VanLehn, K. & van de Sande, B. (2009). Expertise in elementary physics, and how to acquire it. In K. A. Ericsson (Ed.) The Development of Professional Performance: Toward Measurement of Expert Performance and Design of Optimal Learning Environments. pp. 356-378. Cambridge, UK: Cambridge University Press.

16

Acquiring Conceptual Expertise from Modeling: The Case of Elementary Physics

Kurt VanLehn and Brett van de Sande

In many domains, the real world is modeled with systems of equations. Such a model uses variables to represent domain properties and equations to represent applications of domain principles. Given a set of true domain relationships expressed as equations, one can deduce new equations from them using only the rules of mathematics, and the new equations will also be true domain relationships. The latter step, wherein mathematical implications are derived from the initial model, can often be done mechanically, for example, by mathematical symbol manipulation programs, spreadsheets, calculators, etc.

Given a real world situation that is amenable to such analysis, experts and novices understand them quite differently. Whereas novices must go through the whole modeling process by writing equations on paper and solving them, experts can generate many conclusions about the same situations without having to commit anything to paper. For the expert, many domain relationships are just "obvious" or can be easily inferred "by inspection."

There are limits to the experts' abilities. Although experts usually cannot mentally infer *quantitative* relationship, such as the exact numerical value for an energy or a velocity, they can infer *qualitative* relationships, such as whether a quantity is zero, increasing or greater than some other quantity. Thus, it is often said that expertise in such domains is characterized by a *conceptual* or *qualitative* understanding of real world situations (VanLehn, 1996). It is sometimes said that they have developed domain-specific *intuitions* (Simon & Simon, 1978). This ability of the experts is called *conceptual expertise*, and it is the focus of this chapter.

Interest in conceptual expertise has increased in recent years with the discovery that in some surprisingly simple situations, novices have intuitions that conflict with the experts' intuitions (McCloskey, Caramazza, & Green, 1980). For instance, suppose a bowling ball and a golf ball are dropped from 2 meters above the Earth's surface. An expert will know immediately that the two balls strike the earth at exactly the same time, whereas novices usually say that the balls land at slightly different times. Novice intuitions about such situations are often quite systematic, leading to the view that novices have alternative beliefs about how the world works, often called *misconception*s. In the last few decades, such misconceptions have been documented in almost every branch of science (Pfundt & Duit, 1998).

Scientists and educators have been appalled at the ubiquity of misconceptions, which appear in the responses of even the best students. That such misconceptions survive high school and university science courses has been viewed as a crisis in science education. Preventing and/or overcoming misconceptions is the explicit goal of countless educational research and development projects. One step toward achieving that goal is understanding conceptual expertise. In short, interest in understanding conceptual expertise has increased with the discovery of ubiquitous, systematic misconceptions among students who have already taken appropriate courses.

In this chapter, we analyze conceptual expertise and propose an account for its acquisition. Our claim, which is not surprising in the context of this book, is that this kind of conceptual expertise arises from extensive practice of a certain kind. Moreover, there is a natural developmental sequence that occurs with every domain principle: from a *superficial* understanding, to a *semantic* understanding and finally to a *qualitative* understanding. Novices often have too little practice of the right kind, so they never reach the semantic stage of understanding, which appears to be the key.

Although we believe this account will hold true in many domains, this chapter discusses it in the context of just one domain: elementary physics – the kind of physics taught in first-year college courses and in advanced high-school courses. Elementary physics is a small and well-defined domain compared to others in the expertise literature, which makes it tractable for laboratory study. Indeed, early in history of expertise research, elementary physics emerged as the premier domain with a number of very influential studies discovering characteristics of expertise that were later found to be general (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982; Larkin, McDermott, Simon, & Simon, 1980).

Apart from its small, well-defined knowledge base, elementary physics is a domain with excellent access to participants at many levels of expertise. In pioneering studies (e.g., Chi et al., 1981; Simon & Simon, 1978), the concepts and the performance of novices were contrasted with those of experts, where the *experts* were college instructors who had been teaching elementary physics for many years, and the *novices* were recent graduates of such physics courses. Some studies reached further back and tested students just beginning a college physics course (*pre-novices*).

Another reason to study expertise in elementary physics is that the task domain has been thoroughly analyzed in order to develop artificially intelligent physics problems solving software (Bundy, Byrd, Luger, Mellish, & Palmer, 1979; de Kleer & Brown, 1984; Forbus, 1985; Larkin, 1981; Larkin, Reif, Carbonell, & Gugliotta, 1988; McDermott & Larkin, 1978; Novak & Araya, 1980); intelligent

Problem:
A bomber is flying at 1000 km/hr at an altitude of 500 m. If the bomber releases a 1000 kg bomb, what is its impact speed? What would the impact speed of a 500 kg bomb be?

Solution:
Let m be the bomb's mass, so m = 1000 kg
Let v1 be the bomb's initial velocity, so v2 = 1000 km/hr = 278 m/s.
Let h be the height of the bomb when it is released, so h = 500 m.
Let v2 be the bomb's final velocity. We need to solve for this.
The initial total mechanical energy of the bomb is KE1 + PE1 = ½*m*v1^2 + m*g*h
The final total mechanical energy is KE2 = ½*m*v2^2,
Because we can ignore air friction, mechanical energy is conserved,
so we can equate the initial and final total mechanical energies.
Thus, ½*m*v2^2 = ½*m*v1^2 + m*g*h.
Solving, we have v2 = sqrt[v1^2 + 2*g*h] = sqrt[(278 m/s)^2 + 2*(9.8 m/s^2)*(500 m)]
So, v2 = 295 m/s.
Because m cancels, a 500 kg bomb would have the same impact speed.

FIGURE 16.1. A quantitative elementary physics problem and its solution.

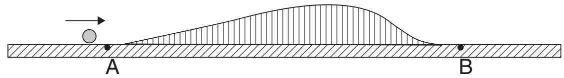
tutoring systems (Jordan, Makatchev, Pappuswamy, VanLehn, & Albacete, 2006; Loftin, Mueller, Way, & Lee, 1991; Murray, Schultz, Brown, & Clement, 1990; Reif & Scott, 1999; VanLehn et al., 2002; VanLehn et al., 2005); computational models of learning (Elio & Scharf, 1990; Jones & Fleischman, 2001; Lamberts, 1990; Reimann, Schult, & Wichmann, 1993; VanLehn & Jones, 1993b; VanLehn, Jones, & Chi, 1992); and qualitative accounts of learning (M. T. H. Chi, 2005; di Sessa, 1993; Sherin, 2001, 2006).

Still another reason for studying physics expertise is that it permits us to study two different indices of mastery of the domain. The first type is *quantitative* problem solving, which involves writing equations and solving them, usually over a period of several minutes. Figure 16.1 shows a quantitative problem and its solution. The second index of expertise is solving problems that involve little math, little writing, and brief solution times. These are called *conceptual* problems or sometimes *qualitative* problems, and they are often devised to display differences between expert and novice intuitions. Figure 16.2 shows some conceptual problems.

