Designs for Knowledge Evolution:

Methods and Measures of a Prescriptive Learning Theory

Daniel L. Schwartz, Taylor Martin, & Na'ilah Nasir

School of Education, Stanford University

Running Head: Knowledge Evolution

Corresponding Author:
Daniel L. Schwartz
School of Education
485 Lasuen Mall
Stanford University
Stanford, CA 94305-3096
Daniel.schwartz@stanford.edu
(650) 736-1514

[To appear in P. Gardenfors & P. Johansson (Eds.), Cognition, Education, and Communication Technology. Mahwah, NJ: Erlbaum.]

Despite the contention that a better understanding of the mechanisms of thought should lead to better models of instruction, the usefulness of cognitive psychology for the development of productive teaching practices is uncertain. A critical challenge for the field is to develop methods and measures that yield prescriptions, not just descriptions, of learning. An impediment to this challenge is cognitive psychology's common methodology of isolating cognitive mechanisms. The method of work separates the learner from access to non-focal cognitive resources, because exposure might contaminate the isolation of a specific mechanism. Learning complex ideas, however, depends on recruiting multiple cognitive (as well as social and motivational) mechanisms and resources. To make robust prescriptive theories, it is important to consider how people integrate processes and resources to facilitate learning.

Our specific example of an integrative prescriptive theory is called Designs for Knowledge Evolution (DKE). We are unclear whether our prescription constitutes a theory, but it does draw inspiration from biological theory, which has a rich vocabulary for describing change. DKE presupposes that the development of understanding involves the co-evolution of different "species" of knowledge in response to environmental demands. For example, the abilities to perceive and communicate co-evolve in a particular task environment; each shapes the other. Smith and Sera (1992) provide a nice metaphor of child development that captures our emphasis on the co-evolution of cognitive resources. In the context of examining how children learn perceptual words, they state that development is:

> "like the evolution and colonization of an island biotope. Perception and perceptual language can be thought of as two species in this biotope. The

> adaptations of each species clearly depend on each other and all other species on the island. No adaptations can be understood in isolation. Moreover, it makes no sense to ask whether one species *determines* the adaptation of the other. The outcome of development, the structure of the island biotope as a whole and the adaptations of the individual species, is best understood as a dynamic system of continual interaction and mutual influence" (p. 140).

As we will add to this story, the evolution metaphor works even better for human learning if we view the co-evolving species of knowledge as moving from environment to environment. Unlike animals on an isolated island, people move, and this movement helps evolve an understanding that can continue to adapt as it moves beyond the original "habitat" of learning.

The chapter comes in three parts. In the first part, we argue that despite its scientific effectiveness, isolating cognitive mechanisms can blind researchers to significant components and indicators of learning. In particular, we highlight that much of the relevant learning research has tended to focus on how people learn from direct experience or how they learn from communicated experience, but not how people co-evolve the two. We also argue that the outcome measures of learning interventions have not sufficiently looked at people's subsequent abilities to adapt to new environments, and this has led researchers to overlook the value of certain forms of instruction. In the second part, we consider more integrative alternatives, and we turn to the work of developing a method for integrative research. We describe the methods and measures of the DKE framework and defend each empirically. In the third part, we combine the methods and measures of DKE and describe the results of a study that taught children descriptive statistics. The study not only measured the students' abilities to apply what they had learned, it also examined their abilities to evolve new knowledge when placed in new contexts. This latter test of "learning at transfer" is extremely important. The goal of most school-based instruction is not simply to train students to solve a specific class of problem efficiently or to transfer a specific procedure whole cloth to a new context. We argue instead that the goal of school-based instruction should be to prepare students to adapt and learn in the future (Bransford & Schwartz, 1999).

CONGITIVE PSYCHOLOGY AND METHODOLOGICAL ISOLATIONISM

An original catalyst for the growth of cognitive psychology was to handle complex forms of learning that behaviorism could not; for example, perceptual learning (Gibson, 1986), language acquisition (Chomsky, 1966), and hypothesis testing (Levine, 1975). The enterprise has been hugely successful, but it has stalled somewhat at the door of classroom education. Cognitive psychology's successes in designing instruction are swamped by criticisms (Cobb, 1992; Lave, 1988) and defensive replies (Anderson, Reber, & Simon, 1996; Mayer, 1996). The state of affairs suggests that the complexity of classroom learning may exceed cognitive psychology, just as learning with understanding exceeded behaviorism.

We attribute some of cognitive psychology's classroom limitations to its "methodological isolationism." We begin our discussion by reviewing two facets of isolationism: the attempt to study cognitive mechanisms in isolation and the attempt to measure learning in settings isolated from resources for continued learning.

Isolating Cognitive Mechanisms

There are many self-acknowledged limitations to cognitive psychology for developing prescriptive classroom learning theories. These include small effect sizes that are not robust to the natural variability of the classroom, the belief that science should avoid prescriptions of what should be, and a limited consideration of contextual sources of information and interpretation that naturally occur in the highly social environment of the classroom. However, we see a more methodological impediment to cognitive psychology's contribution to a prescriptive learning theory. Cognitive psychologists frequently attempt to dissociate and isolate cognitive systems. They distinguish working and long term memory, implicit and explicit memory, declarative and procedural knowledge, metacognition and problem solving, visual and verbal processing, and many other subsystems. It is a highly analytic endeavor, and the double dissociation is the most prized experimental demonstration. Once psychologists have distinguished a particular cognitive mechanism, they study this mechanism, often to the exclusion of others.

The method of dissociation is very effective. It has revealed distinct cognitive mechanisms and is beginning to locate their neurological basis. However, an emphasis on one system, often to the neglect of another, may not be the best way to develop a prescriptive theory. Learning complex topics, like calculus or car repair, involves many cognitive systems. Isolating one system for study does not explain how that system integrates with others, nor how its development depends on other systems. This shortcoming has been noted by the instructional psychologist Robert Glaser who stated that, "Even if we accept that it will be difficult to achieve a unified theory of learning, we should attempt to discover grounds for the integration of key aspects of human competence that are considered separately" (1992, p. 255).

At the risk of over-simplification, we can illustrate our point with a high-level division that has run through the research literature for many years. This division separates theories that emphasize the acquisition and application of first-hand knowledge and theories that emphasize the acquisition and application of second-hand knowledge. First-hand theories focus on direct experience and second-hand theories focus on descriptions of experience (i.e., communicated knowledge).

First-hand theories depend on people directly interacting with the phenomena of interest. Before children develop skills of interpreting and generating descriptions, they engage their world directly. First-hand theories largely focus on perception and action, and they tend to be more individualistic than second-hand theories, because they emphasize direct, personal experience. Examples include Shepard and Cooper's (1986) and Kosslyn's (1980) studies of imagery, because their emphasis was on how people internally represent perceptual phenomenon. Similarly, Piaget's first-hand theory examined how children abstracted understanding based on their actions with and perceptions of the immediate world. Though Piaget and other first-hand theorists acknowledge the significance of second-hand sources of knowledge and interpretation, it is not the focus of their research and they typically do not contribute to second-hand theorizing and research. For instance, while Piaget studied children's egocentric speech, he did so to make the point that young children are cognitively egocentric, rather than exploring the influence of language and description on cognition.

First-hand theories are highly relevant to learning. No amount of reading is sufficient to learn to drive a car. People need a chance to turn the wheel, feel the acceleration, and hit the breaks in real time. Nevertheless, first-hand theories alone are insufficient for prescribing instruction. For example, embodied theories of cognition emphasize the mental simulation and organization of bodily experience (Barsalou, 1999; Glenberg & Kaschak, 2002; Lakoff & Nunez, 2000). They are a response to the amodal, symbol-processing theories common to early information processing, and they argue that expressions like, “he pushed his way to the top of the company,” or “a force applied to an object…” gain meaning from bodily experiences of pushing. But, they have minimal advice for when external symbols, or simply being told something, should offer significant support for learning.

Second-hand theories of knowledge acquisition depend on individuals interpreting descriptions, often in the absence of the original referent. The second-hand information comes through symbolic forms like language and mathematics, and more recent research is examining multi-media. Second-hand theories centrally locate communicable symbolic structures. For example, Anderson’s (1983) theory of how people convert declarative knowledge or instructions (words) into procedural knowledge (actions) emphasizes the “internalization” or “processing” of second-hand knowledge to permit meaningful action.

Second-hand theories are also highly relevant to learning. An incalculable amount of people’s knowledge comes second-hand from books, and understanding how this happens is important. Nevertheless, we doubt that second-hand theories are sufficient for prescribing instruction. Models of text comprehension (e.g., Mannes & Kintsch, 1987), for example, often focus on the relations between the words within a passage. They are an attempt to describe how people use verbal and textual devices (e.g., pronoun position, capitalization, sentence ordering) to integrate sentences to "comprehend" a text. Such theories, however, provide limited guidance for when people should experience a situation instead of just read about it.

The division between first- and second-hand theories has a strong empirical basis; cognition is not uniform. For example, we asked people to reason about pouring liquid from glasses (Schwartz & T. Black, 1999). There were two glasses of equal height, but different diameters, as shown in Figure 1. Each glass had a line drawn the same distance from the rim to indicate the level of water (though there was no actual water in the glasses). In the Second-Hand condition, subjects could only look at the glasses and had to reason by describing what would happen. They had to decide if the two glasses would start to pour at the same angle of tilt, or whether one glass would pour sooner than the other. In the First-Hand condition, people held a glass (without water), closed their eyes, and tilted the glass until they "saw the water reach the rim of the glass in their imagination." We measured the angle of tilt, and repeated the process with the second glass. We administered the tasks with three different cup shapes using three different subject pools.

[Figure 1 – will the glasses pour at the same angle or at different angles?]

People performed below or at chance in the Second-Hand condition. Moreover, when people worked in pairs and increased their reliance on communicated descriptions, they were never correct (Schwartz, 1999). The First-Hand condition presents a different picture of people's knowledge. Except for one person, everybody we tested correctly showed that the glass with a wider opening would start to pour sooner than a comparable glass with a narrower opening. Figure 2 shows their average tilts for three shapes of glass. This research indicates a dissociation between the processes referred to by first-hand and second-hand theories, and evidently, the processes do not always coordinate with one another (the same people were accurate when they titled but inaccurate when they reasoned verbally). The challenge of this chapter is to acknowledge and take advantage of these different processes.

