhaus, C., Koerber, S., & Sodian, B. (2017). Scientific thinking in elementary school: Children's social cognition and their epistemological understanding promote experimentation skills. *Developmental Psychology*, *53*(3), 450–462.

Penner, D. E., Giles, N. D., Lehrer, R., & Schauble, L. (1997). Building functional models: Designing an elbow. *Journal of Research in Science Teaching*, *34*(2), 125–143.

Penner, D. E. & Klahr, D. (1996a). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development*, *67*, 2709–2727.

(1996b). When to trust the data: Further investigations of system error in a scientific reasoning task. *Memory and Cognition*, *24*, 655–668.

Penner, D. E., Lehrer, R., & Schauble, L. (1998). From physical models to biomechanics: A design-based modeling approach. *Journal of the Learning Sciences*, *7*(3–4), 429–449.

Piaget, J. (1929*). The child's conception of the world*. London: Kegan Paul.

(1970). *Genetic epistemology*. New York: Columbia University Press.

Piekny, J. & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, *31*, 153–179.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching*, *49*, 363–393.

Raghavan, K. & Glaser, R. (1995). Model–based analysis and reasoning in science: The MARS curriculum. *Science Education*, *79*(1), 37–61.

Rakison, D. H., Lupyan, G., Oakes, L. M., & Walker-Andrews, A. S. (2008). Developing object concepts in infancy: An associative learning perspective [Monograph]. *Monographs of the Society for Research in Child Development*, *73*(1), i–127.

Renken, M. D. & Nunez, N. (2010). Evidence for improved conclusion accuracy after reading about rather than conducting a belief-inconsistent simple physics experiment. *Applied Cognitive Psychology*, *24*, 792–811.

Richland, L. E. & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, *35*(1), 28–43.

Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, *64*, 1617–1636.

Samarapungavan, A., Mantzicopoulos, P., & Patrick, H. (2008). Learning science through inquiry in kindergarten. *Science Education*, *92*, 868–908.

Santau, A. O., Maerten-Rivera, J. L., Bovis, S., & Orend, J. (2014). A mile wide or an inch deep? Improving elementary preservice teachers' science content knowledge within the context of a science methods course. *Journal of Science Teacher Education*, *25*, 953. https://doi.org/10.1007/s10972-014-9402-3

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, *32*, 102–119.

Schuster, D., Cobern, W. W., Adams, B. A., Undreiu, A., & Pleasants, B. (2017). Learning of core disciplinary ideas: Efficacy comparison of two contrasting modes of science instruction. *Research in Science Education*, *43*, 1–47.

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). The control-of-variables strategy: A meta-analysis. *Developmental Review*, *39*, 37–63.

Schwichow, M., Zimmerman, C., Croker, S., & Härtig, H. (2016). What students learn from hands-on activities. *Journal of Research in Science Teaching*, *53*, 980–1002.

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York: Basic Books.

Siler, S. & Klahr, D. (2012). Detecting, classifying and remediating children's explicit and implicit misconceptions about experimental design. In R. W. Proctor & E. J. Capaldi (eds.), *Psychology of science: Implicit and explicit processes* (pp. 137–180). New York: Oxford University Press.

Siler, S., Klahr, D., & Matlen, B. (2013). Conceptual change in experimental design: From engineering goal to science goals. In S. Vosniadau (ed.), *Handbook of research on conceptual change*, 2nd edn. New York: Routledge.

Singer, S. R., Nielsen, N. R. & Schweingruber, H. A. (eds.) (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering.* Washington, DC: National Academies Press.

Sloutsky, V. M. & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*(2), 166–188.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, *60*(2), 143–171.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, *62*, 753–766.

Sutherland, S. L. & Friedman, O. (2013). Just pretending can be really learning: Children use pretend play as a source for acquiring generic knowledge. *Developmental Psychology*, *49*(9), 1660–1668. http://doi.org/10.1037/a0030788

Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, *24*(4), 357–368.

Triona, L. & Klahr, D. (2008). Hands-on science: Does it matter what the students' hands are on? *The Science Education Review*, *6*(4), 126–130.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, *51*, 1–10.

Uttal, D. H., Fisher, J. A., & Taylor, H. A. (2006). Words and maps: Children's mental models of spatial information acquired from maps and from descriptions. *Developmental Science*, *9*(2), 221–235.

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain, and Education*, *9*(2), 100–106.

Vosniadou, S. (ed.) (2013). *International handbook of research on conceptual change*. 2nd edn. New York: Routledge.

Vosniadou, S. & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535–585.

(1994). Mental models of the day/night cycle. *Cognitive Science*, *18*, 123–183.

Wiser, M. & Smith, C. L. (2008). Learning and teaching about matter in grades K-8: When should the atomic-molecular theory be introduced. In S. Vosiadou (ed.), *International handbook of research on conceptual change* (pp. 205–239). New York: Routledge.

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience*, *1*(8). http://dx.doi.org/10.3389/neuro.09.008.2007

Zimmerman, C. & Croker, S. (2014). A prospective cognition analysis of scientific thinking and the implications for teaching and learning science. *Journal of Cognitive Education and Psychology*, *13*, 245–257.

Zimmerman, C. & Cuddington, K. (2007). Ambiguous, circular and polysemous: Students' definitions of the "balance of nature" metaphor. *Public Understanding of Science*, *16*, 393–406.

Zimmerman, C. & Klahr, D. (2018). Development of scientific thinking. In S. Ghetti (ed.) *The Stevens' handbook of experimental psychology and cognitive neuroscience*, *Vol. 3: Developmental & Social Psychology* (pp. 223–248). New York: Wiley.

Zimmerman, C., Raghavan, K., & Sartoris, M. L. (2003). The impact of the MARS curriculum on students' ability to coordinate theory and evidence. *International Journal of Science Education*, *25*, 1247–1271.