CHAPTER 2

Adaptive Expertise as Acceleration of Future Learning

A Case Study

Kurt VanLehn and Min Chi

This chapter begins with an extensive examination of the various ways that adaptive expertise can be measured. Most of them have fairly well-known theoretical explanations, which are reviewed briefly. On the other hand, theoretical explanations are not easily found for one particularly valuable manifestation of adaptive expertise: acceleration of future learning. Acceleration of future learning is valuable because the growth of knowledge anticipated for the twenty-first-century demands that experts be able to learn new task domains quickly. That is, their training *now* should raise their learning rates *later*: It accelerates their future learning.

We present a case study where accelerated future learning was achieved. The trick was to use an intelligent tutoring system that focused students on learning domain principles. Students in this condition of the experiment apparently realized that principles were more easily learned and more effective than problem schemas, analogies, and so forth. Thus, when given the freedom to choose their own learning strategy while learning a second task domain, they seem to have focused on the principles of the new task domain. This caused them to learn faster than the control group, who were not focused on principles during their instruction on the initial task domain. In short, the metacognitive learning strategy/policy of focusing on principles seems to have transferred from one domain (probability) to another (physics), thus causing accelerated future learning of the second task domain (physics).

A Framework for Understanding Adaptive Expertise

Adaptive expertise has been discussed for at least twenty-five years (Hatano & Inagaki, 1986), and there seems to be consensus on what differentiates it from routine expertise. A widely quoted definition is: “Whereas routine experts are able to solve familiar types of problems quickly and accurately, they have only modest capabilities in dealing with novel types of problems. Adaptive experts, on the other hand, may be able to invent new procedures derived from their expert knowledge" (Holyoak, 1991, p. 312). According to Hatano and Inagaki (1986), a routine expert has mere procedural skill whereas an adaptive expert has some type of conceptual understanding of the procedures as well. Whatever this conceptual understanding may be, it allows the adaptive expert to be more effective in nonroutine situations than the routine expert.

This introduction will present a framework for understanding adaptive expertise by first presenting a fairly mundane classification system for studies. It classifies studies first along their design (cross-sectional vs. training) and second along their method of assessment. Although these are textbook classifications that can be applied to almost any experiment, they set the stage for speculating on the underlying knowledge structures that differentiate adaptive expertise from routine expertise. It will turn out that almost every type of study is easily interpreted using familiar theoretical concepts such as schemas and domain principles. However, one type of study (a training study that uses acceleration of future learning as its assessment) is not so easily explained. Thus, we conducted a study of that type, and describing it occupies the rest of the chapter.

Dimension 1: Cross-Sectional versus Training Studies

Two study designs are commonly used for understanding adaptive expertise: cross-sectional and training. The cross-sectional design is like an expert-novice study, except that the groups being compared are both experts: routine versus adaptive experts. For instance, when Barnett and Koslowski (2002) compared the behaviors of two groups of experts – experienced restaurant managers and business consultants – they found that the business consultants were more adaptive than the restaurant managers. Other studies compare experts with varying degrees of experience (e.g., Lesgold et al., 1988).

The other common design is a training study, where experimental and control groups are both trained to mastery on a small task domain, then tested for adaptive expertise. For instance, Pandy, Petrosino, Ausin, and Barr (2004) studied adaptive expertise in biomechanics that developed after a training period of about three hours.

The advantage of cross-sectional studies is that they focus on knowledge and skills that unarguably qualify as expertise, whereas training studies must use small task domains and short training periods in order to be feasible. On the other hand, training studies focus on instructions and conditions that can cause trainees to acquire adaptive expertise as opposed to routine expertise. This is an important question that cross-sectional studies address only indirectly at best. The chapter focuses exclusively on training studies of adaptive expertise.

Dimension 2: Methods for Assessing Adaptive Expertise

The difference between adaptive and routine expertise can be subtle, so methods of assessment are an important dimension for classifying studies of adaptive expertise. This section presents a simple taxonomy, which is summarized later in Figure 2.2.