On the face of it, the conceptual problems should be the easier of the two because they take less time, involve no math, and so on. A consistent finding in the literature of physics expertise is, however, that even novices who approach experts in their ability to solve quantitative problems are often more like prenovices in their ability to solve conceptual problems (Hake, 1998; Hestenes, Wells, & Swackhamer, 1992).

Most physics education researchers assume that current physics instruction is at fault, and they have invented and sometimes evaluated many instructional innovations. Although there have been improvements, Hake (1998) and others conclude that after decades of instructional innovation, the problem is still not solved.

1. A steel ball rolls along a smooth, hard, level surface with a certain speed. It then smoothly rolls up and over the hill shown below. How does its speed at point B after it rolls over the hill compare to its speed at point A before it rolls over the hill?



A. Its speed is significantly less at point B than at point A.
B. Its speed is very nearly the same at point B as at point A.
C. Its speed is slightly greater at point B than at point A.
D. Its speed is much greater at point B than at point A.
E. The information is insufficient to answer the question.

2. Two steel balls, one of which weighs twice as much as the other, roll off a horizontal table with the same speeds. In this situation:
A. Both balls impact the floor at approximately the same horizontal distance from the base of the table.
B. The heavier ball impacts the floor at about half the horizontal distance from the base of the table than does the lighter.
C. The lighter ball impacts the floor at about half the horizontal distance from the base of the table than does the heavier.
D. The heavier ball impacts the floor considerably closer to base of the table than the lighter, but not necessarily half the horizontal distance.
E. The lighter ball impacts the floor considerably closer to base of the table than the heavier, but not necessarily half the horizontal distance.

FIGURE 16.2. Two qualitative physics problems.
(*Source:* Problem 1 courtesy of Prof. David Hestenes, Arizona State University.)

Our hypothesis is that conceptual expertise is comprised of a qualitative understanding of domain principles, and that extensive practice is needed for a learner to develop through the stages of superficial understanding and semantic understanding before finally arriving at a qualitative understanding. Instructional innovations that fail to provide sufficient practice of the right kinds take students only part of the way toward conceptual expertise.

This chapter will expand upon this hypothesis by presenting a cognitive task analysis and learning mechanisms that are consistent with findings from the elementary physics literature. Because the chapter presents no new findings, it is purely theoretical.

The chapter has several sections, one for each of these questions:

1. What knowledge comprises conceptual expertise in physics?
2. What happened to novices' misconceptions?
3. How does expertise affect quantitative problem solving?
4. How can conceptual expertise be learned?

The first three sections develop an account for the existing body of expert–novice findings. The key idea is that experts have acquired a class of knowledge, called *confluences*, that novices lack. The fourth section proposes an explanation for how conceptual expertise can be learned, based on two well-known learning mechanisms, induction from examples and EBL. The key idea is that learners go through three stages of understanding per principle: superficial, semantic, and qualitative.

This chapter does not attempt to argue systematically for the generality of its claims beyond elementary physics. However, many of the expert–novice phenomena observed in physics have also been observed in other task domains, which increases the plausibility that elementary physics will again prove prototypical of other task domains. As an example of such an expert–novice finding, consider the Chi, Feltovich, and Glaser (1981) discovery that physics novices classify problems by surface features, whereas physics experts prefer to classify problems by deep features. This expert–novice difference has been found in other task domains as well, such as chemistry (Kozma & Russell, 1997), management organizational problems (Day & Lord, 1992), genetics (Smith, 1992), programming (Weiser & Shertz, 1983), counseling (Mayfield, Kardash, & Kivlighan, 1999), and fishing (Shafto & Coley, 2003). As a second example, consider Priest and Lindsay's (1992) finding that experts could mentally generate and orally explain a plan for solving an elementary physics problem, but novices could not. This expert–novice difference has been found in other task domains as well (Ericsson, 2006; Ericsson & Lehmann, 1996; Heyworth, 1999). In short, although parts of this proposal for conceptual expertise are clearly specific to physics alone, it is likely that other parts are more general and apply to mastery of several different domains of knowledge-based expertise. Some speculations on generality are included in a final discussion at the end of the chapter.

WHAT KNOWLEDGE COMPRISES CONCEPTUAL EXPERTISE IN PHYSICS?

In order to uncover the knowledge that comprises conceptual expertise, this section works backward from the problems used to assess it. It asks: What knowledge is needed in order correctly solve these problems? This section is a summary of a computer model, Cascade, that Ploetzner and VanLehn (1997) showed could correctly solve many problems on the widely used Force Concept Inventory (Hestenes et al., 1992). However, the basic categorization of knowledge presented here appears in many other accounts of physics knowledge, including both those backed by computer modeling (Bundy et al., 1979, Elio & Schart, 1990; de Kleer & Brown, 1984; Forbus, 1985; Jones & Fleischman, 2001; Lamberts, 1990; Larkin, 1981; Larkin et al., 1988; McDermott & Larkin, 1978; Novak & Araya, 1980; Reimann et al., 1993; VanLehn & Jones, 1993b; VanLehn et al., 1992) and others (M. T. H. Chi, 2005; di Sessa, 1993; Hestenes, 1987; Sherin, 2001, 2006).

Concept inventories are usually multiple-choice tests with questions such as those in Figure 16.2 and the following one (from Prof. Robert Shelby, personal communication):

> A dive bomber can release its bomb when diving, climbing, or flying horizontally. If it is flying at the same height and speed in each case,

in which case does the bomb have the most speed when it hits the ground?

A. Diving
B. Climbing
C. Flying horizontally
D. It doesn't matter. The bomb's impact speed is the same in all three cases.
E. More information is needed in order to answer.

Most novices and pre-novices choose A as their answer. The experts prefer D, as they recognize this as an application of Conservation of Mechanical Energy. Now consider this problem (from Hestenes et al., 1992):

A book is at rest on a table top. Which of the following force(s) is(are) acting on the book?

1. A downward force due to gravity.
2. The upward force by the table.
3. A net downward force due to air pressure.
4. A net upward force due to air pressure.

Experts would answer 1, 2, and 4, whereas novices tend to answer 1 and 3. Although the dive bomber problem requires applying a principle, the book problem only requires identifying forces.

These problems illustrate two aspects of conceptual expertise. The book problem illustrates what Hestenes (1987) calls the "description phase" of modeling. The expert decides how to describe the physical world in terms of ideal objects, relationships, and quantities, such as point masses, forces, energies, accelerations, etc. Likewise, solving the dive bomber problem begins with a description phase, where the expert constructs an idealized model by deciding whether to neglect friction, ignore rotation of the Earth, treat the bomb as a point object, and so on. At the end of the description phase, the stage is set for applying principles, but no principles have been applied.

