The Psychology of Scientific Thinking:

Implications for Science Teaching and Learning

Junlei Li

David Klahr

Carnegie Mellon University

Li, J. & Klahr, D. (2006). The Psychology of Scientific Thinking: Implications for Science Teaching and Learning. In J. Rhoton & P. Shane (Eds.) *Teaching Science in the 21st Century*. National Science Teachers Association and National Science Education Leadership Association: NSTA Press.

For correspondence, please contact

Junlei Li

Department of Psychology

Carnegie Mellon University

5000 Forbes Avenue

Pittsburgh, PA 15213

junlei@andrew.cmu.edu

The Psychology of Scientific Thinking: Implications for Science Teaching and Learning

Science education has two primary aims: to teach children about our accumulated knowledge of the natural world and to help them employ the methods, procedures, and reasoning processes used to acquire that knowledge – in other words, to "think scientifically". The content of science education is comprised of examples sampled from the vast and ever expanding collection of knowledge in different domains and disciplines. The process of scientific thinking is generalized from the practices shared across domains and disciplines. Such a generalized form of scientific thinking has received increasing attention from researchers in psychological science. In this chapter, we suggest ways in which the emerging psychological understanding of scientific thinking can inform practical questions in science education.

Our effort to examine science education through the lens of the psychology of scientific thinking is prompted, in part, by the *No Child Left Behind Act* (NCLB) of 2001, in which the phrase "scientifically based research" appears more than 100 times (Traub, 2002). In addition to NCLB, the publication of the National Academy of Science's *Scientific Research in Education* (SRE, Shavelson & Towne, 2002) and its sequel, *Advancing Scientific Research in Education* (Towne, Wise, & Winters, 2004), have generated extensive discourse regarding the definition of scientific research as well as its validity and relevance in educational settings (Berliner, 2002; Darling-Hammond & Youngs, 2002; Erickson & Gutierrez, 2002; Feuer, Towne, & Shavelson, 2002; Pellegrino & Goldman, 2002; St. Pierre, 2002). The key message shared by NCLB and SRE is that educational practice should incorporate "what works" [1] solutions validated by scientifically based research. The problem is that few of the scientifically rigorous psychological studies of scientific thinking were ever intended to provide practical recommendations for what works in the classroom. Instead, these psychological inquiries have asked "what is" scientific thinking in children, laymen adults, and present and past scientists (Klahr & Simon, 1999). Can the psychological investigations of scientific thinking inform practical questions in science education? We think this question is worth exploring for both psychological researchers and science educators.

Given the limited space and scope of this chapter, we choose to delve deeply into just one theoretical description of scientific thinking. For more comprehensive reviews of relevant psychological research, see Kuhn (1997), Lehrer & Schauble (in press), Metz (1995, 1997), and Zimmerman (2000). Our particular theoretical framework can be viewed as an elaboration of an insight voiced by Einstein in his characterization of the relationship between scientific thinking and everyday thinking.

> The scientific way of forming concepts differs from that which we use in our daily life, not basically, but merely in the more precise definitions of concepts and conclusions; more painstaking and systematic choice of experimental material, and great logical economy. (Einstein, 1936/1950, p. 98)

In proposing "precise definitions of concepts" for the scientific thinking process itself, we characterize scientific thinking as a set of cognitive processes underlying general "problem solving" but constrained by the unique features of the scientific problem (Klahr, 2000; Klahr & Dunbar, 1988; Newell & Simon, 1972; Simon, Langley, & Bradshaw, 1981). We use this framework to connect the basic research on scientific thinking with the practice of science teaching, particularly with reference to “inquiry”, the advocated goal and means of science education by National Science Education Standards (NSES, National Research Council, 1996).

Psychological Research and Science Teaching

Psychologists who study scientific thinking seek answers to questions such as, “What is thinking?” and “How does knowledge develop?” Teachers and educational developers seek answers to questions such as, "How should we teach (a particular aspect of) science?" and "Do some teaching methods work better than others?" The former questions – in pursuit of fundamental understanding of human cognition – are characteristic of basic research in the science laboratory. The latter questions – aiming towards usability and efficacy – are characteristic of an engineering design, aimed at solving real-world problems. Ideally, these two types of questions and their respective approaches are complementary in both advancing our understanding of the natural world and engendering practical applications. Scientific research provides causal and correlational information (e.g., factors that stimulate the development of thinking, initial knowledge of children at various ages) helpful to engineer effective applications. In return, real-world implementations reveal obstacles and challenges that energize and enlighten scientific research (Stokes, 1997). However, the integration of educational research and practice has fallen short of its potential for reciprocal, productive, and sustainable interaction (Brown, 1992; Hiebert, Gallimore, & Stigler, 2002; Lagemann, 1996, 2000; Strauss, 1998).