[Figure 2 – 3 glass tilts]

First-hand and second-hand theories attempt to isolate the cognitive processes that people apply to perceptual-motor content and to symbolic content, respectively. This divide and conquer strategy is a superior way to make scientific progress, until researchers exclusively rely on one or the other to explain complex cognition. First- and second-hand theorists often attempt to generalize their favored mechanism to explain complex cognition and learning. In our case, when we began research on people's abilities to imagine complex devices, we were first-hand theorists. We assumed people's perceptual experiences were the wellspring of all understanding. Through a process of abstraction, perception-based knowledge would turn into explicit description. However, we have failed to find evidence of bootstrapping from first-hand knowledge to second-hand knowledge.

For example, we asked people to reason about chains of gears (Schwartz & J. Black, 1996). We asked, "Imagine a horizontal chain of five gears. If the gear on the far right tries to turn clockwise, what will the gear on the far left do?" Over multiple trials with differing numbers of gears, most people induced a parity rule; for even chains of gears, the first and last gears turn opposite directions. Before people learned the parity rule, they used hand gestures to portray the gears. They depended on their first-hand knowledge. Once they induced the parity rule, they stopped gesturing and relied on second-hand descriptions of the gears, for example, "All the odd gears turn the same direction." At first glance, the results appear to show that people abstracted their first-hand knowledge of object interactions into second-hand descriptions of the larger pattern of gear behavior. Though the gestures seemed to play a necessary role in the rule induction, they were not sufficient. Second-hand representations and processes were critical to the transition from gestural simulations to rule-based descriptions. We examined the differences between people who did and did not induce the rule. People who verbally described the gears' positions with numbers induced the parity rule 100% of the time, whereas people who did not count the gears only induced the rule 26% of the time.

In another study (Schwartz, 1995), we examined the differences between 10th-grade children working alone or in pairs. The children who tried to solve the gear problems alone only induced the parity rule 14% of the time. In contrast, children who worked in pairs induced the rule 58% of the time. This is well above the probability that any of the pairs would have included at least one individual who might have solved the problem in isolation. Evidently, first-hand experiences were not sufficient to induce the rule – people also needed to integrate descriptions using number, often in communication with others.

We have found similar results when looking at the development of proportional reasoning over physical situations. For example, when young children were encouraged to use mathematics to explain problems about a balance scale or involving juice and water, they developed a more complex understanding of torque and concentration than if they just relied on first-hand experience (Schwartz, Martin, & Pfaffman, in press; Schwartz & Moore, 1998). Although it seems obvious in retrospect, brute induction over first-hand experience is not enough to propel the learning and development of complex ideas. People need the structure provided by communicated cultural forms that have been invented over the years (like mathematics) to help organize complexity. At the same time, second-hand knowledge, without some grounding in first-hand experience, does not get people very far either. Children who merely memorize math facts can end up knowing that 3+4 = 7 yet not know that 7 is greater than 4. Understanding requires first- and second-hand experience, and it is important to figure out how to effectively combine the two through instruction.

Isolating the Outcomes of Learning

A second manifestation of methodological isolationism involves measures of learning. One problem is that researchers often measure outcomes associated with a single cognitive mechanism. For example, there are measures of memory that distinguish between subsystems ranging from implicit and explicit memory to semantic and episodic memory. The design of precisely targeted measures is an important skill developed by cognitive psychologists, and it leads to significant progress. However, tests of specific memory functions are only proxy measures of deep understanding, because deep understanding also includes the ability to perceive, plan, act, and transfer. It is important for education to avoid an over-reliance on memory tests, unless the goal is to train people to remember specific procedures and facts under narrow retrieval conditions. As we demonstrate below, memory tests can fail to differentiate those who understand from those who do not.

A second (but related) problem with many measures of learning is that they tend to employ sequestered problem solving (Bransford & Schwartz, 1999). After participants in an experiment learn a target concept, they attempt unaided problem solving or retrieval. Like a jury, they are sequestered from contaminating sources of information as they complete tasks that measure their learning. This way, the researchers can be sure that any differences between learning conditions are due to the experimental treatments and not some learning that surreptitiously occurred during the assessment.

An important question is whether the goal of education is to prepare people to solve problems without access to resources. For example, typical transfer experiments ask whether participants who have learned a task or concept can apply their learning to a new problem without access to additional resources for learning how to solve the problem. The assumption appears to be that transfer involves the direct application of a confined body knowledge without any adaptation or growth of that knowledge in a new situation. This assumption seems warranted when the goal is efficiency and the context of application is highly similar to the context of instruction. However, for situations of adaptation, a reliance on tests of sequestered problem solving seems ecologically suspect (Schwartz, Bransford, & Sears, in press). As an everyday example, when students apply the arithmetic they learn at school to go grocery shopping, they need to learn about the particular characteristics of shopping (e.g., best buy comparisons), and they need to adapt their algorithms to work without pencil and paper. If they only transferred their paper and pencil algorithms, they would make very slow shoppers. Hatano and Inagaki (1986) differentiate adaptive experts, who continue to learn as the times change, from routine experts, who apply the same skill over and over. For most school-based instruction, the goal is to prepare people to adapt to new situations, and therefore, it makes sense to assess the quality of the instruction by students' abilities to learn given resources in a new environment.

In addition to questions of ecological validity, tests of sequestered problem solving raise a methodological concern. Tests of unaided problem solving can be obstacles to determining effective strategies for putting learners on a trajectory towards adaptive expertise. They can blind us to the value of activities that prepare students for future learning but do not immediately show benefits on tests of sequestered problem solving. An example of this point comes from a study that looked at children learning to add fractions (Martin & Schwartz, 2004). Over three days of guided discovery, 5th-grade students learned fraction addition with physical manipulatives that they could move around to aid in their computations and conceptualization. Half of the students worked with tiles pieces, and half of the students worked with pie wedges (e.g., 1/4 wedges, 1/8 wedges, etc.). Given feedback, we found that students in both conditions learned to do fraction addition problems with their material, and the groups did not exhibit any significant differences. In a transfer phase, students from both conditions tried to solve new fraction addition problems without any feedback. Students tried to solve problems in their heads. When they could not solve the problems in their head, we provided them with new manipulatives they had not seen before (e.g., fraction bars that indicate fractions by their length, and beans and cups). To successfully solve the problems with the new materials, the children had to adapt to the new environment and figure out how to use the materials to support their reasoning. In this setting the tile students did much better than the pie students. Over several trials, the tile students learned how to use the new materials, whereas the pie students did not.

The fraction study provides an important lesson. When the students worked with their original learning material or had to solve problems in their head without access to new resources for learning, they looked the same across the groups. However, when we looked at how well they performed in a context with new resources (i.e., new materials), the benefits of the tiles became apparent. Had we only relied on sequestered problem solving, we would have overlooked the value of the tiles for preparing students to learn during transfer. Tests of sequestered problem solving can be too blunt an instrument for measuring early stages of learning and for evaluating the effectiveness of one instructional treatment over another.

INTEGRATIVE APPROACHES TO COGNITIVE RESEARCH

While debates about the primacy of first- and second-hand processes can successfully advance psychological theory and evidence, these debates will not explain how people integrate the two processes. At the level of educational experience, learners require both, whatever the ultimate atomic structure of cognition may be. Tests of isolated problem solving, without evaluations of subjects' abilities to learn in new situations, can advance cognitive psychology. However, isolating learners from an environment of learning during a test misses important aspects of the prior knowledge that prepare people to learn, especially when the goal of instruction should be to prepare people to learn from the resources available in their next classes or once they leave school altogether.

In this section, we consider more integrative approaches to cognitive psychology that we believe can advance both cognitive research and the development of prescriptive learning theories. We review a few integrative approaches, and then we describe and test Designs for Knowledge Evolution. Afterwards, we describe our efforts to develop dynamic assessments (Feuerstein, 1979) that examine whether students are prepared to evolve their knowledge in new situations and learn.

Some approaches to integrative research

Robbie Case (Case & Okamoto, 1996) provides an excellent example of an integrative model of learning. Case's work arose from the careful analysis of children's natural patterns of development. His theory describes development as the integration of core conceptual structures, rather than as the maturation or enhancement of a single knowledge structure. One example comes from his analysis of the development of children's understanding of counting and simple addition. In the early stages of the process, children move their fingers down a line of objects. At the same time, they say a counting word for each action. The interweaving of action, perception, and language permits the child to develop a differentiated understanding, for example, of the cardinal value of five and the ordinal value of the 5^{th} position. Case's model provides important guidelines about developmental readiness for learning in mathematics, the significance of some representations over others, and the importance of games that lead to children's fluency in traversing the many representations and manifestations of quantity.

Another approach to developing an integrative research agenda comes from work on scaffolding. Scaffolding research examines the material situations and the social mediations that lead to successful learning. The method of work is analogous to naturalists who look for native plants that have medicinal properties. Once they discover a naturally occurring medicinal plant, chemists can distill the active ingredients into a potent medicine. Similarly, scaffolding research often builds on everyday instances of material and social supports to seed the development of more precise instructional technologies. The central idea of scaffolding is that, with mediation, students can complete mature activities that they cannot complete themselves. A common explanation for the effectiveness of scaffolding is that it permits the learner to complete the activity, and over time, the learner internalizes the context and practices of the scaffold. The cognitive mechanisms responsible for internalization are themselves often treated as black boxes, which seems acceptable for the level at which these theories operate.

Scaffolding can take many forms. In one form, learners complete an authentic activity with additional physical support. Training wheels is a canonical example. In another form, the support comes from the social structure and active mediation of more knowledgeable others. Apprenticeship is a good example, and this approach has been nicely generalized to instruction. In reciprocal teaching (Palinscar & Brown, 1984), for example, teachers provide models for how to ask important questions as the teacher and students work jointly to comprehend a text. Gradually, students take on more of the responsibility for asking questions, until such time that they can complete the task without the help of the teacher.

A common ingredient of scaffolding is that learners take on partial roles that allow them to learn the form of the activity even though they may not fully understand its function (Saxe, 1991; 1999). For example, in a study of learning in the game of dominoes (Nasir, 2000), we found that experts managed to scaffold novices' partial moves while still maintaining an enjoyable game for themselves. The experts and novices co-created a game structure in which novice players were allowed to choose which tile to play, and on the next move, the expert partner determined where to place the tile. A particularly interesting form of scaffolding involves learner's identity as a social participant (Lave & Wenger, 1991; Nasir, 2002). For example, a young child might wear a carpenter's belt filled with plastic tools. Though the child cannot use the tools to participate in the activity of building an authentic structure, the plastic tools scaffold the child's identity as a builder and conceivably position him or her to learn carpentry in the future.