The Hatano/Holyoak definition of adaptive expertise refers to a single type of assessment, wherein participants are asked to solve problems that are novel and cannot be solved with routine problem-solving procedures. In the context of training studies, such problems are often called far-transfer problems. As participants are solving the far-transfer problems, they are not allowed to access instructional resources. That is, they are *sequestered*, as are American juries, which led Bransford and Schwartz (1999) to call this "sequestered problem solving."

Bransford and Schwartz point out that transfer can also be measured by replacing sequestered problem solving with a situation where students can learn, a technique called *dynamic assessment* (Haywood & Tzuriel, 2002). That is, participants learn twice. First, they acquire some prior knowledge, then they are instructed in a second task domain as their progress is monitored.

Progress monitoring typically measures the rate of learning, the amount of help required during learning, or both. If the two task domains have the right relationship, then adaptive experts should master the second task domain faster and with less help than routine experts.

In addition to measures of transfer, some studies of adaptive expertise have used measures of the conceptual quality of solutions (e.g., Martin, Petrosino, Rivale, & Diller, 2006). This assessment assumes the theoretical position mentioned earlier, namely that routine experts have mere procedural skills whereas adaptive experts have a conceptual understanding of their procedures. The assessment procedure is merely to ask participates to solve problems that both routine and adaptive experts find challenging but solvable. Students either speak aloud or provide rich explanations of their reasoning in some other way. The participants' utterances are coded for conceptual depth and quality.

So far, three classes of assessment have been defined: (1) sequestered assessment: solving far-transfer problems without instructional resources or other help; (2) dynamic assessment: measuring how fast students learn or how much help they need while learning; (3) direct conceptual assessment. Although there is little more to say about the third type of assessment, there is much to discuss about the first two, so the discussion is broken into two subsections.

Sequestered Assessment of Adaptive Expertise

One way to measure adaptive expertise involves training students on one kind of task then giving students transfer tasks to solve in a sequestered situation. Bransford and Schwartz (1999) call this type of transfer Direct Application.

Direct Application also can be subclassified according the relationship between the transfer tasks and the training tasks. A standard theoretical framework is to assume that students acquire problem-solving schemas that have retrieval cues, slots, and solution procedures, and these correspond to difficulties with *access*, *application*, and *adaptation* of the schemas. Let us examine each briefly.

Access-impeding transfer problems have features that are not typical of the training tasks, and some students have trouble even noticing that their schemas are relevant. For instance, Gick and Holyoak (1983) showed that students who knew how to apply Dunker's convergence schema in one context (marching armies across weak bridges over a circular moat) failed to notice its application to an X-ray convergence problems. However, if reminded of the schema, they were easily able to solve the X-ray problem.

Application-impeding transfer occurs when a known schema has several slots, and students have difficulty figuring out which slots go with which objects in the problem. It can also occur when students have accessed two or more similar schemas and are not sure which one to apply. For instance, several studies first taught students schemas for permutation and combination (Catrambone, 1994; Ross, 1987; VanderStoep & Seifert, 1993). These two schemas have similar applicability conditions and each schema has two numerical slots. The training problems are designed so that irrelevant features (e.g., students choosing lockers; balls placed in urns) are strongly correlated with the choice of schema and the filling of slots. Students are trained to mastery on such problems, so in a sense, they have become routine experts on these schemas. The transfer problems use new features or reverse the correlations, and this causes such students difficulty.

Lastly, adaptation-impeding transfer problems require that students modify the solution procedure. This is the only assessment method mentioned by the Hatano/Holyoak definition. To do such modifications, the expert must access some other kind of knowledge than the schema itself. As an illustration of such "extra knowledge" and how it can be used to modify a procedure, let us consider an especially obvious case. Kieras and Bovair (1984) taught two groups of subjects some complex procedures for operating an unfamiliar device (the phasor cannons on the Starship *Enterprise*). One group of subjects could access a schematic of the device as they learned. They were able to invent shortcuts to the procedures. The group that knew nothing of the schematic performed like routine experts; they could follow the procedures accurately and quickly, but could not modify them. Clearly, the additional knowledge possessed by the adaptive experts was the schematic of the device.