At least in the area of psychology, one primary reason for this lack of integration and mutual influence derives from inherent incompatibilities between the goals and methods of basic psychological research and those of science teaching in classrooms (Klahr & Li, 2005). For example, research in the psychological laboratory has tried to distinguish between domain-specific knowledge and domain-general strategy. Researchers often studied one type of knowledge or the other, but not both simultaneously (Kuhn & Angelev, 1976; McCloskey, 1983; Tschirgi, 1980). The complexity of scientific thinking was divided into a more manageable set of distinguishable processes (e.g., hypothesis generation, experimentation, evidence evaluation) and then each process was studied in relative isolation (Zimmerman, 2000). Furthermore, researchers were typically – albeit not exclusively – interested in examining what, how, and why children think, rather than in how to teach children to think more scientifically (Metz, 1995, 1997). In stark contrast, science teachers generally do not seek to isolate every minute aspect of science in everyday teaching: one does not teach process without any content, or vice versa. Integration across the traditional boundaries of content knowledge and process skills is explicitly emphasized in the standards of science teaching (American Association for the Advancement of Science, 1993; NRC, 1996). In addition, the standards and accountability reforms, by their very definitions, ask the teachers to help students master a predefined set of educational objectives within a predefined period of time (e.g., in grade blocks, 5th to 8th grade) regardless of each child's natural competency or developmental pace. These highly contrastive goals and approaches of researchers and teachers make it difficult to translate researchers' understanding of children's competence (and the lack of competence) in scientific thinking into actual educational practice. Such difficulty is evidenced by much of the misinterpretation and misapplication of psychological research in science education (Metz, 1995, 1997).

In summary, psychological research on scientific thinking cannot readily provide the kind of "what works" information demanded by the NCLB model of research to practice pipeline. To overcome this impasse, we want to explore the possibility that basic psychological research can inform practice without necessarily having to produce "what works" prescriptions. In order to design effective teaching practices, teachers need to understand the goal they are teaching towards. Psychological research, as we have described, adopts goals and employs methodologies for the explicit purpose of seeking such an understanding. Informed by the piece-meal understanding of scientific thinking produced by earlier studies, psychologists over the last two decades have sought to unravel the interdependencies among domain-specific knowledge, experimental strategy, hypothesis generation, and experimentation (Klahr & Dunbar, 1988; Kuhn, Garcia-Mila, Zohar, & Anderson, 1995; Schauble, 1990, 1996; Schauble, Klopfer, & Raghavan, 1991). A much more integrated psychological account of scientific thinking has emerged to describe the interplay between content knowledge and process skills in children, adults, and practicing scientists (Klahr & Simon, 1999; Zimmerman, 2000). We see an opportunity to synthesize the emerging integrative account of scientific thinking with what NSES has broadly defined as the inquiry goal and approach of science education.

Translating psychological theories into educational implications is a daunting and undervalued task. The topic domains studied by researchers, even when relevant, constitute a small subset of the content standards to which teachers are held accountable. Of the few studies which do involve training children to think more scientifically, even fewer training procedures have been validated beyond the one-on-one laboratory studies into actual classrooms (e.g. Chen & Klahr, 1999; Toth, Klahr, & Chen, 2000). In psychological journal articles about scientific thinking, it has been and is still common to find brief and generic “educational implications” sections relegated to the very end of each paper, occupying the space of a footnote rather than that of considerable substance. In this chapter, we expand such a “footnote” with more substantive and pragmatic analyses.

Scientific Thinking as Problem Solving

We first describe a theoretical model of scientific thinking: Scientific Discovery as Dual Search (SDDS). The SDDS model describes scientific thinking by integrating domain-specific knowledge (content) and domain-general strategy (process). This theoretical perspective has guided two decades of empirical research in our own research program (Klahr, 2000; Klahr & Dunbar, 1988; Klahr, Fay, & Dunbar, 1993). While it is by no means the only theory of scientific thinking, SDDS is emerging as a promising framework to synthesize the accumulated research on scientific thinking (Zimmerman, 2000).

The defining feature of SDDS is its conceptualization of scientific thinking as a complex problem-solving process, involving the coordination of hypothesis-search and experiment-search. The claim that scientific discovery is a type of problem solving is neither controversial nor informative unless we go beyond a generic interpretation of "problem solving" as a synonym for "thinking." SDDS is based on the theoretical perspective of Newell and Simon (1972) that defines a problem as consisting of an *initial state* and a *goal state*, between which may exist a hierarchy of intermediate *subgoal* states. For example, to find out which factors influence the period of a pendulum, a middle-school child's initial state may consist of some hunches about pendulums. The goal state is to find out explicitly which hunch is right. The subgoal states may include specifying a hypothesis ("I think the weight of the bob matters"), testing the hypothesis, and then evaluating the outcomes ("The period did not change when I changed the bob's weight.") To accomplish each subgoal, the child needs to know a few *operators*: the set of permissible transformations from one state to another. To test whether the weight of the bob matters, it would be helpful to know how to design a controlled and informative experimental comparison and make a precise prediction from the hypothesis being tested. To determine whether the experimental outcome matters one way or the other, it would be helpful to differentiate experimental error (e.g., random counting errors) from experimental effect (e.g., the difference actually created by changing a variable) and guard against one's own confirmation bias in data interpretation. Executed proficiently, a sequence of operators can result in a solution path from the initial state to the goal state. Operators have *constraints* that must be satisfied before they can be applied. For example, not any controlled experiment would do if your goal is to test the effect of the bob weight. The experiment must compare the weight difference rather than some other variables, like the string length. The total set of states, operators, goals, and constraints is called a *problem space*. The problem solving process can be conceptualized as a search for a path in the entire problem space that progresses from the initial state to the goal state.

Each of the elements of SDDS – initial states, goal states, operators, and constraints – can vary along a well- to ill-defined continuum in a classroom or professional scientific task. For example, in a "mix these chemicals and see what happens" scenario, one has a well-defined initial state (i.e., the set of chemicals), but an ill-defined goal state (i.e., what one is expected to find). Or, in a "use strings and weights to find out what determines the period of a pendulum" scenario, one has well-defined initial and goal states (i.e., the lab materials and the purpose), but ill-defined operators (i.e., how one should design experiments to isolate the causal variable).