Let us coin the term "description phase knowledge" and agree that conceptual problems, such as the book problem, test whether students have such knowledge. Such knowledge appears as a distinct type in several computational models of physics problem solving (e.g., Bundy et al., 1979; Ploetzner & VanLehn, 1997; VanLehn et al., 1992), whereas in other models, it is represented in the same formalism as other knowledge. All these models interleave the application of description phase knowledge with other knowledge in order to be consistent with human data. When problems are exceptionally tricky, experts verbalize applications of description phase knowledge and the applications are intermingled with application of other knowledge rather than being done as a distinct phase (Larkin, 1983).

Although description phase knowledge alone suffices for solving some problems, it does not suffice for the dive bomber problem. For the dive bomber problem, experts must both recognize that Conservation of Mechanical Energy can apply and draw conclusions from its application. In the Hestenes (1987) terminology, these comprise the formulation and ramification stages of modeling. It is widely believed, and consistent with much work on the modeling of expertise (VanLehn, 1996), that the knowledge driving these stages is organized into principle *schemas* (Chi et al., 1981; Dufresne, Gerace, Hardiman, & Mestre, 1992; Larkin, 1983; VanLehn, 1996). Thus, we use "principle schemas" to refer to it.

Schemas have three important parts: applicability conditions, bodies, and slots (Russell & Norvig, 2003). The information for deciding whether a schema applies is called the *applicability conditions* of the schema, and the information that draws conclusions comprises the *body* of the schema. A schema also has *slots*, which are filled with objects from the problem situation and indicate how the principle is mapped onto the situation. Applicability conditions determine which objects can fill the schema's slots. For instance, in applying Conservation of Mechanical Energy to the dive bomber problem, the schema's applicability conditions decide that the principle should be applied to the bomb, not the plane; that is, the bomb should fill one of the slots of the principle's schema. Let us examine applicability conditions a bit further, then consider the bodies of schemas.

Verbal protocols taken as experts read ordinary problems suggest that they recognize the applicability of schemas rapidly, sometimes after reading just a few words of the problem (Hinsley, Hayes, & Simon, 1977). This suggests that their knowledge includes simple applicability conditions for recognizing commonly occurring special cases. For instance, one such applicability condition is:

> If there is a moving object, and we have or need its velocity at two time points, time 1 and time 2, and there are no non-conservative forces acting on the object between those two time points, then we can apply Conservation of Mechanical Energy to the object from time 1 to time 2.

This applicability condition could be used for the dive bomber problem, and expert verbal protocols suggest that it is indeed used for many Conservation of Mechanical Energy problems (Chi et al., 1981).

Applicability conditions delineate only the *possible* principle applications. This would suffice if the question merely asked, "Which principles apply when a dive bomber drops a bomb?" (c.f., Owen & Sweller, 1985; Sweller, Mawer, & Ward, 1983). However, the dive bomber question did not ask about principle applications, but about the speed of the bomb when it reached the ground. To answer such questions, the expert must use knowledge from the *body* of the principle's schema. For this example, the expert might use this knowledge component:

> If Conservation of Mechanical Energy is applied to two objects, A and B, and the mass, initial velocity, initial height, and final height have the

same values for object A as for object B, then objects A and B must have the same final velocity as well.

The second type of knowledge component is a rule inferred from the principle's equations, KE1 + PE1 = KE2 + PE2, where KE stands for kinetic energy and PE stands for potential energy, and the numbers distinguish the two time points. A rule inferred from a qualitative interpretation of an equation is called a *confluence* (de Kleer & Brown, 1984). A confluence is stated in terms of a qualitative value system, such as {positive, negative, zero, nonzero} or {increasing, decreasing, constant, non-constant}. For instance, if the algebraic form of a principle is $X = Y + Z$, then a confluence based on the value system {increase, decrease, constant} is "If X increases and Y is constant, then Z increases." Another confluence for the same equation and same value system is, "If Z decreases and Y decreases, then X decreases." There are 19 more such confluences. For any equation and any qualitative value system, there are a finite number of confluences, and it is straightforward to work them all out.

To summarize, the hypothesis is that experts solve conceptual problems by applying three types of knowledge component: (1) description phase knowledge, (2) *applicability conditions*, which have the form, "if <condition> then <principle application>," and (3) *qualitative confluences*, which have the form, "if <principle application> and <quantity has qualitative value>, <quantity has qualitative value>, … then <quantity has qualitative value>." The latter two are parts of principle schemas. Experts have a great deal of knowledge besides these three types, but for answering simple conceptual problems, they probably only need these three.

Although concept inventories are convenient and widely used to assess conceptual expertise, there are other methods as well. Thus, we need to check that the three types of knowledge mentioned above will succeed on these less-common assessments as well.

Several assessments involve showing participants some ordinary *quantitative* problems, such as the one shown in Figure 16.1, and asking them *not* to solve the problems but instead to:

1. Sort the problems into clusters of similar problems (Chi et al., 1981; Chi et al., 1982),
2. Describe their basic approach or plan for solving the problem (Chi et al., 1982; Priest & Lindsay, 1992), *or*
3. Pick which of two other problems is similar to the given problem (Hardiman, Dufresne, & Mestre, 1989).

The common finding among all these studies is that experts can mentally generate a plan for solving a problem. The plan identifies the major principle applications, which are then used for problem clustering and problem similarity judgments. When the experts explain their choices in these tasks or their basic approach to solving a problem, they nearly always mention principles.

On the other hand, novices' basic approaches seldom mention application of principles. Accord to Chi et al. (1981, p. 142):

> when asked to develop and state "a basic approach," [novices] did one of two things. They either made very global statements about how to proceed, "First, I figured out what was happening … then I, I started seeing how these different things were related to each other … I think of formulas that give their relationships and then … I keep on relating things through this chain…." or they would attempt to solve the problem, giving the detailed equation sets they would use.

This suggests that experts' schemas include *planning confluences*, where a planning confluence is a confluence (a non-numerical interpretation of an equation) that uses the value system {known, sought}. For instance, generic planning confluences for $X = Y + Z$ include "if X and Y are known, then Z is known," and "if X is sought and Y is known, then Z is sought." As another example, a planning confluence useful for Figure 16.1 is:

> If the definition of kinetic energy applies to an object at time 1, and the velocity of the object at time 1 is sought, then its kinetic energy at time 1 should be sought.

In summary, the principle schemas of experts, but not novices, have planning confluences of the form, "if <principle application> and <quantity is known/sought>, <quantity is known/sought>, … then <quantity is known/sought>."