Scaffolding research offers a nice instance of studying a pre-existing, integrative learning activity to build a prescriptive theory. Scaffolding involves perceiving and acting, plus the communication and interpretation of other people's understandings that can lend significance and structure to first-hand experiences. However, the study of scaffolding has some limitations for the development of prescriptive learning theories. One limitation is that it has followed the natural tendency for precise terms to become amorphous when taken up broadly. Scaffolding has become so pervasive that anything that supports learning is labeled a scaffold. Another, and more foundational limitation is that scaffolding tends to focus on mature performance, so that the measure of successful scaffolding is whether a learner can complete a task unaided. From our perspective, one goal of instruction should be to prepare people to learn in the future. Because of its emphasis on performance, not all scaffolding and apprenticeship models include mechanisms to support future learning and adaptation beyond the performance of the original task.

Final in our survey of integrative methods is work on multiple representations. Research along these lines stems from the belief that the juxtaposition of different representations will lead to a deeper understanding. Kaput (1995), for example, has developed yoked computer simulations. Students see the movement of an object like a car on a computer screen while yoked graphs simultaneously show graphs of the car's acceleration, velocity, and distance. A similar method allows children to make the movement themselves while showing the plots on the computer screen (Nemirovsky, Tierney, & Wright, 1998). These types of juxtapositions often build on important intuitions and assumptions. For example, in the yoked simulations, the underlying assumption appears to be that meaning arises by finding the similarities between first-hand experiences of an event and second-hand representations of the same event. Students, for example, learn that their bodily acceleration maps onto a steeper slope in the velocity graph. However, by itself, the mere mapping of the similarities between representations may not be the most effective way to learn and understand. For example, work on analogical mapping proposes that people learn by mapping a known structure into an unknown structure. For this to be effective, people need to have the known structure to begin with. What if they do not? Ideally, research on multiple representations will develop a principled account for how people can best integrate different forms of knowing and representation, especially when that knowledge is immature to begin with.

Designs for knowledge evolution

DKE is an explicit attempt to join first- and second-hand experiences into a prescription for learning. It begins with well-documented mechanisms that generate specific forms of learning. It then proposes a framework in which the multiple representations and mechanisms can interact to co-evolve a well-rounded understanding that supports future learning and adaptation. We first describe a mechanism for developing second-hand knowledge and then a mechanism for developing first-hand knowledge. We then present our method of integration and an initial test in the domain of statistics instruction. Afterwards, we return to the question of preparing people for future learning.

A powerful and natural mechanism by which people come to describe the world is through the construction of mental and symbolic models. Vosniadou and Brewer (1992), for example, asked young children to draw pictures of the earth. They found that children spontaneously constructed coherent, albeit unconventional, models. For example, children combined their first-hand experience of a flat earth with their second-hand knowledge that the world is round to draw round discs or a flat earth resting in a bowl. These constructions illustrate that people are natural model builders, and it follows that this could be a useful mechanism for enhancing learning.

In the case of the earth, the drawings presumably reflected the internal models the children spontaneously constructed. People are also good at intentionally constructing external models to serve as explicit second-hand descriptions. We asked 7th-grade children to construct visual representations of causal pathways like, "X can communicate the disease to Y," "Q can get the disease from R," "F gets infected by Y" (Schwartz, 1993). Their task was to make representations that could solve problems like, “If X has the disease, what else can get the disease?” The children were quite inventive at building models, and most children represented the many-to-one and one-to-many relationships needed to solve the problems. Figure 3 shows a representative selection of the visualizations the children developed. Interestingly, the opportunity to construct visual models had a lasting effect. Several weeks later, embedded in a class activity, the children spontaneously transferred the idea of using visual representations for a novel problem, though there were no prompts or cues to do so. Even more impressive, over half of the students also tried to invent new visual representations for problems that did not have the same “causal pathway” structure. Appropriate opportunities to build models can prepare students to adapt new structures in novel settings.

[Figure 3 – children’s visual models]

In addition to a mechanism for developing second-hand knowledge, we also need a mechanism for developing first-hand knowledge. A critical form of first-hand knowledge is the ability to perceive. Contrary to the common assumption that perception is untutored, people learn to perceive. For example, novices cannot taste the subtle flavors that differentiate two wines, whereas experts can. A common “expert trap” is to assume that novices can see what the expert refers to. Students and teachers, for example, can use the same words with very different meanings in mind. For example, in a psychology course, teachers may present an instance of recognition memory, label it for the students, and hear the students use the words “recognition memory.” This does not mean the students have noticed what constitutes the phenomenon “recognition memory.”

They may only see the vague phenomenon of "remembering things," and fail to distinguish between recognition memory and free recall, for example.

A significant body of research describes learning to perceive in terms of noticing what differentiates things from one another (E. Gibson, 1969; Marton, this volume). Beiderman and Shiffrar (1987) demonstrated that people who have to determine the sex of baby chicks learn to differentiate the males and females by discerning the distinctive and often subtle features that uniquely identify each sex. In contrast, novices cannot distinguish between male and female chicks, because they do not see the key features. A powerful way to help people notice is to have them examine contrasting cases (Bransford, Franks, Vye, & Sherwood, 1989; Gibson & Gibson, 1955). For example, wine tasting classes ask people to compare one wine against another, so that people can isolate what makes each wine distinctive. Howard Gardner (1982) describes an art exhibit that juxtaposed original paintings and forgeries. At first people cannot tell the difference, but over time, they begin to notice the features that identify the original. Dibble, as reported in Gibson (1969), even found that opportunities to examine contrasting cases of letters enabled people to subsequently recognize the letters better than copying them.

The goal of DKE is to unite the mechanisms of perceptual learning for developing first-hand knowledge with model building for developing second-hand knowledge. Each mode of understanding has different characteristics and supports different insights and inferences. To bring the two ways of knowing into productive interaction, it is not sufficient to simply juxtapose them. The goal of learning is not to find the correlations or mappings between the two, such that they are simply isomorphs in different modalities. Instead, the goal is to find out how the two forms of knowing can complement one another to make a more profound and multi-faceted understanding. For example, in the previous example of people learning about gears, the gestured simulations of the gears provided access to primitive physical intuitions of force and movement, whereas the mathematics provided access to highly structured representations. In combination, they generated a generalized symbolic parity rule grounded in physical experience.

Our approach to developing a prescriptive learning theory is to use the unique strengths of each form of knowing to illuminate the other. We design instruction so that the two forms of knowing co-evolve as they adapt to new task environments. The key features of DKE draw on some of the concepts found in evolutionary theory:

1. <u>The process begins by students producing a new species of symbolic model that can respond to differences between contrasting situations.</u> This permits students to simultaneously perceive what is significant about the features of the contrasting cases and develop a structured account of what they perceive.

2. <u>Students test their models across contexts of contrasting pairs</u>. As they confront new contexts, some models fail and students notice properties of the new contexts and attributes of their models that "selected" against survival.

3. <u>Students mutate new models that can survive in the new context</u>. The new models evolve from the understanding developed from previous models, even if the students need to abandon the form of their earlier model and try a new "genetic" line of models.

4. <u>Students juxtapose their respective models to notice their "survival" value</u>. Whereas the contrasting cases introduce environmental variation, the juxtaposed models

introduce species level variation. Noticing the varying quality and useful features of different models introduces selective pressures to help students generate useful models.

Across the multiple contexts and opportunities for co-evolution, the learner comes to perceive important features of the problem domain while evolving models that can adapt to those and future features. We can best illuminate the first three aspects of DKE with a brief study that helped students learn about the statistical concept of variability (Moore & Schwartz, 1998). We presented college students with a sequence of tightly focused contrasting cases. They had to invent formulas to capture what is different about each pair of cases. For example, the first pair of contrasting cases presented the two distributions: {1 3 5 7 9} versus {3 4 5 6 7}. We pointed out that the two distributions have something in common, namely, the average. We explained that the average is a convenient way to characterize what is common about the distributions. It is much easier to communicate the averages than the complete distributions, especially when the number of items gets very large. We then asked the students to notice that there is also a difference between the two sets of numbers, called the "spread," and to invent a formula that can capture what is different. The students typically invented a range formula that subtracted the smallest number from the largest. We then presented a new contrasting case: {1 3 3 3 9} versus {1 3 5 7 9}. Students saw that their range formula did not differentiate the two data sets. They came to perceive that "spread" is not simply measured by the end values; it involves density as well. They had to evolve their original model to handle the new context of contrasting cases. As the process of contrasting cases plus invention continued, students noticed additional features and developed models that were robust to those features. For example, we presented the contrast: {1 3 5} versus {1 1 3 3 5 5}. This helped the students notice that distributions also have different sample sizes and that their formulas needed to accommodate this possibility.

Students rarely invented the conventional solution agreed upon by experts; namely, the variance formula. However, as we demonstrate below, inventing models over contrasting cases prepares students to understand the statistical formulas at a deep level when they become available. For example, students appreciate that dividing by "n" elegantly solves the problem of different samples sizes. For now, we simply show that the process of co-evolving models and perceptions across contrasting cases helped students become aware of the aspects of context that their representations must handle to be useful. In turn, this awareness provided them with a better understanding of the work that a symbolic model does and the situations to which it refers. We compared the students who completed the DKE activities and never learned a formal solution with two other groups of students. One group of students learned the procedure for computing variance from a worked example and applied it to each of the contrasting data sets in turn. The other group of students had taken a semester of college statistics. A few weeks after our intervention, embedded in a regular class, the students from all three groups saw the problem shown in Table 1. In this problem, an industrialist has made the claim that blue people are smarter than green people, and therefore it is better to hire blue people. To support his claim, the industrialist offers the result of an intelligence test that shows that blue people have a higher average IQ than green people. He also points to the work of many other researchers who found the same result. The students saw an example of the distributions the industrialist had found. The students were asked to write as many arguments as they could think of to disagree with the industrialist.