Figure 1 and Table 1 summarize how the three phases of SDDS – search hypothesis space, test hypothesis, and evaluate evidence – iteratively advance the scientific thinking process. The output from *search hypothesis space* is a fully specified hypothesis, which is passed forward as the input to *test hypothesis*. The output of test hypothesis is a description of evidence for or against the current hypothesis, based on the match between the prediction derived from the current hypothesis and the actual experimental result. Next, *evaluate evidence* decides whether the cumulative evidence – as well as other considerations – warrants the acceptance, rejection, or continued consideration of the current hypothesis. The rejection or continuing consideration of a hypothesis starts the entire process all over again.

The progression within phase and across phases is driven by a coordinated search in two problem spaces: the hypothesis space and the experiment space. Each space has its own initial states, goal states, operators, and constraints. The hypothesis space starts with an initial state consisting of prior content knowledge. Its goal state is to produce a fully specified and testable hypothesis that can account for some or all of that knowledge in a concise or universal form. When prior knowledge is not sufficient to evoke a testable hypothesis, one may supplement prior knowledge by generating additional experimental outcomes. This is the first of the two places where the search in the experiment space is coordinated with search in the hypothesis space. For brevity, we use the term “experiment” to refer to both active experimental manipulations of variables (e.g., a physics experiment) and procedures of passive observations (e.g., a geological field expedition or astronomical data collection). Within the search hypothesis phase, the goal of searching experiment space is to produce outcomes that will allow some generalization towards a testable hypothesis. The method of experimentation is not tightly constrained. For example, if your goal is to develop some hunches about what makes a pendulum swing faster, then putting everything you think would speed up the swing into the same pendulum design is an efficient and acceptable strategy. The critical testing of these hunches comes later.

Once a testable hypothesis is formed, the problem solving process proceeds to test the hypothesis. This is the second place where searching in the experiment space is required, but the constraint is much more stringent. The goal state in the experiment space is to produce informative and logically unambiguous outcomes against which the hypothesis’ prediction is compared. Generally, the aforementioned engineering approach (i.e., building a fast car) with confounding variables would not be acceptable and would most certainly be uninformative.

Based on cumulative experimental outcomes, the evidence evaluation process makes a decision regarding the acceptance, rejection, or continued consideration of the hypothesis. Both the evidence and prior knowledge constrain decisions in this process. When children, adults, and practicing scientists are reasoning about real world contexts, their prior knowledge imposes strong theoretical biases on their reasoning (Brewer & Chinn, 1994; Tschirgi, 1980). These biases influence not only the initial strength with which hypotheses are held – and hence the amount of disconfirming evidence necessary to refute them – but also the features in the evidence that will be attended to and encoded. In scientific reasoning, such confirmation bias is seen paradoxically: necessary, so that one does not readily change a reasonable theory based only on small amounts of contradictory data; and problematic, when one refuses to change a theory despite the preponderance of evidence.

Following this theoretical introduction, we anticipate a pragmatic question: “So what?” Haven’t science educators always regarded science as “problem solving”, albeit in a generic and less operational sense? Technical jargon aside, what does this model practically add to the 6-step scientific method found on posters in nearly every grade-school science classroom or the more elaborate definition of inquiry or scientific process found in science standards or textbooks? We readily admit that the integration of this theoretical perspective with educational practice is at an early and exploratory stage. In our own research program, we have only begun to conduct laboratory and classroom studies towards this effort since the late 1990s (Chen & Klahr, 1999; Klahr & Li, 2005; Klahr & Nigam, 2004; Masnick & Klahr, 2003; Toth, Klahr & Chen, 2000; Triona & Klahr, 2003). However, based on our experience observing and teaching in the elementary school classrooms, we believe that educational practice can be much informed by the understanding of “what is scientific thinking”. We proceed to suggest why SDDS differs from traditional educational conceptualizations of the scientific method and how it may serve as an organizing framework to connect psychological research to inquiry-based science teaching.

Theory, Standards, and Science Teaching

To connect research to practice, we relate the rest of our discussion to the KWHL chart (Table 2), a tool originally intended to help teachers plan instruction (Kujawa & Huske, 1995; Ogle, 1986). We suggest how SDDS and relevant research can inform the first three questions of the KWHL chart with respect to science education. The fourth question, “what students learned”, is explored in a separate report (Klahr & Li, 2005).

The "K" – What do students know?

Of the four KWHL questions, this is the one for which psychological research has provided the most extensive answers. Many psychological investigations explicitly address this question within one or more aspects of scientific thinking. The emerging consensus by researchers is that even kindergarten children can demonstrate partial competency in various aspects of scientific thinking. A short list of examples includes: preschoolers infer causal relations from data patterns (Gopnik, Sobel, Schultz, & Glymour, 2001); children in 1^{st} and 2^{nd} grade understand the difference between a conclusive and an inconclusive test of a simple hypotheses (Sodian, Zaitchik, & Carey, 1991); children from 6^{th} to 9^{th} grade provide and incorporate causal explanations into their evaluation of evidence (Koslowski, 1996). More impressively, elementary school age children are ready to develop and refine their thinking when provided with carefully designed instruction and/or task environment (Klahr & Chen, 2003; see also Kuhn, 1997; Metz, 1995, 1997). The cumulative research data convincingly argues against the once commonly held notion that elementary or middle school children lack developmental readiness to engage in abstract reasoning about evidence and hypotheses and must be taught science using only "developmentally-appropriate" tasks, such as concrete manipulation of objects, procedural execution, categorization, classification, and description (Metz, 1995; Lehrer & Schauble, in press).