When experts are required to operate or troubleshoot a physical device or system, it is widely believed that they must understand its anatomy (structure) and physiology (function) in addition to mastering the standard operating procedures. Such "mental models" are a standard explanation for adaptive expertise in medicine, troubleshooting, and other domains where participants work with physical systems or devices.

Some mathematical and scientific task domains include explicit domain principles that are composed to form schemas. For instance, in kinematics, there are three common schemas for identifying acceleration. They apply when objects are (1) moving in a straight line and speeding up, (2) moving in a straight line and slowing down, and (3) moving in a circle. When experts are asked to identify the acceleration of a pendulum during the ascending part of its swing, routine experts apply the third schema, circular motion, which produces an incorrect answer (Reif & Allen, 1992). Adaptive experts produce a correct answer by retreating to first principles, namely the definition of acceleration as the first derivative of velocity. This behavior suggests that the "extra knowledge" that adaptive experts use is merely knowledge of the first principles of the task domain. This obviously is not quite right, as all experts "know" the first principles of their task domain; apparently, that knowledge is more inert for routine experts than adaptive experts. Perhaps the adaptive expert's retrieval cues for accessing first principles are more general, stronger, or somehow better than those of the routine expert. Anyway, we have at least the outline of an explanation for adaptive expertise in principle-rich task domains when such expertise is assessed by adaptation-impeding far-transfer problems.

As illustrated by these two types of task domain, adaptive experts modify their schemas' procedures by using extra knowledge. For procedures that operate or troubleshoot systems, that extra knowledge was a mental model: the structure and function of the system. For principle-rich math and science task domains, it was noninert knowledge of the domain's first principles. Unless a task domain's procedures are completely arbitrary, they have some rational derivation, so adaptive experts may simply possess noninert knowledge of the original derivations of the procedures that they follow, whereas routine experts have somehow lost access to such knowledge or never have learned it.

Dynamic Assessment of Adaptive Expertise

Dynamic assessment features two learning periods, and the second learning period is monitored to provide an assessment of the knowledge produced during the first learning period. Logically, transfer measured via dynamic assessment can be divided into two types based on the students' performance during the second learning period. *Savings* occurs when some students have a head start in learning. *Acceleration of Future Learning* (AFL) occurs when prior knowledge increases the rate at which some students learn, relative to others. These are independent measures, so students could exhibit both Savings and AFL. For instance, suppose we are comparing two methods to train a skill, A and B. Figure 2.1 maps our terminology onto the relative performance of A (dashed line) and B (solid line) during the second learning period. The next few paragraphs present more concrete illustrations of Savings and AFL, albeit not in the context of measuring adaptive expertise.

Singley and Anderson (1989) monitored students as they learned the Microsoft Word text editing. Half of the students were familiar with WordPerfect, which is similar to Word, and the other half were familiar with

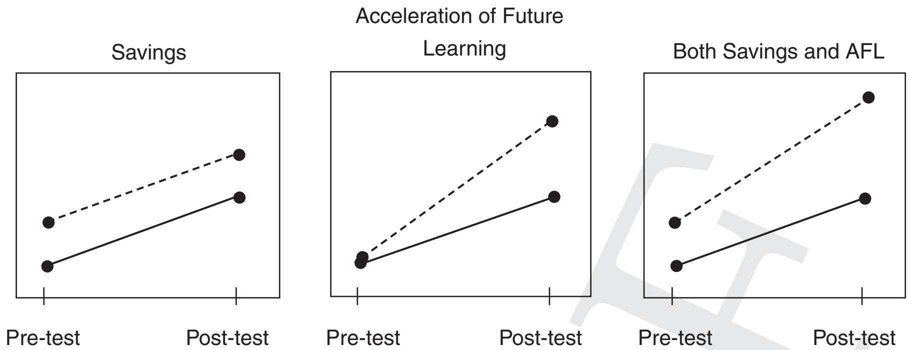

Figure 2.1. Performance during the second learning period for three types of transfer.