There is evidence that these confluence are actually used by experts. Chi et al. (1981, p. 124) found that their experts took longer than novices to choose a cluster for a problem (45 seconds per problem versus 30 seconds). This is consistent with experts using the extra time to plan a solution to some problems. Moreover, Larkin (1983) found that when experts are given unfamiliar problems, their verbal protocols are peppered with statements of the form "<quantity> is known, so …" and "we need <quantity>, so …." When students' principle schemas have correct applicability conditions but lack planning confluences, then certain kinds of quantitative problems can fool them into applying principles that are irrelevant to solving the problem (M. Chi & VanLehn, 2008).

The question addressed by this section is, "what knowledge comprises conceptual expertise in elementary physics?" and the proposed answer is, "description phase knowledge, and for each principle, mastery of its applicability conditions, its qualitative confluences, and its planning confluences." This claim should be understood as an approximation. There are other knowledge components that do not fit these categories and yet they are part of conceptual expertise.

For instance, experts tend to know which quantities typically cancel out during calculations. It turns out that simple reasoning with confluences is incomplete – there are solutions to systems of confluences that cannot be found by the simple algorithm (Forbus & de Kleer, 1993). One method, called "plunking" (Forbus & de Kleer, 1993), for increasing the number of solutions found, is to assume that a quantity will cancel out and then check (somehow!) that it does. This appears to be what experts do, but the evidence is only anecdotal at this time.

Finding sound and complete algorithms that solve large systems of confluences is just one of the problems studied in the field of qualitative reasoning (QR) about physical system. The field has an extensive literature, a textbook (Forbus & de Kleer, 1993), and periodic conferences(http://www.cs.colorado.edu/~lizb/qr08.html is the web site for the 22nd conference, which occurred in 2008). For the QR community, this whole section is old news. For them, it is axiomatic that conceptual expertise includes at least qualitative principle schemas and description phase knowledge. Moreover, QR applies to a wide variety of science and engineering domains. The default assumption is that if one can model something with differential equations, then it can probably also be modeled with QR.

WHAT HAPPENED TO THE NOVICES' MISCONCEPTIONS?

On conceptual problems, many different novices give the same incorrect answer. For instance, on the dive bomber problem, most novices think that releasing the bomb when the dive bomber is diving will maximize the final velocity of the bomb. Such systematic, incorrect responses have been collected and codified as *misconceptions* (also called alternative conceptions). A common misconception is: "If an object is moving in a certain direction, there is a force acting on it in that direction." Many misconceptions have been inferred (Pfundt & Duit, 1998) and there has been much research on their epistemology and ontogeny (e.g., Chi, in press; di Sessa, 1993; Vosniadou & Brewer, 1992).

Presumably, physics experts once had misconceptions, which have disappeared as the experts achieved conceptual expertise. Because graduate student teaching assistants still hold many misconceptions (Hestenes & Wells, 1992; Hestenes et al., 1992), the disappearance of misconceptions may occur somewhat late in development. If we nonetheless assume that experts lack misconceptions, we can speculate as to the reasons for their demise.

A key observation, made by many (e.g., Hestenes et al., 1992; Ranney & Thagard, 1988) is that abstract, general misconceptions are incompatible with conceptual expertise. For instance, suppose someone who believes that motion always implies force in the direction of motion gradually acquires a qualitative confluence of Newton's Second Law, that a zero acceleration (i.e., constant velocity) implies a zero net force. These two beliefs produce contradictory predictions about many familiar situations. As budding experts begin to master such confluences, they may notice the contradictions. This probably weakens their belief in the misconceptions and narrows the situations where the misconceptions' predictions are preferred. Figuratively speaking, misconceptions don't ever die, they just get beaten in so many situations by confluences that they retire.

Increasing conceptual expertise may also modify misconceptions and/or the conditions under which they are retrieved or believed. As an example of modification, Sherin (2006) suggests that his students' vague belief that force implies motion (di Sessa's [1993] force-as-mover p-prim) was specialized to become the correct belief that "force implies change in velocity."

HOW DOES EXPERTISE AFFECT QUANTITATIVE PROBLEM SOLVING?

As mentioned earlier, science educators use two common indices of mastery: quantitative problem solving (Figure 16.1) and conceptual problem solving. So far, we have discussed only conceptual problem solving as it occurs either during concept inventories whose multiple-choice problems do not involve quantities (Figure 16.2), or during laboratory tasks where subjects sort, compare, or discuss, but did not solve, quantitative problems. Are there also differences between experts and novices on the second indicator, the solving of quantitative problems, such as the one in Figure 16.1?

An early finding in the expert–novice literature was that as experts solved quantitative problems, they wrote down or mentioned equations in different order than pre-novices. At first, this phenomenon was characterized as forward (experts) versus backward (pre-novices) problem-solving strategies (Larkin et al., 1980; Simon & Simon, 1978). However, later analyses characterized the orderings as grouped by principle schemas (experts) versus algebraic chaining (pre-novices) (Larkin, 1983). For instance, Larkin (1983, p. 89) says of the experts that, "in all cases, the work associated with one schema is completed before work associated with another is begun."

However, these early studies used pre-novices, that is, students who were just beginning their study of college physics. The pre-novices made so many errors that it was often difficult to compare their work to the work of experts. Moreover, the pre-novices averaged about 40 seconds per equation, whereas the experts averaged about 5 to 10 seconds per equation (Larkin, 1981), which gives one an idea of just how different their behaviors are.

On the other hand, several studies of *novices* (i.e., students who had just finished a physics course) showed that their behavior on quantitative problems was remarkably similar to expert behavior. Chi et al. (1981) found no difference in speed between experts and novices, and only a small difference in accuracy.

More tellingly, Priest and Lindsay (1992) found no difference between experts and novices in the *order* in which equations were written.

This suggests that both experts and novices have well-developed knowledge components for the equations associated with principles, whereas pre-novices are still struggling to learn the equations. As computer modeling has shown (Klenk & Forbus, 2007; Larkin, 1981), many physics problems can be solved quite efficiently given only a thorough knowledge of the equations, which would explain why novices behave so much like experts despite their lack of conceptual knowledge.

On the other hand, although experts and novices display similar equation ordering, speed, and errors, their mental processes seem quite different. When participants describe their reasoning either during or after problem solving, experts display clear plans for solutions whereas novices do not (Priest & Lindsay, 1992). This makes sense, given that experts, but not novices, tend to succeed at the planning tasks discussed earlier. All these findings are consistent with the assumption that experts have mastered many planning confluences and novices have not.

HOW CAN CONCEPTUAL EXPERTISE BE LEARNED?

The preceding sections argued that all the expert–novice findings can be explained by assuming that novices lack the expert's description phase knowledge, confluences, and applicability conditions. This section indicates possible methods of learning each of these three types of knowledge. It draws on machine learning and more specifically on the Cascade model of physics learning (VanLehn, 1999; VanLehn & Jones, 1993b; VanLehn et al., 1992). It describes how conceptual expertise *can* be learned, in that it presents machine-learning methods that can output the appropriate knowledge when given the appropriate training. It also compares the prescribed methods to current practices in physics education. In particular, it describes what would need to be changed in order to increase the number of college students who achieve conceptual expertise by the end of a year-long introductory course.