Children's developmental readiness for scientific thinking is impressive, but it should not be mistaken for robust competence. Many studies have found significant gaps between children's and adults' scientific thinking, even though adults are far from being proficient across many aspects of reasoning (Fay & Klahr, 1996; Klahr, Fay, & Dunbar, 1993; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn, Garcia-Mila, Zohar & Anderson, 1995). Children struggle much more than adults in evaluating evidence that contradict their prior beliefs about causal relationships

(Amsel & Brock, 1996; Kuhn et. al., 1988); children lack conscious awareness that contradictory evidence should result in a reexamination of prior hypotheses (Kuhn et. al., 1988); children's use of experimentation to test a hypothesis is often driven by an "engineering" approach to create desirable effects. For example, when children were asked to test which features cause a car to go fast or slow, they designed experiments to create the fastest car instead (Schauble, 1990).

What are the practical implications of these research findings? We suggest that SDDS can be used as an instructional analysis framework to map research findings onto classroom practices. To illustrate, we found two instances of "search experiment space" in an NSES 5th – 8th grade sample lesson. In the sample lesson, a teacher facilitates lesson activities where students are to discover which factors determine the period of a pendulum (NRC, 1996, pp. 146 – 147). To start the lesson, the teacher first engages the students to search the hypothesis space to suggest potential causal variables. Instead of asking the students to evoke such hypotheses directly from prior knowledge, she guides the students to set up pendulums, count swings, and discuss why pendulums swing at different rates. Viewed through a SDDS lens, this teaching strategy is productive in two ways: 1) it enables the students to generate testable hypotheses from data when their prior knowledge may only offer partially specified hypotheses; 2) it helps the teacher to anticipate how the students' causal beliefs may subsequently affect their choice of experimentation and evidence evaluation strategies. Because generating explanatory hypotheses is a natural inclination found in middle-school children (Koslowski, 1996), the teacher could afford to focus less on instruction and more on facilitating the exploration and discussion.

The sample lesson began sounding "too good to be true" when the students, having thus searched the hypotheses space, spontaneously proceeded to test each hypothesis by designing unconfounded experiments (i.e., where only the focal variable is varied and all others held constant). The psychological research suggests that students' strategy would falter precisely at the point of searching experiment space under the constraining goal of hypothesis testing (Chen & Klahr, 1999; Schauble, 1990; Schauble et. al. 1991). Connecting such research findings to this particular junction of inquiry may help the teacher to anticipate the less optimal but more likely scenario in which students, instead of carefully isolating variables, confound experimental variables either in "favor" of their prior beliefs or to create what they consider a "favorable" effect. For example, students would set up a fast swinging pendulum with short string, high release, and light weight bob to compare against a slow swinging pendulum with all the opposite features, but only to conclude that the high drop point is what makes the difference. The stated goal of this NSES sample lesson was for "students to develop an understanding of variables in inquiry and how and why to change one variable at a time" (NRC, 1996, p. 146). This second instance of search experiment space is the critical junction where psychological research would suggest more teacher guidance or instruction. Yet the narrative for this lesson segment was so idealistically understated that it mentioned neither teacher instruction nor guidance.

Understanding what students know in generic terms of developmental "readiness" or "appropriateness" is insufficient to implement effective inquiry in science education. Using the sample lesson above, we suggest that SDDS could serve as a framework to organize the teacher's knowledge of children's particular strengths and weaknesses. Such knowledge needs to come from both research and classroom practice, as science teachers accumulate a wealth of understanding from experience. What SDDS can help to facilitate is a task analysis process by which a teacher structurally examines each component of scientific thinking involved in the lesson plan (e.g., search experiment space, search hypothesis, evaluate evidence) and asks, "Is this well-defined or ill-defined in the students' minds?" If hypotheses are ill-defined, the teacher may choose to add experimentation to allow students to explore and explain. If the experimental strategy to test a hypothesis is ill-defined, the teacher may choose to offer more explicit instruction or guidance. If the strategy to evaluate evidence is influenced by students' strong prior beliefs, the teacher may anticipate students' biased interpretation of contradictory data and help students become consciously aware of their bias. Such a task analysis helps the teacher to anticipate trouble spots, understand dependencies of one thinking process on another, and adjust the degree of constraint and guidance in instruction.

The "W" – What do we want students to know?

NSES uses the term "inquiry" to broadly define both the ends and means of science education and emphasize the integration of individual process skills with content knowledge. This matches the theoretical motivation of SDDS, which is to integrate various scientific thinking skills studied in isolation into a coherent whole along with content knowledge. In fact, the list of "science as inquiry" objectives for 5^{th} through 8^{th} grade content standards (NRC, 1996, pp. 143 – 148) contains many of the same aspects of scientific thinking described by SDDS. So what can SDDS add to the inquiry standards beyond what was already stated in NSES?