a line-oriented editor. Singley and Anderson observed Savings but not AFL That is, the students who knew WordPerfect did much better than the other students on a pre-test of Word skills, and they maintained exactly that advantage and no more throughout their training on Word. As Thorndike, Anderson, and many others have pointed out, this kind of transfer is easily explained by assuming that the two bodies of knowledge (Word and WordPerfect usage, in this case) share some elements. When students have mastered one, they have a head start on learning the other.

An example of AFL is Slotta and Chi's (2006) study. They taught half their students the emergence schema, which is an explanatory schema for certain types of scientific phenomena. Students were then taught about the flow of electricity in wires, an emergent phenomena. The electricity pre-test scores of the two groups were not significantly different, but the post-test scores of the emergence group were much higher than the post-test scores of the control group. Thus, this is evidence of AFL without Savings.

AFL can also be caused by teaching metacognitive skills and learning-to-learn skills (Hattie, Biggs, & Purdie, 1996). These are domain-independent skills (e.g., self-monitoring, note taking, self-explanation), in that they can in principle accelerate the learning of almost any content. If domain-independent skills were the whole explanation for the competence of adaptive experts, then such people would be universal experts: Adaptive experts could learn *any* content much more rapidly than routine experts. Thus, the secret of adaptive expertise probably lies more toward domain-specific knowledge, such as the emergence schema taught by Slotta and Chi.

A type of AFL, called Preparation for Future Learning (Bransford & Schwartz, 1999; Schwartz & Bransford, 1998; Schwartz, Bransford, & Sears, 2005), focuses on just one task domain but still uses two learning periods. During the first learning period, students engage is some kind of preparatory activity that does not involve explicit domain instruction. During the second learning period, their knowledge is dynamically assessed. For instance, during the first learning period, Schwartz and Martin (2004) had the experimental group try to answer the question, "Was Bill a better high-jumper than Joe was a long-jumper?" Students had not yet been taught about z-scores, but they were given the appropriate distributions of high jumpers and long jumpers' performance, so they could have invented the z-score concept, although none did. Thus, this preparatory exercise did not contain explicit instruction on z-scores, nor did students successfully discover the z-score concept. On the other hand, the control group engaged in a filler activity during the first learning period. During the second learning period, both groups were taught explicitly about z-scores. The experimental group's learning was faster during the second learning period than the control group's learning. This experiment demonstrated Preparation

- Direct measures of conceptual understanding
- Sequestered assessment
  - Access-impeding problems
  - Application-impeding problems
  - Adaptation-impeding problems
- Dynamic assessment
  - Savings
  - Acceleration of future learning (AFL)
    - Transfer of meta-cognitive or learning-to-learn skills
    - Transfer of explicitly taught knowledge
    - Preparation for future learning (PFL)
    - Other

Figure 2.2. A taxonomy of assessments of adaptive expertise.

for Future Learning. Unlike the first earlier types of AFL where the to-be-transferred knowledge was explicitly taught (e.g., a metacognitive skill, a learning-to-learn skill, or Chi's emergence schema), Preparation for Future Learning does not teach anything explicitly, so it is less clear what knowledge is being transferred. What transfers may not be knowledge per se, but instead be interest, commitment, or some other persistent affective state.

In summary, we have defined a taxonomy of commonly used adaptive expertise assessments, shown in Figure 2.2. As mentioned earlier, AFL is an especially important manifestation of adaptive expertise, because as the amount of knowledge increases in the twenty-first century, experts will repeatedly be called on to master new knowledge. Moreover, as the work on Preparation for Future Learning shows, there are aspects of AFL that are not well understood. Thus, we chose to focus on AFL for the study described next.