Learning Description Phase Knowledge

Description phase knowledge is essentially a set of classifications or categories. They recognize instances of categories such as forces, magnetic fields, pressure, etc. A piece of description phase knowledge says that if certain conditions exist, then a certain object or property exists as well. In machine learning, a piece of knowledge in the form "if <conditions> then <instance of class exists>" is called a classifier (Russell & Norvig, 2003).

A simple way to learn classifiers is by induction from labeled examples. For physics, an example is just a physical situation, such as a block sliding down an inclined plane. The label indicates whether the classifier applies (positive example) or does not apply (negative example). For instance, the conditions under which a normal force exists can be induced from situations where the learner is told that a normal force is present (positive examples) or told that it is absent (negative examples). Many cognitive mechanisms suffice for performing classifier induction, which is also called concept formation or category learning in the psychology (Ashby & Maddox, 2005; Medin & Ross, 1989).

Learning a classifier can take hundreds of examples, but an instructor can dramatically decrease the number of examples required via several pedagogical methods. One is to teach just one classifier at a time and to use the simplest examples possible. For instance, when teaching students to recognize forces, instructors should ideally show a situation with just one force and as few distracting details as possible. If the situation physically requires multiple forces, then the instructor should explicitly indicate which parts of the situation support the existence of each force. Students should also practice identifying concepts in isolation, and they should get feedback on their performance. By drawing forces (Heller & Reif, 1984) and energies (Van Heuvelen & Zou, 2001) in isolation, and not as part of solving a larger problem, students could probably induce those key description phase concepts with only a few dozen situations. Unfortunately, such exercises were uncommon (Hestenes, 1987), so most students may have to acquire description phase concepts by analysis of feedback on their solutions to larger problems, which could slow their learning significantly.

A second way to speed up induction is to include ample negative examples. Textbooks seldom present a situation where there are no forces, then ask the student to identify all forces. Such exercises should help.

A particularly useful technique is to present minimally contrasting pairs of examples: One example is positive, the other is negative, and they differ in only one critical feature. For instance, one can contrast a projectile moving along a curved constant speed path with one moving at constant speed along a straight path, and ask students to identify accelerations. Another contrasting pair shows two situations, both positive, but differing in a critical feature. For instance, students often believe that only animate agents can exert forces, so one can show a situation where a person's hand (animate agent) supports a motionless book, versus a situation where a table (inanimate agent) supports a motionless book. The instructor points out that there is a normal force acting on the book in both situations despite the fact that one agent is animate and the other is not. Minimally contrasting pairs are only moderately common in current physics instruction.

Learning Applicability Conditions

Applicability conditions determine when a principle schema can be applied. They also determine its possible slot fillers, for example, whether to apply Conservation of Mechanical Energy to the bomb or the dive bomber.

Applicability conditions are also classifiers, so they too can be induced from labeled examples. For instance, the dive bomber problem is a positive example for Conservation of Mechanical Energy, but a negative example for Newton's First Law. Because applicability conditions must also fill slots, they are first order categories, so more intricate induction methods may be needed (Muggleton, 1992).

Learning such applicability conditions would be simple if students were given isolated training examples instead of examples of quantitative and qualitative problem solving, which include application of schemas as only a small part of their solution. That is, given a situation, students would be shown which principles applied to which object. Later, they would be asked to list all the principle applications for a given situation and would get feedback on their choices. When Dufresne, Gerace, Hardiman, and Mestre (1992) added classification training to quantitative problem solving, conceptual expertise increased.

Learning Confluences

In contrast to the two types of knowledge discussed so far, which can be acquired by simple mechanisms, confluences can be acquired in moderately complex three-stage process. The stages correspond to three different ways of understanding the fundamental equations behind the principles: *superficially*, *semantically*, and *qualitatively*.

The learning process starts with the students acquiring a *superficial* understanding of the principle's equations. This is what novice students do now, much to the chagrin of instructors. For Conservation of Mechanical Energy, students might literally encode KE1 + PE1 = KE2 + PE2 as a string of characters. Indeed, when asked to state the principle, they may say, "Kay ee one plus pea ee one equals…." Moreover, when such students are asked, "But what is kay ee one?" they do *not* say, "kinetic energy at time one" but instead would probably reply, "Kay ee equals one-half em vee squared." Students with a superficial understanding of the principles' equations have not integrated the semantics of the generic variables into the equation.

Explanation-based learning (EBL; see Russell & Norvig, 2003) can be used to construct an equation containing expressions that refer to quantitative properties of the objects, times, etc., to which that the principle is being applied. That is, the terms inside the equations are not symbols or character strings, but are instead referring expressions similar to the mental representations of noun phrases. Moreover, embedded inside these referring expressions are references to the slots of the schemas. Thus, an equation like KE1 + PE1 = KE2 + PE2 is now understood by the student as:

> "The kinetic energy of <the object> at <time 1> + the potential energy of <the object> at <time 1> = …"where <the object> and <time 1> slots.

When the principle is applied to the dive bomber problem, <the object> is filled by "the bomb" and <time 1> is filled by "the moment that the bomb is released" so that the a problem-specific semantic equation reads:

"The kinetic energy of the bomb at the moment it is released + the potential energy of the bomb at the moment it is released = …"

It should be easy for students to acquire semantic equations since they are usually given all the information they need in the text. For instance, a statement of Conservation of Energy might include:

> … KE1 + PE1 = KE2 + PE2, where KE1 denotes the kinetic energy of the object at time 1, and PE1 denotes ….

All the student has to do is to integrate the phrase "where <symbol> denotes <expression>" into the equation. This is just what EBL would do.

However, physics students can solve quantitative problems via purely algebraic, shallow, analogical methods (Klenk & Forbus, 2007; Larkin, 1981 VanLehn, 1998; VanLehn & Jones, 1993a, 1993b). If this is the only training they get, then they have no incentive to formulate semantic equations. The superficial versions will do just fine.

In short, although students have opportunities to construct semantic equations, ordinary homework does not encourage or require it. This may explain why at the end of the semester, only some students have reached a semantic stage of understanding on some principles.

Semantic equations would be simple to learn if students were given exercises that required use of semantics. For instance, Corbett et al. (2006) gave students a problem and an equation that applied to that problem, then had students type in English descriptions for variables and expressions in the equation. Here is one of the Corbett et al. problems:

> The Pine Mountain Resort is expanding. The main lodge holds 50 guests. The management is planning to build cabins that hold six guests each. A mathematical model of this situation is $Y = 6X + 50$. What does X stand for? What does Y stand for? What does $6X$ stand for? What does 50 stand for?