The most important contrast between SDDS and inquiry in NSES is how differently each organizes the elements of scientific thinking. NSES, as well as the many state science standards developed during the standards reform, uses lists (bulleted, numbered, or simply organized as such) to define inquiry objectives. SDDS adopts a hierarchical and cyclical representation. Such representational differences are not superficial, because they lead to substantively different conceptualizations of the inquiry goal. Using lists to present complex information has been criticized as being too generic, leaving critical relationships unspecified and critical assumptions unstated (Shaw, Brown, & Bromiley, 1998; cited by Tufte, 2003). While the vagueness and unstated assumptions of NSES as a whole have been critiqued (Donmoyer, 1995; Rodrigeuz, 1997; Shiland, 1998), we as psychological researchers are most concerned by the lack of operational specificity in NSES regarding the relationships among the inquiry components. Without relationships, entities such as prediction, experimentation, hypothesis generation, and evidence evaluation are liable to be treated as a laundry list, with no sense of relative sequences or mutual contingencies. In fact, a common practice adopted by educational developers, publishers, and teachers in response to the standards reform was to literally make and complete checklists to show how curriculum and teaching "align" with the list items in the standards.

We believe that the list representation of inquiry and the subsequent checklist implementation of inquiry standards overlook the fundamental structure of the nature of scientific thinking and ultimately mislead instruction. Without understanding the contingent relationships among the problem states and operators within scientific thinking, a teacher could easily engage a process skill out of context, a practice which NSES explicitly deemphasizes. We find examples of such practice even among the sample lessons included in NSES which are intended to convey exemplary inquiry-based teaching. In a density lesson unit (NRC, 1996, pp. 150 – 153) used to illustrate inquiry teaching standards for 5th through 8th grades, the teacher asks students to make predictions for novel demonstrations on four separate occasions in three lessons. Does that indicate the lessons have helped students to "develop … predictions … using evidence" and "recognize and analyze … predictions"? (NRC, 1996, pp. 147 – 148) Recall that in SDDS, prediction is an operator under the "test hypothesis" phase (see Figure 1), which requires a fully specified and testable hypothesis as input (from the phase "search for hypothesis") and produces as output (towards the phase "evaluate evidence") a matching between the prediction and the experimental outcome. In each of the four cases where the teacher asks students to make predictions, the teacher presents students with new and interesting experimental setups (i.e., the teacher conducted "search experiment space" for the students) without suggesting or asking students to formulate any specifiable hypotheses (i.e., the students are given experiments, but are not required to first “search hypothesis space”). Naturally, the students’ responses focus on the outcomes (e.g., which of several objects will sink in a column of layered liquids) and not on the hypotheses that might explain their predictions (e.g., relative density and buoyancy). While we acknowledge that such practices can be very interesting and engaging to the students, we question whether they improve students’ scientific thinking processes or content knowledge. In this particular instance, the lesson simply did not press the students to link their predictions with some testable hypotheses. Thus, even when the evidence contradicts or confirms prediction, it would unlikely result in students’ revising or affirming their explanatory hypotheses, but more likely push them further along a course of guessing by trial and error. Such use of “prediction” reminds us of the “Will it float?” segment from David Letterman’s Late Show, in which people make guesses about various objects before dropping them into the water, rarely justifying or explaining the underlying reasons. The segment is entertaining, though the people making the guesses do not seem to improve their accuracy over time.

We use this example to illustrate the instructional consequences of viewing scientific thinking as an unordered list of relevant skills, rather than as a set of problem solving goals and operators that are mutually contingent and constrained within explicitly structured relationships. While we offer a critique of NSES with regards to its presentation of “inquiry” objectives, we appreciate that the intent of NSES was to counter the practice of rigidly defining scientific thinking as a sequence of unalterable steps (e.g., state the problem, collect data, communicate results, etc.), as is still seen on posters titled “The Scientific Method” in many science classrooms. Despite our concerns with inquiry standards and some sample lessons, we generally agree with the emphasis NSES places upon integrating various process skills. We suggest that, for the practitioner who aims to understand and implement inquiry, a theoretical model such as

SDDS is a well-specified middle ground between the inflexible 6-step scientific method and the vague bullet lists used by science standards.

The "H" – How do students learn and develop "scientific thinking"?

As researchers, we readily admit that, of the four KWHL questions, the "H" is the one least answerable by the psychological studies of scientific thinking. As we described earlier, most psychological research does not directly ask "what works". Recognizing this limitation, we refrain from discussing "prescriptions" for science teaching, but suggest how SDDS and psychological research can help in examining the "what works" question.

How does one decide the fit between an instructional method and a particular learning goal? In such analysis, SDDS can serve as a framework for connecting psychological research to educational practice. For example, if the goal is to develop students' understanding of the control of variable strategy (as in the NSES sample lesson described earlier, NRC, 1996, pp. 146 – 147), should a teacher let students discover on their own, or guide the students by asking them to justify their experimental design and conclusions, or offer explicit instruction on the concept of good experimental design? Viewing this problem through SDDS lens, we suggest that the discovery process does not provide nearly as much corrective feedback to the search for experiments as it does the search for hypotheses. If a hypothesis is wrong, then evidence collected via proper experimentation would at least contradict its prediction. This offers the learner feedback about the quality of the hypothesis. But if the learner chooses an incorrect experimental strategy that intentionally favors the confirmation of a prior belief, the subsequent "confirmation" would satisfy the goal of confirming one's prior belief and mask the deficiency of the experimental strategy. Therefore, in order to develop sound experimental strategy, the teacher must compensate for such lack of feedback inherent within the discovery process. In our own research, we have found that a brief period of explicit instruction, combined with probing questions and hands-on experiments, is more effective in helping students learn to control variables than either probing or hands-on experiments without explicit instruction (Chen & Klahr, 1999; see review of replication studies by Klahr & Li, 2005). Other researchers have found that, across a period of sustained engagement, children improve their use of controlled comparisons to make inferences without explicit instruction, provided that the discovery task was designed to make desirable outcomes difficult to attain without first understanding of underlying causal relationships (Schauble, 1990, 1996; Schauble, Klopfer, & Raghavan, 1991). Sustained engagement and task complexity can thus offer performance feedback for students to improve experimental strategies.