[Table 1 – blue green problem]

When we examined the arguments presented by the students, the results were striking. The students from the DKE condition noticed that the averages were misleading. They saw past the symbolic measure (i.e., the average), and perceived the bi-modal distribution of the green people. For example, one student stated, "The average is wrong here. Nearly half the green people are smarter than all but the top few blue people." Over 95% of the DKE students noticed that there were green people that were smarter than the blue people. In contrast, less than half of the students in each of the other two groups noticed this. Instead, they tended to accept the interpretation of the average. These students exclusively made arguments like, "IQ tests are not fair," or "IQ tests do not mean they won't be good workers." So, even though the DKE students had worked with fairly limited and abstract sets of data, they were prepared to think deeply about the meaning and applicability of the average and to perceive the quantitative phenomenon to which the average refers. In contrast, the students had been directly taught statistics accepted the average at face value and did not consider whether it was a fair summarization of the data.

Measures of preparation for future learning

We now return to our second concern with methodological isolationism. We argued that much of cognitive psychology uses a sequestered problem-solving paradigm that measures the effects of learning with tests of unaided problem solving or memory. We argued that the subsequent ability to learn with resources could be a more important and sensitive indicator of the effectiveness of instruction. We suggested that an alternative to sequestered problem solving is a dynamic assessment that measures preparation for future learning. In this section, we demonstrate one measure of preparation for future learning. We used it to compare two methods of instruction. Students completed one of two instructional treatments. Afterwards, we measured how well they subsequently learned from a new information resource. The students who learned more from the resource told us something about which treatment better prepared students for future learning.

Our studies examined whether assessments of preparation for future learning would reveal important information missed by assessments of sequestered problem solving (Schwartz & Bransford, 1998). The studies occurred in the context of teaching college students about memory phenomena including false recognition, primacy, recency, ordered recall and so on. In one study, some of the students analyzed simplified data sets from classic memory experiments. By design, experiments generate contrasting data sets that help illuminate the consequences of different treatments. Table 2 provides a sample of the data sets the students analyzed. A careful examination of the data reveals multiple contrasts that can help students perceive what is significant in the results. The students' task was to analyze the data sets and graph the important patterns they found. We did not tell the students the purpose of the experimental designs; they had to discern and decide which patterns in the data they thought were important. The other students did not work with the contrasting cases. Instead, they read a modified book chapter that described the same studies, showed the graphed results, and explained their theoretical significance. Their task was to write a two-page summary of the important ideas in the chapter. A few days later, the students in both groups heard a common lecture that explained the experiments, the results, and the theories that were designed to accommodate the results. The question is whether both groups of students had been equally prepared to learn from the lecture. We also included a third group that did not hear the lecture. This group also completed the contrasting cases activity, but instead of hearing the lecture, they analyzed the data sets a second time looking for any patterns they may have missed. All told, there were three conditions: Contrasting Cases + Lecture, Summarize Chapter + Lecture, Double Contrasting Cases.

[Table 2 – Examples of simplified data sets]

To assess whether the students learned from the lecture, we employed two measures about a week later. One measure used a recognition test that included claims repeated in the book chapter and the lecture. For example, "When people understand something they have read, they tend to remember it verbatim. True or false?" The second measure was a prediction task that used the description of a novel experiment. The students' task was to predict as many of the outcomes from the experiment as possible. There were eight distinct predictions that could be applied from the materials they had worked with beforehand.

On the recognition test, the two conditions that heard the lecture looked about the same, but the Double Contrast condition did poorly. By this assessment, the contrasting cases activity appears useless. However, the results on the prediction task reveal a different story. Figure 4 shows that the Double Contrast students again did badly. However, the Summarize + Lecture students did equally badly. The Contrasting Cases + Lecture students did quite well, producing over twice as many correct predictions as students in the other conditions. By this result, the contrasting cases were very important for learning from the lecture. Students, who had read about the descriptions of the experiments instead of analyzing them first hand, did not learn very well from the lecture. In contrast, students who had analyzed the contrasting cases learned a great deal from the lecture. We know they learned from the lecture because the Double Contrast group that did not hear the lecture did badly.

[Figure 4 – time for telling results]

One important lesson from this study is that the activity of analyzing contrasting cases would have looked useless if we had not measured its effects on students' subsequent abilities to learn. Assessments of preparation for future learning can reveal levels of knowing that are imperceptible to sequestered forms of assessments. A second lesson is that assessments of memory, particularly recognition memory, can be misleading. On the test of recognition memory, students who summarized the chapter performed the same as students who analyzed the contrasting cases and heard the lecture. The benefits of the contrasting cases appeared when students had to transfer to evaluate a new situation and perceive its significant features.

A final lesson is that lectures can be an effective method of instruction if people are prepared to understand the significance of what the lecture describes. We have met researchers who believe that "telling" is inconsistent with theories of constructivist pedagogy. We have also seen instructors who refuse to tell students an answer for fear of violating effective principles of constructivist instruction. Constructivism, however, is a theory of knowledge growth and not a prescriptive theory of instruction. According to constructivism, all knowledge is constructed, whether the building blocks of knowledge come from first- or second-hand experience. Given appropriate experiences, people can be very effective at constructing knowledge (or as we prefer to say it, "effective at evolving knowledge"), even if they are sitting quietly listening to a lecture. The question is what activities prepare students to continue to evolve their knowledge.

INTEGRATING METHODS AND MEASURES TO TEACH DESCRIPTIVE STATISTICS

In the preceding sections, we offered two alternatives to methodological isolationism. One alternative was the development of studies that examine the integration of different forms of knowing rather than their isolation, and we outlined Designs for Knowledge Evolution as a promising instance. Our second alternative was to measure students' preparedness for future learning rather than relying solely on measures of isolated performance. We provided an instance of a dynamic assessment in the context of teaching theories of memory that uncovered important indicators of learning that were obscured by measures of unaided performance. We now bring these two proposals together in a single study that involved 9^{th}-grade students whom we taught descriptive statistics.

Statistics is a notorious instance where people have trouble learning the formulas and the phenomena to which they refer. A large body of cognitive research on statistical understanding has documented misconceptions about probability and statistical inference. Tversky and Kahneman (1973), in particular, have shown that people borrow non-probabilisitic reasoning methods to solve probability problems, and this leads to faulty inferences and misconceptions. The research on "judgment under uncertainty" has powerful implications for policy and practice, but it has not provided much insight into instruction (but see Nisbett, Krantz, Jepson, & Kunda, 1983). This may be because it is very difficult to remediate people's faulty heuristics, or it may be because the design of the research was not learning focused. In either case, George Cobb (1997) has proposed that it is better to avoid statistical inference in the early phases of instruction and focus on descriptive statistics. We have adopted Cobb's wisdom.

The co-evolution activities

To examine the value of DKE, especially as a preparation for future learning, we conducted a two-week study involving eight classrooms of 160 public school students (Schwartz & Martin, 2004). The instruction consisted of two three-day cycles. In cycle one, students learned about graphing and central tendency. In cycle two, they learned about formulas and variance. The study also included a third abbreviated cycle that implemented an experimental manipulation to evaluate whether DKE prepared students to learn. We provide examples of each cycle, which all students completed, and then we develop a description of the experimental treatment in cycle three.

In the first cycle, students worked with contrasting data sets that highlighted how the value of a central tendency measure depends on the context of application. For example, students decided which of two climbing ropes should get the higher rating. They received data from "load tests" that indicated the weights at which BlueStar and RedGrip ropes over multiple trials. Students typically graphed the mean load of each rope, but eventually they began to realize that for a climbing rope, the minimum load at which it will break is a safer measure than the mean. After the rope activity students further evolved their knowledge by working with contrasts where the spread of grades differed in two chemistry classes. To provide a concrete instance of the materials, Table 3 shows the data and assignment for this activity. Finally, for the third problem set, students had to decide if a drug was more effective than a placebo. In this case, the drug led to bi-modal effects such that a simple comparison of the means would be misleading.

[Table 3 – class grades task]

Students worked on each activity in small groups for about 30 minutes. Their task was to invent a graphical representation that would help justify their decisions. They were told that their graphs had to be obvious enough that another student would be able to understand what it represented and what decision the group had made. Students drew their finished graphs on the blackboard. Other students were chosen at random to come to the board and explain a graph and its implied conclusion, as if they had been part of the group. The need to make graphs that could "stand independently" of the person who made the graph, encouraged the students to develop more precise and complete second-hand description, and alerted them to the importance of communicable knowledge. In addition, just as the contrasting data sets helped students perceive important properties of distributions, the contrasting graphs that filled the board helped students notice important aspects of second-hand representations.

The cycle on graphing and central tendency introduced the students to the idea of creating their own procedures and representations. This was important because these students had typically received procedures rather than evolved them. Throughout, the instructor merely facilitated and clarified student work and presentation. It was only after completing the three activities that the instructor gave a brief lecture on conventional graphing solutions (e.g., a histogram and box plot), and students practiced for about 15 minutes. When presented with conventional graphical representations, the representations were offered as solutions that experts had invented over the years to capture important aspects of distributions. So, rather than creating a rhetorical "guided discovery" task, the students were in the position of evaluating how successful they thought the expert's solutions were. (Many preferred their own solutions.)

The second cycle, which targeted variance, began with the problem of inventing a "reliability index" for baseball pitching machines. The students worked with the grids in Figure 5. Each grid represents a different pitching machine. The X in the center is the target. Each of the black dots represents the location of one pitch from the respective machine. The grids hold many contrasts that helped students notice the distribution characteristics that their formula "index" would need to capture. Students presented their solutions on the board as before. Many students drew a box around the dots and either found the area or perimeter of the box. Other students used the Pythagorean theorem to measure the distance from a randomly chosen dot to all the other dots. Of all the groups across all the classes, only one group came up with a general and consistent solution; they measured the distance from each dot to every other dot and summed the distances. Even though no students generated a general solution, this activity plus another using more focused contrasting cases, prepared the students to learn from the subsequent lecture. The instructor presented the mean deviation formula in a five-minute lecture. Students practiced the formula with a new set of data for about 15 minutes.