Answering each question with a menu or typing should cause students to construct semantic versions of equations, and that may explain why the Corbett et al. instruction was successful compared to ordinary quantitative problem-solving practice.

The third and last stage of learning principle schemas is to construct *confluences* from the semantic equations. Again, this can be done via EBL, but it requires some background knowledge, called a generic confluence (Forbus & de Kleer, 1993). A *generic confluence* matches the form of a semantic or algebraic equation but has no domain content itself. For instance, one such generic confluence is,

> Given that the equation $?W + ?X = ?Y + ?Z$ applies to two objects, if $?W$, $?X$, and $?Z$ are the same for both objects, then so is $?Y$.

where $?W$, $?X$, $?Y$, and $?Z$ are intended to match semantic (or algebraic) expressions. For instance, they would match terms in the semantic equation of Conservation of Mechanical Energy:

> The kinetic energy of <the object> at <time 1> + the potential energy of <the object> at <time 1> = the kinetic energy of <the object> at <time 2> + the potential energy of <the object> at <time 2>

and produce a confluence, which is part of the qualitative understanding of the principle:

> If the kinetic energies of <the two objects> are the same at <time 1>, and the potential energies of <the two objects> are the same at <time 1>, and the potential energies of <the two objects> are the same at <time 2>, then the kinetic energies of <the two objects> are the same at <time 2>.

This confluence provides the key step for solving the dive bomber problem when <the two objects> is filled by the bomb from the diving plane and the bomb from the climbing plane.

In order to do such learning, students should solve qualitative problems, such as the dive bomber problem. However, they must already know the semantic equations for principles and the appropriate generic confluences. The generic confluences can either be taught explicitly or induced from experience in mathematics. When the student has both semantic equations and generic confluences, they can be applied to solve qualitative problems. EBL can then abstract the problem-specific parts away, leaving a physics-specific confluence, which is added to the principle's schema. In other words, the desired principle-specific confluences are probably acquired by specialization of generic confluences, which are themselves acquired from mathematics practice.

A similar proposal was articulated by Sherin (2001; 2006), who points out that certain knowledge components exist midway between physics-rich, principle-specific confluences and physics-free, generic confluences. He calls these knowledge components *symbolic forms*. For instance, one symbolic form, called "balancing," says that if a situation can be analyzed as two opposing force-like entities, X and Y, that are in balance, then $X = Y$, where the "=" should be understood as both a qualitative and algebraic relationship. This knowledge component has some physics content, but not as much as, say, the confluences for Newton's First Law. If students possess knowledge of symbolic forms, they may be able to use EBL to specialize them to principle-specific confluences. According to Sherin, symbolic forms develop out of di Sessa's (1993) p-prims, which are components of intuitive physics possessed even by young children. Generic confluences and symbolic forms provide two routes to the same destination: the confluences that comprise a qualitative understanding of principles.

EBL essentially just moves knowledge around, combining parts of two knowledge components to form a new knowledge component. Machine learning of one confluence is easy and requires just one example. However, there are many possible confluences to learn. For instance, given an equation with three terms (e.g., A = B*C) and a number system with two values (e.g., {zero, nonzero}), there are 24 different possible confluences such as "If B is zero and C is non-zero then A is zero" and "If A is zero and B is non-zero then C is zero." It is not clear how many qualitative physics confluences need to be constructed for a student to achieve conceptual expertise, so it is not clear how many conceptual problems they need to solve.

They probably do not need to practice each confluence separately, as there can be hundreds of confluence *per principle*, and not even experts have solved that many conceptual problems. It is more likely that experts only possess confluences for a few common cases. When they lack the appropriate confluence, they just use the semantic equation and the generic confluence instead.

Indeed, when Dee-Lucas and Larkin (1991) taught students using semantic versions for equations written in English, the students did better on conceptual problems than students taught the same material with mathematical equations. This suggests that the students already had generic confluences (which is likely, given that they were students in a highly selective technical university), and that they constructed confluences on-the-spot while answering the conceptual questions.

If learning a confluence is so easy, why don't today's students acquire conceptual expertise? Although a likely response would be that students may not be getting enough conceptual problem-solving practice, the prevalence of conceptual problems increased dramatically since 1980 and this does not seem to have cured the problem. It seems likely that students are getting conceptual problems but are solving them by some method that avoids conceptual learning. Because conceptual problems have such simple answers, it is relatively easy to memorize them. For instance, students may learn that "when two objects are dropped from the same height, they hit the ground together." This is just a slight generalization of a common conceptual problem (posed earlier with a bowling ball and a golf ball), so it is often called a *problem* schema. A problem schema suffices only for a very narrow set of problems, whereas a *principle* schema is more general. Indeed, students often answer one problem on a concept inventory correctly but miss another that, to a physicist, seems nearly identical. For instance, even if a student gets the bowling ball and golf ball problem right, they may give an incorrect answer on:

> Suppose a bowling ball and a golf ball are released at the same time from around shoulder height, but the bowling ball is somewhat higher than

the golf ball when they are released. Will the bowling ball catch up to the golf ball before they hit the ground?

The problem schema is too specific to apply to this problem.

To prevent memorization of conceptual problems as conceptual problem schemas, students should probably only be given such practice *after* they have mastered the relevant semantic equations. Many instructors feel that because conceptual problems are easier (for them), they should come before quantitative problem solving. Indeed, many high-school courses teach only conceptual problem solving. The claim here is that students should get conceptual problems only after achieving a semantic stage of understanding, which can be done with the training outlined above.

This section addressed the question, "What learning mechanisms suffice for acquiring conceptual expertise?" The proposed answer is that two learning mechanisms are involved. Induction (also called concept formation or category learning) suffices for learning description phase knowledge and applicability conditions. EBL (a form of partial evaluation or knowledge compilation) suffices for learning confluences. Both learning mechanisms can be sped up by giving learners specific types of training. Such training is not currently part of college instruction, although most of it was successful in laboratory experiments. A key experiment would be to assemble all these types of training into a multi-week experimental curriculum, and compare it to standard instruction over the same period of time.

CONCLUSION

We have argued that conceptual expertise in elementary physics consists of mastery of description phase knowledge, applicability conditions, and confluences. Description phase knowledge and applicability conditions can be induced from certain types of examples. Confluences can be learned from equations during a three-stage process: Learners first acquire a *superficial* understanding of the equation; then they construct a *semantic* version of the equation via EBL; and finally they construct a *qualitative* version via EBL, which is comprised of multiple confluences.

Because current physics instruction does not contain the right sort of training, only a few students acquire a semantic understanding of some principles, and very few attain a qualitative understanding of any principle. The training that students currently receive is mostly practice in solving quantitative and conceptual problems. Quantitative problems can be solved with only a superficial understanding of equations, and conceptual problems can be solved by memorizing problem schemas. An interesting experiment would be to replace most of the conventional problem-solving practices with the training recommended above and see if that allows more students to achieve conceptual expertise with no increase in training time.