These examples make a broader point about searching for “what works” in science education. Recently, the long advocated “hands-on science” approach has been challenged by proponents of “direct instruction”. The debate over instructional approaches was particularly polarized in deciding whether California’s science curriculum should require either a maximum or a minimum of 25% hands-on science (Adelson, 2004; Begley, 2004; California Department of Education, 2004; Cavanagh, 2004; “Stand and deliver…or let them discover?”, 2004). We disagree with the perception that instructional approaches such as hands-on science, discovery learning, and direct instruction are mutually exclusive rather than intersecting and complementary. As our examples illustrate, (1) explicit instruction can be particularly useful when the hands-on experience itself offers little corrective feedback; (2) sustained exploration can be effective if the discovery task is explicitly designed to afford performance feedback; and (3) a thoughtful combination of explicit instruction and self-guided exploration, both within the context of hands-on experience, can effectively help children develop both process skills and content knowledge.

Examining the policy debate itself through a SDDS lens, one finds that despite the passionate advocacy on both sides, it is hardly a well-defined debate. The "direct instruction" label is sometimes intended to mean a highly-specified instructional procedure, mainly for reading and math, developed by Engelmann and colleagues (Engelmann & Carnine, 1991); at other times it is used much more diffusely to mean a wide range of teacher-controlled talking, showing, questioning, and demonstrating. "Hands-on science" and "discovery learning" are even less well defined terms that usually allude to various mixtures of physical manipulation, guided inquiry, and student exploration. Such lack of precise operational definitions has allowed earnest, passionate, but empirically ungrounded debates to flourish in the area of science instruction policy. Only a handful of empirical studies have directly compared an operationally defined exemplar of one approach versus another in science education (cf. Klahr & Nigam, 2004). Thus, it seems scientifically premature to even begin a debate (i.e., entering the evaluate evidence phase of SDDS) when the terms of the debate has not yet passed the muster of the "search hypothesis space" phase. Perhaps, in addition to applying SDDS towards science teaching, we may be informed by applying SDDS to the debates about science teaching as well!

The Teacher and the Researcher

If we envision the end goal of improving science education as painting a masterpiece, then the painter (whether teacher or researcher) needs to possess not only great techniques but perceptive vision. A painter spends just as much time seeing the picture as painting it. The champion of "inquiry" in education, John Dewey, remarked, "Abstract thought is imagination seeing familiar objects in a new light." (cited by Prawat, 2002, p. 20) In that spirit, we have described a relatively abstract theoretical framework which has guided and synthesized the psychological studies of scientific thinking. We suggest how such a model may be applied to the examination of the means and ends of science education. By viewing scientific thinking as a model of problem solving, one can use the model's descriptive power (the contingencies, relationships, inputs and outputs) to analyze a topic area, evaluate instruction, and integrate available research findings into instructional design.

While we recognize that our suggestions are theoretical in nature, we arrived at this tentative stage by placing ourselves and our research under the environmental constraints of the classrooms and the policy constraints of standards and accountability (Klahr & Li, 2005). By attempting to translate our rather abstract and theoretical formulation of scientific thinking into instructional analysis, we hope to encourage teachers to become researchers, not in the sense that teachers should do our kind of research in classrooms, but that teachers would supplement their professional knowledge by viewing scientific thinking from the researchers' perspective. We also hope that we have continued along the paths trodden by other basic researchers to put theoretical findings to the test of educational relevancy and usefulness. Before drawing conclusions about "what works" in science education, teachers and researchers may be well served by seeing scientific thinking in a different light – and we could all begin by seeing it through the other's eyes.

Acknowledgements

The work reported here was supported in part by grants from NICHD (HD 25211), NIH (MH19102), the James S. McDonnell Foundation (96-37), the National Science Foundation (BCS-0132315), and the Cognition and Student Learning program at the Institute of Education Science (R305H030229).

References

Adelson, R. (2004). Instruction versus exploration in science learning. *Monitor on Psychology, 35(6)*, 34-36.

American Association for the Advancement of Science (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

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

Begley, S. (2004, December 10). The best ways to make schoolchildren learn? We just don't know. *The Wall Street Journal Online,* p. B1. Retrieved December 10, 2004 from http://online.wsj.com/article/0,,SB110263537231796249,00.html.

Berliner, D.C.(2002). Educational Research: The Hardest Science of All. *Educational Researcher, 31(8)*, 18-20.

Brewer, W.R. & Chinn, C.A. (1994). The theory-laddeness of data: An experimental demonstration. In A. Ram and K. Eiselt, eds., *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society,* pp. 61-65. Hillsdale, NJ: Erlbaum.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of Learning Sciences, 2(2),* 141-178.