[Figure 5 – baseball grids]

The final day of instruction implemented the experimental design. All the classes received a scenario in which they had to compare high scores from two distributions, for example, comparing test scores from two different tests. They had to decide which score was higher, even though the two test distributions had different means and variances. The appropriate solution to this type of problem is to use standardized scores, which measure how many deviations a score is from the mean. In lay terms, they had to "grade on a curve" and compare where each high score appeared on its respective curve. There were two conditions. In the Invention condition, the students had to invent their own solution to the problem. In the Tell-and-Practice condition, the students were told how to solve the problem graphically by marking deviation regions on a histogram and comparing the "deviations" of the high scores (i.e., the normalized score). Students worked for about 25 minutes. There were no presentations, and the instructor never presented the conventional solution for computing and comparing standardized scores. The question was whether the Invention or Tell-and-Practice students would be more prepared to learn the conventional solution when embedded as a learning resource during the post-test. Our prediction was that the Invention students would be better prepared to learn, because they had grappled first-hand with the contrasting data sets and they had invented possible models.

Assessments of performance and readiness to learn

The students took a 40-minute test on a broad array of measures before and after the instructional intervention. The test included several different types of measures to capture the benefits of the first two instructional cycles on graphing and measures of variability. For example, there were computation items, graphing problems, word problems, symbolic insight problems (e.g., "Why does the variance formula divide by "n"?). Students performed quite well at posttest, and when measured a year later, they still outperformed college students who had taken a semester of statistics (for more details, see Schwartz & Martin, 2004). Evidently, the DKE curriculum had prepared the students to learn from the brief lectures. For example, the majority of students knew that "dividing by 'n'" solves the problem of comparing different sample sizes; it finds the average of how far each data point deviates from the mean. This information was only presented once as one of many points in the brief lecture. We also included an "adaptation" problem. Students had to find a way to estimate variability in bi-variate data shown in a scatter plot, though they had only learned about univariate data. This is a difficult adaptation of the mean deviation, and the frequency of students who gave adequate responses only improved from 10% on the pretest to 34% on the posttest. On the other hand, college students who had a full semester of college statistics were only able to solve the problem 12% of the time. Compared to the college students, the DKE students had developed an adaptive base of knowledge.

For the experimental comparison of the Invention versus Tell-and-Practice conditions (which only occurred on the last day of instruction), the test evaluated how well students had been prepared to learn. Of particular focus was whether the students could learn how to compute and compare standardized scores, given summary descriptive measures. (Recall that in the class work, the students had worked with raw data not descriptive statistics). To examine whether students were prepared to learn from resources, our test included a dynamic assessment composed of two items. The "resource item" built into the test came in the form of a worked example. It provided step-by-step instructions for computing standardized scores to compare individual data from different distributions (e.g., is Betty better at assists or steals). As part of the test, students had to follow the worked example to complete a second example at the bottom of the same page. The “target transfer problem” appeared a few pages later in the test. The problem included averages, deviations, and scores of an individual in each of two groups (e.g., two biology classes). The average score and variance differed between the groups and the students had no raw data to work with. The question was whether the students would understand the implications of the procedure that appeared in the worked-example resource problem and transfer it to solve the target problem.

To ensure that students were actually learning from the worked example and using it to solve the transfer problem, the study used a 2 x 2 between-subjects experimental design. One factor was whether students were in the Invention or Tell-and-Practice condition. The second crossed-factor was whether the resource worked example appeared in the posttest. Half of the students from each condition received the step-by-step resource item for how to compute and compare standardized scores, and half did not. If students were prepared to learn from the embedded resource, then students who received the resource should do better on the target transfer problem than students who did not receive the embedded resource.

[Figure 6 – standardized score results]

Figure 6 shows the students’ performance on the target transfer item at posttest broken out by condition and whether the worked example resource appeared in the test. Students received credit if they gave either a quantitatively correct or qualitatively correct answer, including graphs. Students in the Invention condition who received the embedded resource for how to compute standardized scores doubled the performance of the other three groups. The Invention students must have learned from the embedded resource example, because Invention students who did not receive the resource performed poorly. In contrast, the Tell-and-Practice students did not show any benefits from having the resource item. Their lesson, though well-designed and highly visual, had not prepared them to learn the significance of the resource item, even though 100% of the students correctly followed the worked example on the resource item.

As fits our overall story, activities that help students perceive important properties of data coupled with activities that encourage the development of models can prepare students for future learning. Had we only measured their ability to perform without an assessment of their abilities to learn, we would have seen little benefit of DKE over the other method of instruction.

SUMMARY AND PROSPECTS FOR COMPUTER TECHNOLOGIES

Many cognitive theories are descriptive. They describe how people think, and in many cases, they provide explanations for the mechanisms of those thoughts. Other theories are normative; they describe the behaviors or knowledge of experts, and they often illuminate the limitations of novices. Fewer cognitive theories are prescriptive. They do not convert their descriptive findings into prescriptions for how to achieve particular normative outcomes. We identified two related methodological obstacles to developing prescriptive theories. One obstacle is that cognitive research tends to isolate mechanisms and disregard the multiple levels of knowing and learning relevant to education. As a broad instance, we described a tendency of the field to separate into first-hand theorists who emphasize direct experience with phenomenon and second-hand theorists who emphasize the acquisition of descriptions of experience. We proposed that these types of theorizing do not resolve the problem of how people integrate different ways of knowing. A second obstacle is that cognitive research typically uses assessments that aim to prove psychological theories rather than improve educational goals. We pointed to tests that isolate subjects from opportunities to learn during the test. We argued that these types of sequestered assessments of unaided performance can miss an important goal of most education, which is to prepare students to learn once they leave the classroom. We also demonstrated that assessments of sequestered problem solving can overlook important levels of understanding that prepare students to learn.

As an example of research designed to generate prescriptive learning theories, we described our efforts to develop and assess Designs for Knowledge Evolution. We borrowed some concepts from evolution to generate activities that help students grow intellectually. Unlike other high level theories such as constructivism, evolutionary theory suggests specific mechanisms of change that can be used prescriptively. In DKE, we encourage students to co-evolve first- and second-hand knowledge jointly across a series of different problem contexts. Students generate and revise models to differentiate contrasting cases. Over time, students develop a first-hand understanding of the important features of the domain, and they learn the work that second-hand representations need to accomplish. This prepares students to learn subsequently, for example, when they hear a formal solution in a lecture and when they follow worked examples embedded in a test. By co-evolving and adapting their knowledge to different contexts, the students exhibited excellent gains on assessments of preparation for future learning.

The current work can translate into instructional technologies. Our goal is a prescriptive learning theory that is easily adapted to current classrooms. For example, we are building technologies that permit students to complete DKE activities as homework before coming to class (instead of the more traditional model where homework follows a lesson). In this way, they will be more prepared to understand the teacher's lessons. One very simple technological implementation involves putting the problem sets on the web, along with appropriate guidance and access to other people's inventions.

We are also examining solutions that integrate additional cognitive mechanisms into knowledge evolution. One example involves Teachable Agents (Biswas et al., 2001). Teachable agents capitalize on the common wisdom that people "really" learn when they teach. So, rather than have the computer teach the students, we have the students teach the computer. Afterwards, students get to see how their agent performs, and they try to learn from its mistakes and improve its performance.

[Figure 7 – Orbo screen shots]

We have built a statistics teachable agent named Orbo[1]. Students "teach by showing." They try to show Orbo how to compute values to differentiate two contrasting distributions. Orbo tries to induce what the students are trying to show and solve subsequent problems generated by other students or teachers who want to test Orbo (e.g., using the web-based problem sets mentioned above). At select points, the computer system can introduce new problems that illuminate weaknesses in what Orbo has been taught. This leads the student to evolve new knowledge and re-teach Orbo. Finally, Orbo can exhibit misconceptions. Like a good tutor, students need to figure out what Orbo is thinking, and this helps the student's clarify their own understanding. Figure 7 describes an example where a student needs to infer what caused Orbo to misunderstand. Additional information about Orbo and another statistics agent, Milo, may be found at <aaalab.stanford.edu>.

Orbo reflects our general methodology of integrating cognitive mechanisms. For Orbo, we are integrating the two previous mechanisms of perceptual learning and modeling with the mechanism of trying to understand another person's thoughts, or in this case, another agent's thoughts. Students need to figure out what Orbo understands so they can teach him most successfully. Trying to infer the intentions behind another person's (or agent's) behavior is surely one of the most fundamental and spontaneous of human capacities, and it should be possible to leverage it for educational purposes.

In our view, there is much promise in our approach of explicitly looking for ways to evolve integrated knowledge both for understanding the nature of learning and for developing instructional theories that will be useful to classroom teachers and to students once they leave the classroom. Our hope is that the example of DKE provides one concrete instance that can help the field move in the direction of increasingly integrative models that are directly relevant to building prescriptive cognitive theories.

REFERENCES

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Cambridge University Press.

Anderson, J. R., Reder, L. M., Simon, H. (1996). Situated learning and education. Educational Researcher, 25(4), 5-11.

Barsalou, L. W. (1999). Perceptual Symbol Systems. Behaviour and Brain Sciences, 22, 577-609.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. Journal of Experimental Psychology: Learning, Memory, & Cognition,13(4), 640-645.

Biswas, G., Schwartz, D. L., Bransford, J. D., & The Teachable Agents Group at Vanderbilt. (2001). Technology support for complex problem solving: From SAD Environments to AI. In K. Forbus & P. Feltovich (Eds.), Smart machines in education (pp. 71-98). Menlo Park, CA: AAAI/MIT Press.

Bransford, J. D., Franks, J. J., Vye, N. J. & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 470-497). NY: Cambridge University Press.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), Review of Research in Education , 24, 61-101. Washington DC: American Educational Research Association.

Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. Monographs of the Society for Research in Child Development, 61(1-2), 1-265.

Chomsky, N. (1966). Cartesian linguistics: A chapter in the history of rationalist thought, New York, Harper & Row.

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics. Journal of Research in Mathematics Education, 23, 2-33.

Cobb, G. W., & Moore, D. S. (November, 1997). Mathematics, statistics and teaching. (pp. 801-823).

Feuerstein, R. (1979). The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques. Baltimore: University Park Press.

Gardner, H. (1982). Art, mind, and brain: A cognitive approach to creativity. New York: Basic Books.

Gibson, E. J. (1969). Principles of perceptual learning and development. New York: Meredith.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment. Psychological Review, 62, 32-51.

Gibson, J.J. (1986). The ecological approach to visual perception. Hillsdale, NJ: Erlbaum.