In what other task domains is this basic path to conceptual knowledge likely to be similar? A key feature of physics is that it is mostly concerned with constructing mathematical models of situations. Such modeling occurs in many other task domains as well, because modeling is such a powerful cognitive tool. One converts a situation into a mathematical model, then "turns the crank" to produce mathematical implications, and these implications turn out to be true of the situation. In elementary physics, the models are systems of algebraic equations. In other task domains, they can be differential equations, causal networks, rule-based systems, etc. The point is that once a situation is represented in a model, it is mechanical to produce implications. However, qualitative approximations to the implications can be produced with simpler methods. That is the major finding of the QR community, a subfield of artificial intelligence. The claim here is simply that conceptual expertise consists of such QR. This is hardly a surprise to the QR community, but perhaps novel to others.

However, the power of modeling is seductive to students. They focus on the "turn the crank" parts of modeling exclusively. They find ways to circumvent the model-construction and model-interpretation processes. In particular, they find ways to solve quantitative exercises by working only with a superficial understanding of the model (e.g., the names of the variables, causal network nodes, etc. and not their meanings). Consequently, many students fail to develop a semantic understanding of the models. Without this semantic understanding, they have no way to deal with conceptual questions. Such questions are constructed so that one cannot write down a mathematical model. The only way to answer them properly is to use semantic understandings of the domain, which these students lack. Since they can't reason properly, they answer using their naïve misconceptions or their memory of previously solved conceptual questions. So the irony is that the power of mathematical modeling to produce important domain conclusions with purely mathematical, non-domain reasoning seduces learners into trying to ignore the semantics of models. Such superficial reasoning works surprisingly often on conventional analysis problems, but fails utterly on conceptual problems. Hence, conceptual expertise indicates mastery of a semantic understanding of models, which in turn can be used for both qualitative and quantitative problem solving.

Our suggestions for increasing conceptual expertise focus on increasing semantic understanding of the models. The suggested training focuses on individual pieces of a quantitative model, such as a vector, a variable, a term, an applicability condition, etc., and drills students on the denotations of each in isolation. Once a semantic understanding of the models has been mastered, it should take only a few conceptual problems to build the requisite qualitative knowledge. Current instructional practices give them too early, before students have the semantic understanding of models that will allow them to construct appropriate qualitative knowledge.

On an even more general level our work suggests that mastery and conceptual knowledge are not passively attained as a function of students' typical activity in the task domain. In particular, the development of mastery and QR within a wide range of domains, such as medicine (Boshuizen, Chapter 17; Davis, Chapter 8), law (Boshiuzen, Chapter 17), military tactics (Shadrick & Lussier, Chapter 13), music, sports, and chess (Ericsson, Chapter 18) do not emerge as automatic consequences of experience, but requires engagement in designed learning environments relying on reflective thinking and deliberate practice. To think and reason in an insightful and expert manner in a domain is, therefore, the fruit of extended efforts and is an observable characteristic of attained mastery in the relevant domain.

ACKNOWLEDGMENTS

This work is supported by NSF under grants 0325054 and 0354420. We thank Anders Ericsson, Robert Shelby, Tim Nokes, Robert Hausmann, and Jodi Davenport for their comments on drafts.

REFERENCES

Ashby, G.F., & Maddox, W.T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.

Bundy, A., Byrd, L., Luger, G., Mellish, C., & Palmer, M. (1979). *Solving mechanics problems using meta-level inference*. In Proceedings of the Sixth International Joint Conference on AI (pp. 1017–1027). San Mateo, CA: Morgan Kaufmann.

Chi, M., & VanLehn, K. (2008). Eliminating the gap between the high and low students through meta-cognitive strategy instruction. In B.P. Woolf, E. Aimeur, R. Nkambou & S.P. Lajoie (Eds.), *Intelligent tutoring systems: 9th International Conference: ITS2008* (pp. 603–613). Amsterdam, NL: IOS Press.

Chi, M.T.H. (2005). Common sense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences*, *14*, 161–199.

(in press). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of Research on Conceptual Change*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M.T.H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Chi, M.T.H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence, vol. 1* (pp. 7–75). Hillsdale, NJ: Lawrence Erlbaum Associates.

Corbett, A., Wagner, A.Z., Lesgold, S., Ulrich, H., & Stevens, S.M. (2006). The impact of learning of generating vs. selecting descriptions in analyzing algebra example solutions. In S.A. Barab, K.E. Hay & D.T. Hickey (Eds.), *The 7th International Conference of the Learning Sciences* (pp. 99–105). Mahwah, NJ: Lawrence Erlbaum Associates.

Day, D.V., & Lord, R.G. (1992). Expertise and problem categorization: The role of expert processing in organizational sense-making. *Journal of Management Studies*, *29*(1), 35–47.

de Kleer, J., & Brown, J. S. (1984). A qualitative physics based on confluences. *Artificial Intelligence*, *24*, 7–83.

Dee-Lucas, D., & Larkin, J. (1991). Equations in scientific proofs: Effects on comprehension. *American Education Research Journal*, *28*(3), 661–682.

di Sessa, A. A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, *10*(2 & 3), 105–225.

Dufresne, R. J., Gerace, W. J., Hardiman, P. T., & Mestre, J. P. (1992). Constraining novices to perform expert-like problem analyses: Effects on schema acquisition. *The Journal of the Learning Sciences*, *2*(3), 307–331.

Elio, R., & Scharf, P. B. (1990). Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization. *Cognitive Science*, *14*, 579–639.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 223–241). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273–305.

Forbus, K. D. (1985). The role of qualitative dynamics in naive physics. In J. R. Hobbes & R. C. Moore (Eds.), *Formal theories of the commonsense world* (pp. 185–226). Norwood, NJ: Ablex.

Forbus, K. D., & de Kleer, J. (1993). *Building problem solvers*. Cambridge, MA: MIT Press.

Hake, R. R. (1998). Interactive-engagement vs. traditional methods: A six-thousand student survey of mechanics test data for introductory physics students. *American Journal of Physics*, *66*(4), 64–74.

Hardiman, P. T., Dufresne, R. J., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition*, *17*(5), 627–638.

Heller, J. L., & Reif, F. (1984). Prescribing effective human problem-solving processes: Problem descriptions in physics. *Cognition and Instruction*, *1*(2), 177–216.

Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, *55*, 440–454.

Hestenes, D., & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, *30*, 159–166.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*, 141–158.

Heyworth, R. M. (1999). Procedural and conceptual knowledge of expert and novice students for the solving of a basic problem in chemistry. *International Journal of Science Education*, *21*(2), 195–211.

Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. In P. Carpenter & M. A. Just (Eds.), *Cognitive processes in comprehension*. Mahwah, NJ: Lawrence Erlbaum Associates.