California Department of Education (2004, January 29). Curriculum commission: Approval of criteria for evaluating k-8 science instructional materials for 2006 primary action. Retrieved on April 7, 2004 from http://www.cde.ca.gov/be/pn/im/documents/infocibcfirfeb04item01.pdf.

Cavanagh, S. (2004, November 10). NCLB could alter science teaching. *Education Week, 24(11)*, p.1, 12-13.

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.

Collins, A. (1998). National science education standards: A political document. *Journal of Research in Science Teaching, 35(7),* 711-727.

Darling-Hammond & Youngs (2002). Defining “highly qualified teachers": What does “scientifically-based research” actually tell us? *Educational Researcher, 31(9)*,13-25.

Donmoyer, R. (1995). The rhetoric and reality of systemic reform: A critique of the proposed national science education standards. *Theory into Practice, 34(1),* 30-34.

Einstein, A. (1950). Physics and reality. In A. Einstein, *Out of my later years* (pp. 59-111). New York: Philosophical Library. (Original work published 1936).

Engelmann, S. & Carnine, D. (1991). *Theory of instruction: Principles and applications*. Eugene, OR: ADI Press.

Erickson, R. & Gutierrez, K. (2002). Culture, rigor and science in educational research. *Educational Researcher, 31(8)*, 21-24.

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

Feuer, M.J., Towne, L. & Shavelson, R.J. (2002). Scientific Culture and Educational Research. *Educational Researcher, 31(8)*, 4-14.

Gopnik, A., Sobel, D., Schultz, L., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37(5)*, 620–629

Hiebert, J., Gallimore, R, & Stigler, W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher, 31(5)*, 3-15.

Klahr, D. & Chen, Z. (2003) Overcoming the "positive capture" strategy in young children: Learning about indeterminacy. *Child Development, 74*, 1256-1277.

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

Klahr, D. & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology, Special Issue, Science Education in Review*.

Klahr, D. & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15(10)*, 661-667.

Klahr, D. & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, *125(5)*, 524-543.

Klahr, D. (2000). *Exploring Science.* Cambridge, MA: MIT Press.

Klahr, D., Fay, A.L. & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *24(1)*, 111-146.

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

Kuhn, D. (1997). Constraints or guideposts? Developmental psychology and science education. *Review of Educational Research, 67(1)*, 141-150.

Kuhn, D. and Angelev, J. (1976). An experimental study of the development of formal operational thought. *Child Development*, *47*, 697-706.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking*. New York: Harcourt, Brace & Jovanovich.

Kuhn, D., Garica-Mila, M., Zohar, A. & Anderson, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, Serial No. 245, *60(40)*, 1-128.

Kujawa, S., & Huske, L. (1995). *The Strategic Teaching and Reading Project guidebook* (Rev. ed.). Oak Brook, IL: North Central Regional Educational Laboratory.

Lagemann, E.C. (1996). Contested terrain: A history of education research in the United States, 1890-1990. *Educational Researcher, 26(9),* 5.

Lagemann, E.C. (2000). *An Elusive Science: The Troubling History of Educational Research.* Chicago: University of Chicago Press.

Leher, R., & Schauble, L. (in press). *Scientific Thinking and Science Literacy: Supporting Development in Learning Contexts.* Handbook of Child Psychology.

Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*, 67-98.

McCloskey, M. (1983) Naive theories of motion, in D. Gentner and A. Stevens (eds), *Mental Models* (pp.229-324). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research, 65(2)*, 93-127.

Metz, K. E. (1997). On the complex relation between cognitive developmental research and children's science curricula. *Review of Educational Research, 67(1),* 151-163.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

No Child Left Behind Act of 2001. (2002). Public Law 107-110-January 8, 2002. 107th Congress. Washington, DC.

Ogle, D. S. (1986). K-W-L group instructional strategy. In A. S. Palincsar, D. S. Ogle, B. F. Jones, & E. G. Carr (Eds.), *Teaching reading as thinking* (Teleconference Resource Guide, pp. 11-17). Alexandria, VA: Association for Supervision and Curriculum Development.

Pelligrino, J.W. & Goldman, S.R. (2002). Be Careful What You Wish For--You May Get It: Educational Research in the Spotlight. *Educational Researcher, 31(8),* 15-17.

Prawat, R.S. (2002, June-July). Dewey and Vygotsky viewed through the rearview mirror—and dimly at that. *Educational Researcher, 31(5),* 16-20.

Rodriguez, A.J. (1997). The dangerous discourse of invisibility: A critique of the national research council’s national science education standards. *Journal of Research in Science Teaching, 34,* 19-37.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31-57.

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

Schauble, L., Klopfer, L., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 18(9),* 859-882.

Shavelson & Towne (2002) *Scientific Research in Education*. National Academies Press

Shaw, G., Brown, R., & Bromiley, P. (1998, May-June). Strategetic stories: How 3M is rewriting business planning. *Harvard Business Review, 76,* 3-8.

Shiland, T.W. (1998). The atheoretical nature of the national science education standards. *Science Education, 82(5),* 615-617.

Simon, H. A., Langley, P., & Bradshaw, G.L. (1981). Scientific discovery. as problem solving. *Synthese*, *47*, 1-27.

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

St. Pierre, E. A. (2002). "Science" rejects Postmodernism. *Educational Researcher*, *31(8)*, 25-27.

Stand and deliver…or let them discover? (2004, November). *District Administration*, p.79.