Glaser, R. (1992). Learning, cognition, and education: Then and now. In H. L. Pick, Jr., P. Van Den Broek, & D. C. Knill (Eds.), Cognition: Conceptual and methodological issues (pp. 239-266). Washington, DC: American Psychological Association.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. Psychonomic Bulletin & Review, 9, 558-565.

Hatano, G. & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, J. Azuma, & K. Hakuta (Eds.), Child Development and Education in Japan (pp. 262-272). New York, NY: W. H. Freeman & Co.

Kaput, J. J. (1995). Creating cybernetic and psychological ramps from the concrete to the abstract: Examples from multiplicative structures. In D. N. Perkins & J. L. Schwartz (Eds.), Software goes to school: Teaching for understanding with new technologies (130-154). London : Oxford University Press.

Kosslyn, S. M. (1980). Image and mind. Cambridge, MA: Harvard University Press.

Lakoff, G., & Núñez, R. E. (2000). Where mathematics comes from: How the embodied mind brings mathematics into being. New York, NY: Basic Books.

Lave, J. (1988). Cognition in practice: Mind, mathematics and culture in everyday life. New York, NY: Cambridge University Press.

Lave, J., & Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge: Cambridge University Press.

Levine, M. (1975). A cognitive theory of learning: Research on hypothesis testing. Oxford, UK: Lawrence Erlbaum.

Martin, T., & Schwartz, D. L. (2004). Physically distributed learning: Adapting and reinterpreting physical environments in the development of the fraction concept. Manuscript submitted for publication.

Mannes, S. M., & Kintsch, W. (1987). Knowledge organization and text organization. Cognition & Instruction, 4(2), 91-115.

Mayer, R. (1996). Learners as information processors: Legacies and limitations of educational psychology's second metaphor. Educational Psychologist, 31(3-4), 151-161.

Moore, J. L., & Schwartz, D. L. (1998). On learning the relationship between quantitative properties and symbolic representations. In A. Bruckman, M. Guzdial, J. Kolodner, & A. Ram (Eds), Proceedings of the International Conference of the Learning Sciences (pp. 209-214). Charolttesville, VA: Association for the Advancement of Computing in Education.

Nasir, N. (2002). Identity, Goals, and Mathematical Understandings: Learning in Basketball and Dominoes. Mathematical Thinking and Learning, v 4 (2 & 3), pp 211-245

Nasir, N. (2000). All Money Ain't Good Money: Strategy development in the game of dominoes. Unpublished Doctoral Dissertation. University of California, Los Angeles.

Nemirovsky, R., Tierney, C., & Wright, T.(1998). Body motion and graphing. Cognition & Instruction, 16(2), 119-172.

Nisbett, R. E., Krantz, D., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 90, 339-363.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension monitoring activities. Cognition and Instruction, 1, 117-175.

Saxe, G.B. (1991). Culture & Cognitive Development: Studies in Mathematical Understanding. New Jersey: Lawrence Erlbaum.

Saxe, G.B. (1994). Studying cognitive development in sociocultural context: The development of a practice-based approach. Mind, Culture, and Activity, 1(3) 135-157.

Saxe, G.B. (1999). Cognition, development, and cultural practices. In E. Turiel (Ed.). Culture and Development: New Directions in Child Psychology. SF: Jossey-Bass.

Schwartz, D. L. (1993). The construction and analogical transfer of symbolic visualizations. Journal of Research in Science Teaching, 30, 1309-1325.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. Journal of the Learning Sciences, 4, 321-354.

Schwartz, D. L. (1999). The productive agency that drives collaborative learning. In P. Dillenbourg (Ed.), Collaborative learning: Cognitive and computational approaches (pp. 197-218). New York: Elsevier Science.

Schwartz, D. L. & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. Cognitive Science, 20, 457-497.

Schwartz, D. L. & Bransford, J. D. (1998). A time for telling. Cognition & Instruction, 16, 475-522.

Schwartz, D. L, Bransford, J. D., & Sears, D. (in press). Efficiency and innovation in transfer. To appear in J. Mestre (Ed.), Perspectives on transfer. Information Age Publishing.

Schwartz, D. L., & Black, T. (1999). Inferences through imagined actions: knowing by simulated doing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 116-136.

Schwartz, D. L., & Martin, T. (2004). Inventing to Prepare for Learning: The hidden efficiency of original student production in statistics instruction. Cognition & Instruction, xx, xxx-xxx.

Schwartz, D. L., Martin, T., & Pfaffman, J. (in press). How mathematics propels the development of physical knowledge. Journal of Cognition and Development, xx, xxx-xxx.

Schwartz, D. L., & Moore, J. (1998). On the role of mathematics in explaining the material world. Cognitive Science, xx, xxx-xxx.

Shepard, R. N. & Cooper, L. A. (Eds.) (1986). Mental images and their transformations. Cambridge, MA: The MIT Press.

Smith, L. B. & Sera, M. D. (1992). A developmental analysis of the polar structure of dimensions. Cognitive Psychology, 24, 99-142.

Suchman, L. (1987). Plans and situated actions: The problem of human machine communication. NY: Cambridge University Press.

Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

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

Vygotsky, L. S. (1978). Mind in Society (Cole, M., John-Steiner, V., Scribner, S., & Souberman, E., Eds.). Cambridge: Harvard University Press.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. BCS-0214549, BCS-0214548, REC-0196238, and ROLE-0231946. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

FOOTNOTES

1. We are deeply indebted to George Chang for developing (and naming) Orbo.

TABLE 1. A test item used to measure students' abilities to perceive and evaluate measures of central tendency (Schwartz & Martin, 2004).

A wealthy industrialist wrote a book describing how to make a business work. He said the single most important task was to hire the smartest people possible. In particular, he suggested hiring BLUE people. To back up his suggestion, he reported the results of a study in which he compared the intelligence of BLUE and GREEN people. In the study, he randomly selected 40 BLUE people and 40 GREEN people. He gave each individual in each group an IQ test. Here are the individual scores and the group averages:

GREEN People Scores
82, 83, 84, 86, 87, 88, 88, 88, 89, 89, 89, 89, 89, 90, 90, 90, 90, 91, 91, 92, 95, 95, 97, 101, 106, 108, 108, 109, 109, 109, 110, 110, 110, 110, 111, 111, 111, 112, 113, 115
GREEN average IQ = 98

BLUE People Scores
85, 93, 96, 97, 97, 98, 98, 99, 99, 99, 99, 100, 100, 100, 100, 100, 100, 101, 101, 101, 101, 101, 102, 102, 102, 102, 102, 102, 103, 103, 103, 103, 104, 104, 104, 105, 106, 106, 107, 111
BLUE average IQ = 101

Based on this data, the industrialist claimed that BLUE people are smarter than GREEN people. One hundred activists across the country were outraged and claimed that the industrialist's results were a fluke. They each conducted their own studies by giving IQ tests to BLUE and GREEN people. To their surprise, the activists came up with results that were nearly identical to the industrialist's -- the industrialist's results were reliable. The industrialist published an article in the *New York Times* reporting the results. He repeated his suggestion, "If you want the smartest people to work for you, hire BLUE people."

How would you argue that the industrialist's conclusions are wrong?
Write as many arguments as you can think of in the next 5 minutes.

TABLE 2. Examples of the simplified and contrasting data sets that students analyzed to prepare them to learn from a lecture (Adapted from Schwartz & Bransford 1998).

Study 1

In this study, psychological researchers brought together five subjects. The researchers read the subjects the following list of 20 words at a rate of one word per three seconds. Here are the words the researchers read to the subjects in the order in which they read them:

car, sky, apple, book, cup, lock, coat, light, bush, iron,
water, house, tape, file, glass, dog, cloud, hand, chair, bag

After the researchers read these words they said: "Recall." When the researchers said "Recall" the subjects wrote down as many of the 20 words as they could remember.

Here are the words the five subjects in the study recalled and the order in which they recalled them:

Sbj #1: bag, hand, chair, dog, car, sky, apple, book, tape, file, house, list, bush
Sbj #2: bag, chair, hand, cloud, sky, car, book, apple, cup, lock, iron, glass
Sbj #3: bag, hand, chair, cloud, sky, car, apple, book, file, bush, coat, iron, tape
Sbj #4: bag, hand, chair, dog, car, sky, apple, water, cup, glass, house, bush, dog, book
Sbj #5: bag, chair, hand, cloud, sky, car, book, coat, water, light, lock, house

Study 2

This study is the same as Study 1, except that the researchers did not tell the subjects to recall the words immediately after reading the list of words. Instead, the researchers asked the subjects to do another task first (i.e., a division problem). This task took 30 seconds. Immediately after this task they were told to recall as many of the words as they could, again in any order they liked.

Here are the words the five new recalled and the order in which they recalled them:

Sbj #6: car, sky, book, apple, bush, house, glass, chair
Sbj #7: car, sky, lock, iron, water, cloud, bag
Sbj #8: car, apple, coat, bag, hand, file
Sbj #9: car, sky, light, cup, tape, dog
Sbj #10: car, apple, cup, water, glass, house

Table 3. An example of a contrasting case the students analyzed to learn about central tendency and graphing (Schwartz & Martin, 2004).

MAKING THE GRADE

Imagine your friend Julie is very worried about getting a good grade in Chemistry. She can take the class from Mrs. Oxygen or from Mr. Carbon. Here are the grades each teacher gave out last year.

Mrs. Oxygen: D+, D+, C-, C-, C-, C, C, B-, B-, B-, B-, B, B, B, B, B+
Mr. Carbon: D+, C-, C-, C-, C, C, C+, C+, C+ A-, A+, A+

Which teacher would you suggest? Create a visual representation of the data to support your position. If your visualization from before does not work, try something new.

FIGURE CAPTIONS

Figure 1. Will the glasses pour at the same or different angles?

Figure 2. The average angles of tilt for subjects who turned the glasses with their eyes closed and imagined they had water (Schwartz & Black, 1999).

Figure 3. Examples of explicit models students invented to solve problems about chains of causality (The construction and analogical transfer of symbolic visualizations, Schwartz, D. L., Journal of Research in Science Teaching, 30. Copyright © 1993 Wiley Periodicals, Inc.).

Figure 4. Students who studied contrasting data sets were more prepared to learn from a lecture than students who read and summarized a book chapter. (Adapted from Schwartz & Bransford, 1998).

Figure 5. An example of an activity where students co-evolved their perception and symbolic characterization of variance (Schwartz & Martin, 2004). Each grid represents a pitching machine. The 'X' is the target, and the black dots indicate where each ball landed. Students had to invent a "reliability index" to characterize each pitching machine.