Jones, R. M., & Fleischman, E. S. (2001). Cascade explains and informs the utility of fading examples to problems. In *23rd Annual Conference of the Cognitive Science Society* (pp. 459–464): Lawrence Erlbaum Associates.

Jordan, P., Makatchev, M., Pappuswamy, U., VanLehn, K., & Albacete, P. (2006). A natural language tutorial dialogue system for physics. In G. Sutcliffe & R. Goebel (Eds.), *Proceedings of the 19th International FLAIRS Conference*. Menlo Park, CA: AAAI Press.

Klenk, M., & Forbus, K.D. (2007). Cognitive modeling of analogy events in physics problem solving from examples. In D.S. McNamara & J.G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1163–1165). Austin, TX: Cognitive Science Society.

Kozma, R.B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching*, *34*(9), 949–968.

Lamberts, K. (1990). A hybrid model of learning to solve physics problems. *European Journal of Cognitive Psychology*, *3*(2), 151–170.

Larkin, J. (1981). Enriching formal knowledge: A model for learning to solve textbook physics problems. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Larkin, J. (1983). The role of problem representation in physics. In D. Gentner & A. Stevens (Eds.), *Mental models*. Mahwah, NJ: Lawrence Erlbaum Associates.

Larkin, J., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Models of competence in solving physics problems. *Cognitive Science*, *4*, 317–345.

Larkin, J., Reif, F., Carbonell, J., & Gugliotta, A. (1988). Fermi: A flexible expert reasoner with multi-domain inferencing. *Cognitive Science*, *12*(1), 101–138.

Loftin, B.R., Mueller, S., Way, B., & Lee, B. (1991). An intelligent tutoring system for physics problem solving. In *Proceedings of the contributed sessions 1991 conference on intelligent computer-aided training* (p. 127).

Mayfield, W.A., Kardash, C.M., & Kivlighan, D.M. (1999). Differences in experienced and novice counselors' knowledge structures about clients: Implications for case conceptualization. *Journal of Counseling Psychology*, *46*, 504–514.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science*, *210*(5), 1139–1141.

McDermott, J., & Larkin, J.H. (1978). Re-representing textbook physics problems. In *Proceedings of the second national conference of the Canadian society for computational studies of intelligence*. Toronto, Ontario.

Medin, D.L., & Ross, B.H. (1989). The specific character of abstract thought: Categorization, problem solving and induction. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 5* (pp. 189–223). Mahwah, NJ: Lawrence Erlbaum Associates.

Muggleton, S. (1992). *Inductive logic programming*. New York: Academic Press.

Murray, T., Schultz, K., Brown, D., & Clement, J. (1990). An analogy-based computer tutor for remediating physics misconceptions. *Interactive Learning Environments*, *1*(2), 79–101.

Novak, G.S., & Araya, A.A. (1980). Research on expert problem solving in physics. In *Proceedings of the first national conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press.

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology*, *77*(3), 272–284.

Pfundt, H., & Duit, R. (1998). *Bibliography: Students' alternative frameworks and science education, 5th ed.*. Kiel, Germany: Institute for Science Education.

Ploetzner, R., & VanLehn, K. (1997). The acquisition of informal physics knowledge during formal physics training. *Cognition and Instruction*, *15*(2), 169–206.

Priest, A.G., & Lindsay, R.O. (1992). New light on novice-expert differences in physics problem solving. *British Journal of Psychology*, *83*, 389–405.

Ranney, M., & Thagard, P. (1988). Explanatory coherence and belief revision in naive physics. In V. Patel & G.J. Groen (Eds.), *The Tenth Annual Conference of the Cognitive Science Society* (pp. 426–432). Mahwah, NJ: Lawrence Erlbaum Associates.

Reif, F., & Scott, L. A. (1999). Teaching scientific thinking skills: Students and computers coaching each other. *American Journal of Physics*, *67*(9), 819–831.
Reimann, P., Schult, T. J., & Wichmann, S. (1993). Understanding and using worked-out examples: A computational model. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 177–201). Amsterdam: Elsevier.
Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.
Shafto, P., & Coley, J. D. (2003). Development of categorization and reasoning in natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 641–649.
Sherin, B. L. (2001). How students understand physics equations. *Journal of the Learning Sciences*, *19*(4), 479–541.
(2006). Common sense clarified: The role of intuitive knowledge in physics problem solving. *Journal of Research in Science Teaching*, *43*(6), 535–555.
Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. S. Siegler (Ed.), *Children's thinking: What develops?* Mahwah, NJ: Lawrence Erlbaum Associates.
Smith, M. U. (1992). Expertise and the organization of knowledge: Unexpected differences among genetic counselors, faculty, and students on problem categorization tasks. *Journal of Research in Science Teaching*, *29*(2), 179–205.
Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, *112*, 629–661.
Van Heuvelen, A., & Zou, X. (2001). Multiple representations of work-energy processes. *American Journal of Physics*, *69*(2), 184–194.
VanLehn, K. (1996). Cognitive skill acquisition. In J. Spence, J. Darly & D. J. Foss (Eds.), *Annual Review of Psychology, Vol. 47* (pp. 513–539). Palo Alto, CA: Annual Reviews.
(1998). Analogy events: How examples are used during problem solving. *Cognitive Science*, *22*(3), 347–388.
(1999). Rule learning events in the acquisition of a complex skill: An evaluation of Cascade. *Journal of the Learning Sciences*, *8*(2), 179–221.
VanLehn, K., & Jones, R. M. (1993a). Better learners use analogical problem solving sparingly. In P. E. Utgoff (Ed.), *Machine Learning: Proceedings of the Tenth Annual Conference* (pp. 338–345). San Mateo, CA: Morgan Kaufmann.
(1993b). Learning by explaining examples to oneself: A computational model. In S. Chipman & A. Meyrowitz (Eds.), *Cognitive models of complex learning* (pp. 25–82). Boston, MA: Kluwer Academic Publishers.
VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *The Journal of the Learning Sciences*, *2*(1), 1–59.
VanLehn, K., Jordan, P., Rose, C. P., Bhembe, D., Bottner, M., Gaydos, A., et al. (2002). The architecture of Why2-Atlas: A coach for qualitative physics essay writing. In S. A. Cerri, G. Gouarderes & F. Paraguacu (Eds.), *Intelligent Tutoring Systems, 2002, 6th International Conference* (pp. 158–167). Berlin: Springer.
VanLehn, K., Lynch, C., Schultz, K., Shapiro, J. A., Shelby, R. H., Taylor, L., et al. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education*, *15*(3), 147–204.
Vosniadou, S., & Brewer, W. F. (1992). Mental models of the Earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535–585.
Weiser, M., & Shertz, J. (1983). Programming problem representation in novice and expert programmers. *International Journal of Man-Machine Studies*, *14*, 391–396.