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Strauss, S. (1998). Cognitive Development and Science Education: Toward a Middle Level Model. In I. Sigel & K. A. Renninger (Eds., W. Damon, Series Ed.) *Handbook of Child Psychology: Vol. 4. Child Psychology in Practice* (pp. 357-400). New York: John Wiley & Sons.

Toth, E., Klahr, D. & Chen, Z. (2000). Bridging research and practice: A cognitively-based classroom intervention for teaching experimentation skills to elementary school children. *Cognition & Instruction*, *18(4)*, 423-459.

Towne, L., Wise, L. L. & Winters, T. M. (2004). *Advancing Scientific Research in Education.* The National Academies Press: Washington, D.C.

Traub, J. (2002, November 10). No child left behind: does it work? *New York Times.* Retrieved December 9, 2002 from http://www.nytimes.com.

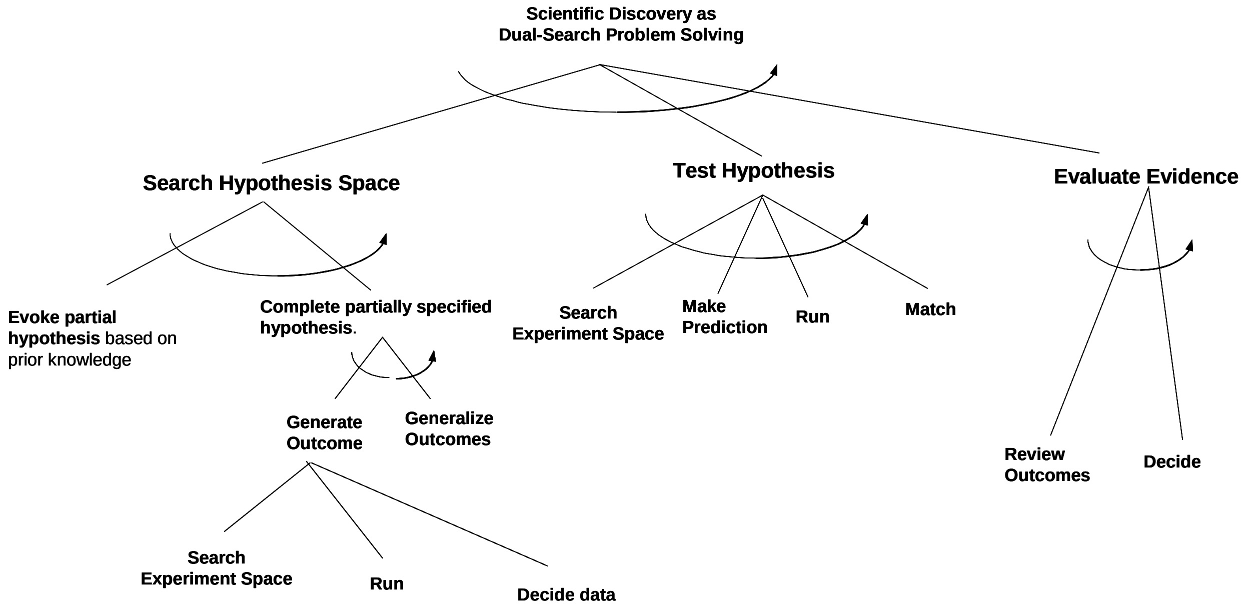
Triona, L. M. & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition & Instruction, 21*, 149-173.

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

Tufte, E.R. (2003). *The cognitive style of powerpoint*. Cheshire, CT: Graphics Press LLC.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20*, 99-149.

Figure 1. *Scientific Discovery as Dual Search* (modified from Klahr, 2000)



Note. See detailed explanations of the terminologies in Table 1.

Table 1: *Scientific Discovery as Dual-Search Problem Solving (SDDS)*

<table>
<tr><td rowspan="5">Search Hypothesis Space</td><td>Evoke partial hypothesis based on prior knowledge</td><td colspan="2">Recall from memory, combined with analogical mapping, heuristic search, priming, and other cognitive mechanisms.</td></tr>
<tr><td rowspan="4">Complete partially specified hypothesis</td><td rowspan="3">Generate Outcome</td><td>Search Experiment Space
Generate some useful data (low constraint) that present intriguing and informative phenomenon. Low constraint: engineering approach acceptable.</td></tr>
<tr><td>Run
Execute experiments or collect data via observation.</td></tr>
<tr><td>Decide data
Accept, reject, or continue collecting data. Rejection reasons include measurement, methodology, or description. Continue collecting if data is not clearly interpretable.</td></tr>
<tr><td>Generalize Outcomes</td><td>If fail to generalize pattern, do more generate outcomes.</td></tr>
</table>

Test Hypothesis	**Search Experiment Space**	High constraint: controlled experimentation, focused data collection, scientific approach preferred, need to inform and discriminate.
	Make Prediction	Constrained by theory, not hunch.
	Run	Execute experiments or collect data via observation.
	Match	Does the prediction match the experimental outcome?
Evaluate Evidence	**Review Outcomes**	Evaluate theory vs. accumulated evidence: how to respond to anomalous data, can theories explain data, are there alternative hypotheses?
	Decide	Accept, reject, modify, or continue?

Table 2

KWHL Chart

What students *know (K)*
What we *want (W)* students to know
How (H) students will find out (about scientific thinking)
What students have *learned (L)*

Note. The "W" question is paraphrased to suit the present discussion. A more common version is: "What students want to know".

Endnote

[1] http://www.whatworks.ed.gov