Figure 6. Students who had evolved their knowledge were more prepared to learn from a resource embedded in the posttest. (Adapted from Schwartz & Martin, 2004)

Figure 7. Orbo: A Teachable Agent for Statistics. In the top panel a student has invented a way to find the spread in a set of data by subtracting adjacent values and shows the agent, Orbo, how to do it with a given data set. In the bottom panel, Orbo applies the method to a new data set. Orbo does not do what the student intended. This leads the student to think what Orbo must have had in mind. The answer is that in the teaching

example, there were only four numbers, so Orbo induced that the procedure only uses four numbers. This alerts students to the importance of a solution that can generalize across sample sizes.

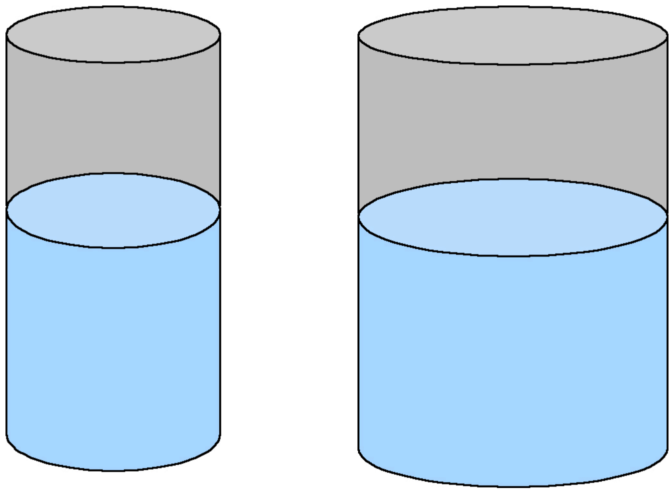

Figure 1.

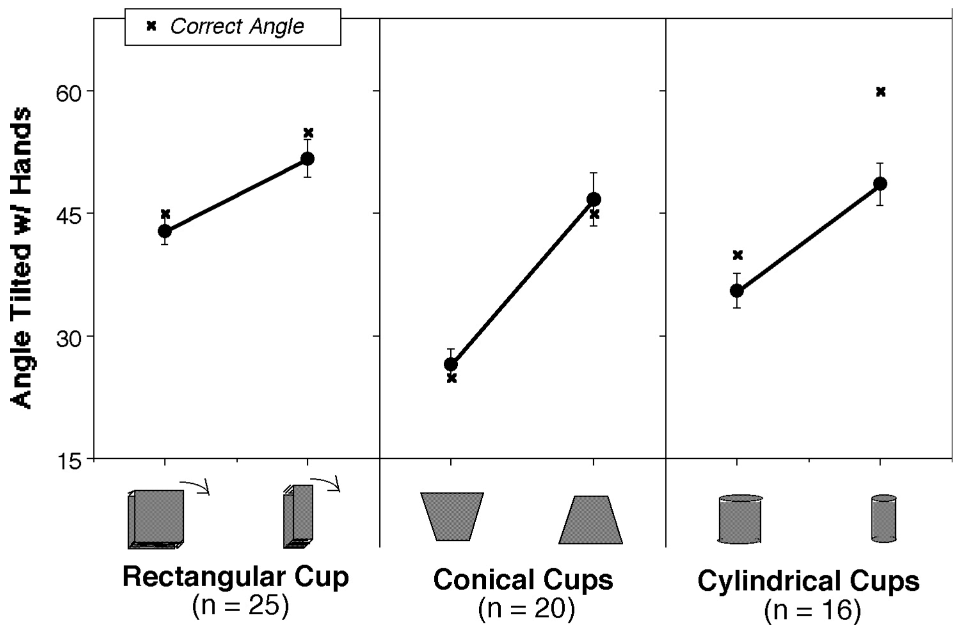
Correct Angle
Angle Tilted w/ Hands
60
45
30
15
Rectangular Cup
(n = 25)
Conical Cups
(n = 20)
Cylindrical Cups
(n = 16)

Figure 2.

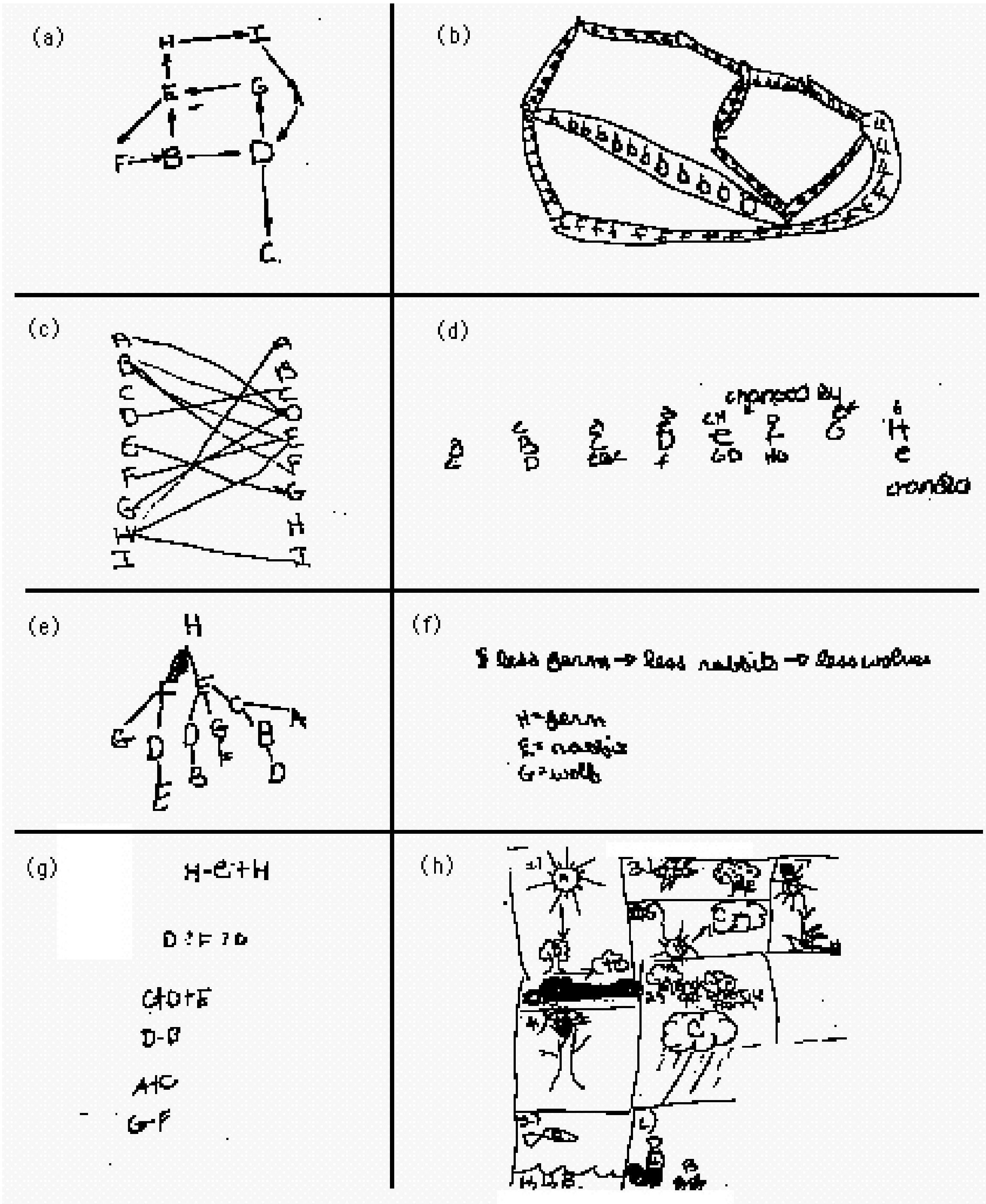

Figure 3.

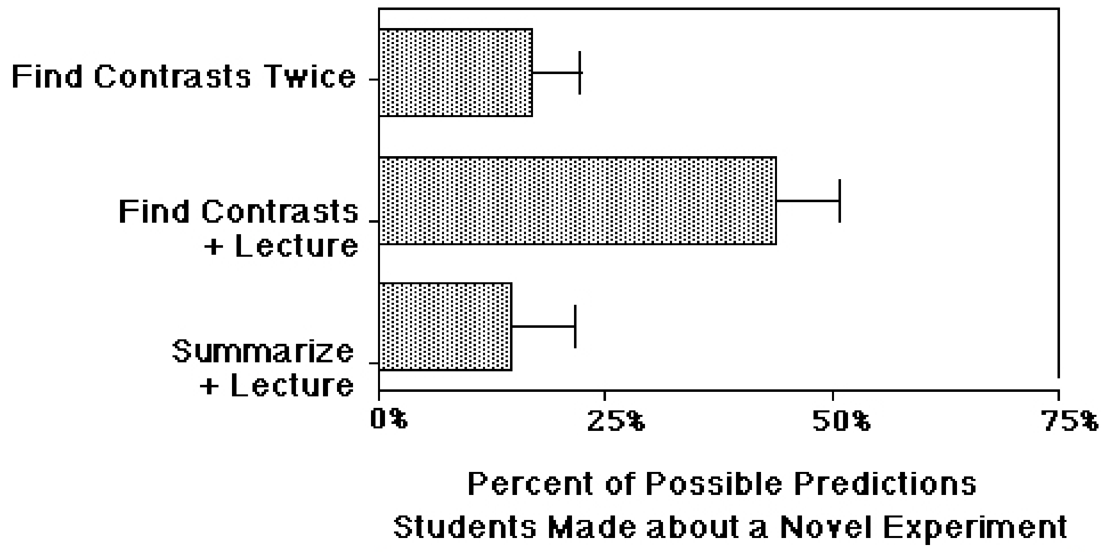
Find Contrasts Twice
Find Contrasts
+ Lecture
Summarize
+ Lecture
0%
25%
50%
75%
Percent of Possible Predictions
Students Made about a Novel Experiment

Figure 4.

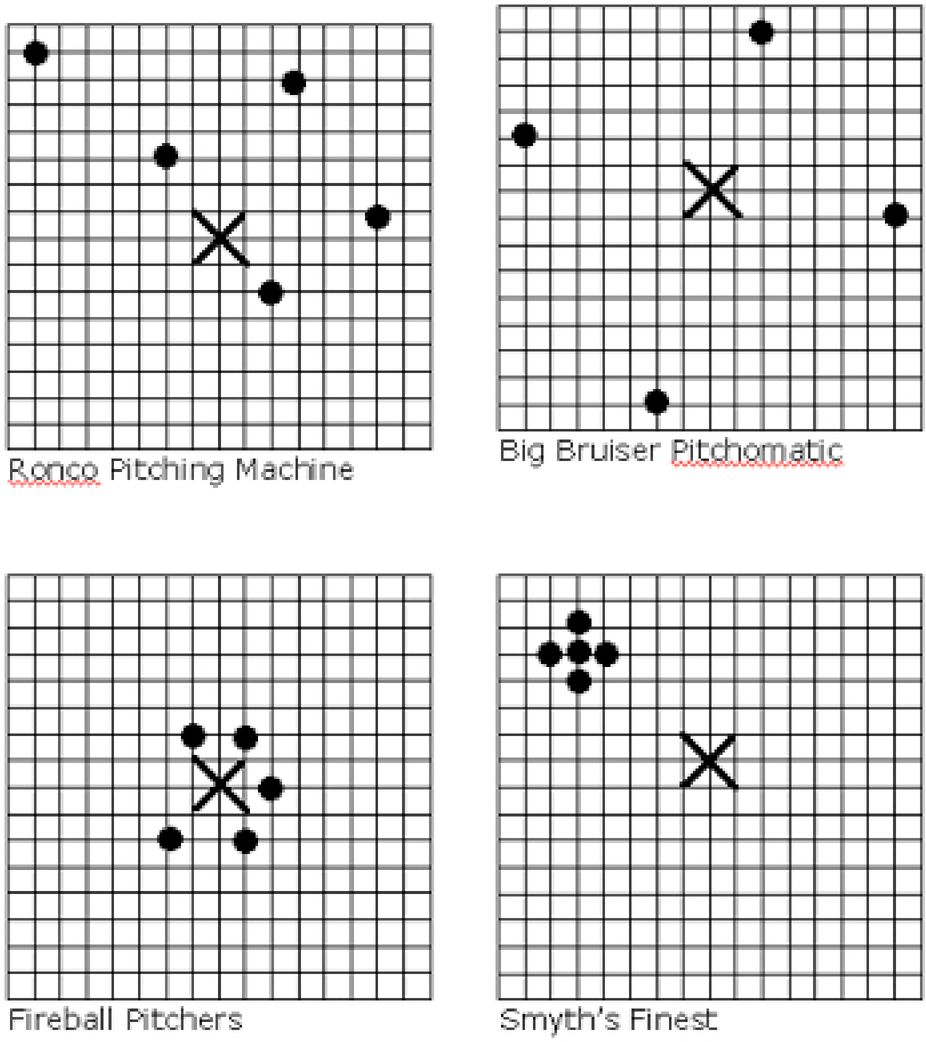
Ronco Pitching Machine
Big Bruiser Pitchomatic
Fireball Pitchers
Smyth's Finest

Figure 5.

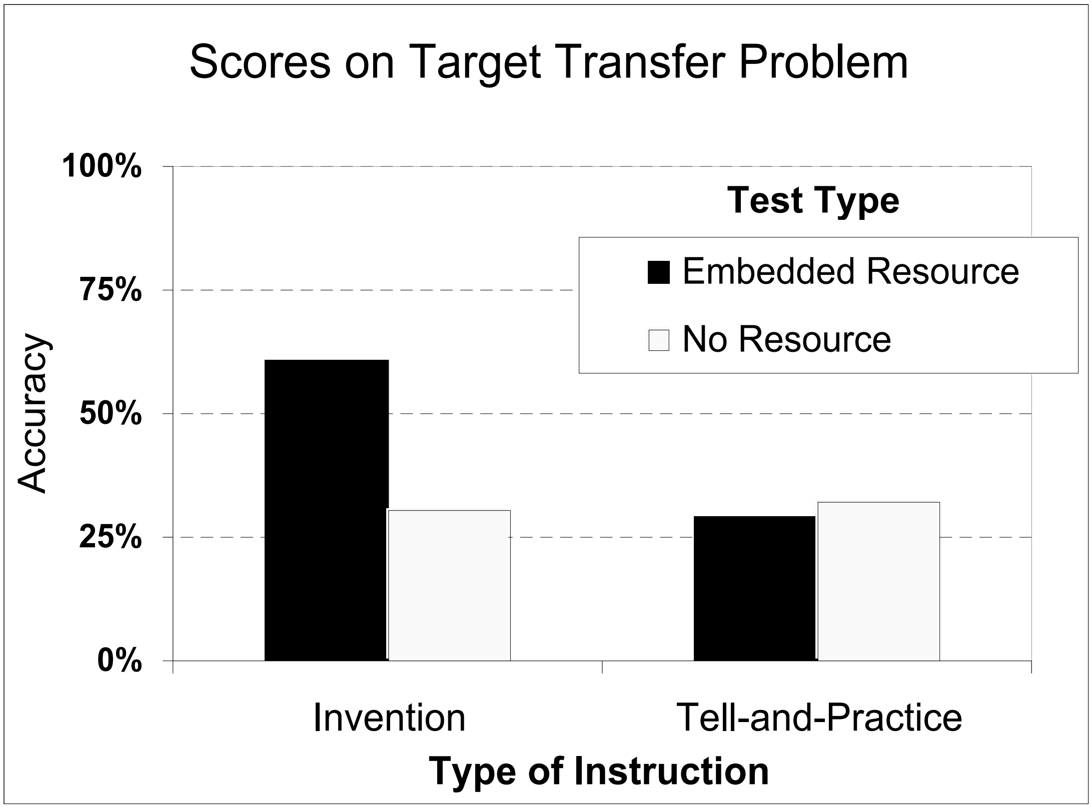
Scores on Target Transfer Problem
100%
75%
50%
25%
0%
Accuracy
Test Type
Embedded Resource
No Resource
Invention
Tell-and-Practice
Type of Instruction

Figure 6.

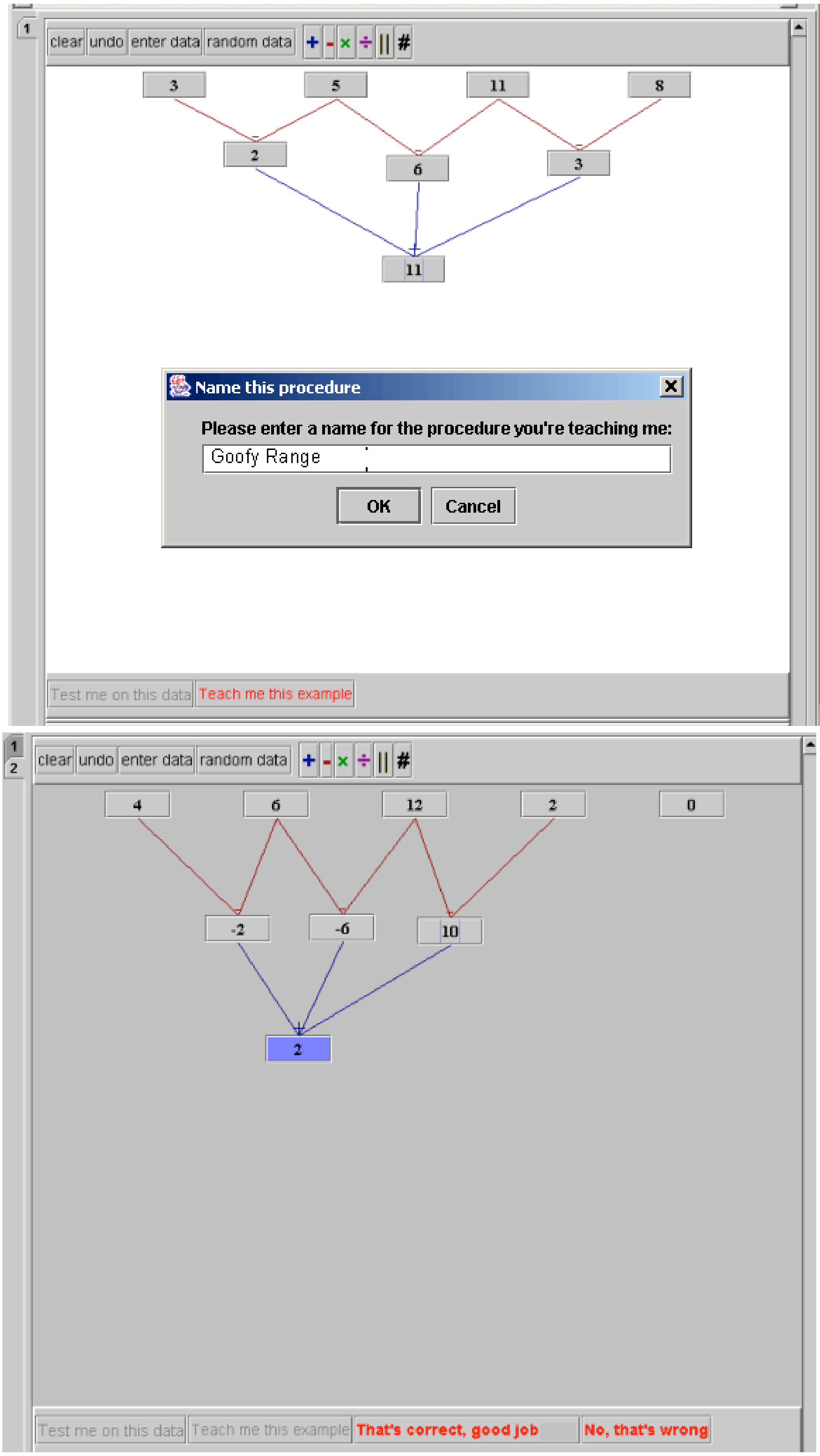
clear
undo
enter data
random data
#
3
5
11
8
2
6
3
11
Name this procedure
Please enter a name for the procedure you're teaching me:
Goofy Range
OK
Cancel
Test me on this data
Teach me this example
clear
undo
enter data
random data
#
4
6
12
2
0
-2
-6
10
2
Test me on this data
Teach me this example
That's correct, good job
No, that's wrong

Figure 7.