The Study's Design and Findings

The study is complex, but it can be easily described now that the taxonomy of studies of adaptive expertise has been presented. This study is a training study rather than a contrast of two types of experts. It used both sequestered and dynamic assessment. The sequestered assessment used application-impeding and adaptation-impeding transfer problems, but it did not use access-impeding transfer problems. The dynamic assessment showed both Savings and AFL.

On all these measures, the experimental group performed significantly better than the control group. In the context of this experiment, adaptive expertise turned out to be a unified phenomenon: Whatever it was that the experimental group learned and the control group did not learn, that knowledge, meta-knowledge, motivation, and so on probably caused them to do better on *every* assessment.

Although the success of the experimental group is surprising, even more surprising is that the knowledge that we taught them, and not the control group, does not seem to be the source of their success. This raises the very interesting question: What did they transfer? Why did they become such adaptive experts?

Rather than extend this introduction, which is quite long already, we will describe the experiment itself. Afterward, we will discuss what it implies for adaptive expertise.

Methods

Many scientific task domains have equation-based principles. For instance, mechanics has a principle called Conservation of Mechanical Energy, which states that under certain conditions, the total mechanical energy of a system does not change over time. This can be more formally expressed by the equation TME1 = TME2, where TME1 and TME2 are the total mechanical energy of the system at times 1 and 2, respectively. Similarly, in the task domain of elementary probability, the addition law says that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

In such task domains, a common way to get students to learn the principles is to give them problems where they are expected to apply some principles, then solve the resulting set of equations algebraically. That is, each application of a principle generates an equation expressed in terms of problem-specific variables. Some of those variables have already been given values in the statement of the problem. The student's job is to write enough equations so that a certain variable, often called the sought variable, can have its value determined algebraically. Such problems are common in textbooks for introductory physics, probability, circuits, thermodynamics, statics, and many other task domains.

Although it is not widely known, there is an effective, complete strategy for solving such problems. It is a specialization of a well-known weak method for deductive problem solving, called backwards chaining (Russell & Norvig, 2003). As far as we know, Bhaskar and Simon (1977) were the first to study its properties, but it is undoubtedly older than that. The basic idea is relatively straightforward:

1. If there are no goal variables left, then go to step 4. Otherwise, pick a goal variable as the target variable.
2. Considering all the principle applications that are possible for this problem and have not yet been done, pick one whose equation contains the target variable. Write that equation down.
3. Some of the variables in the equation may have known values and some may be goal variables now or in the past. If there are any others that are neither known nor goal variables, then add them to the set of goal variables. Go to step 1.
4. Solve the set of equations that have been written down. This is most easily done via substitution, working backward through the list of equations.

We call this strategy the Target Variable Strategy, and we implemented an intelligent tutoring system called Pyrenees (VanLehn et al., 2004), which teaches students how to apply it. Pyrenees also teaches students the task domain's principles. From the students' point of view, learning the principles is probably much more difficult than learning the Target Variable Procedure. There are often dozens of principles, and some of the principles have subtle conditions on when they may and may not be applied. We view Pyrenees mostly as scaffolding for teaching principles. The Target Variable Strategy is part of that scaffolding.

The hypothesis tested by this experiment is that the Target Variable Strategy can be transferred. That is, when students learn one task domain using the Target Variable Strategy and Pyrenees, then they will more rapidly learn a second task domain without the aid of Pyrenees because they will apply the Target Variable Strategy by themselves. The hypothesis really hinges on motivation. That is, after several hours of following the Target Variable Strategy while using Pyrenees, they have probably mastered the procedure. However, they may or may not be motivated to use it in a new task domain. To paraphrase Pintrich and de Groot (1990), they have the *skill* but do they have the *will*?

Design

Two task domains, probability and physics, were taught. As students learned the first task domain, probability, half the students learned the Target Variable Strategy. The rest of the students were free to solve problems any way they liked. We call these two groups the strategy and no-strategy groups. The second task domain, physics, was taught the same way to all students and was used to

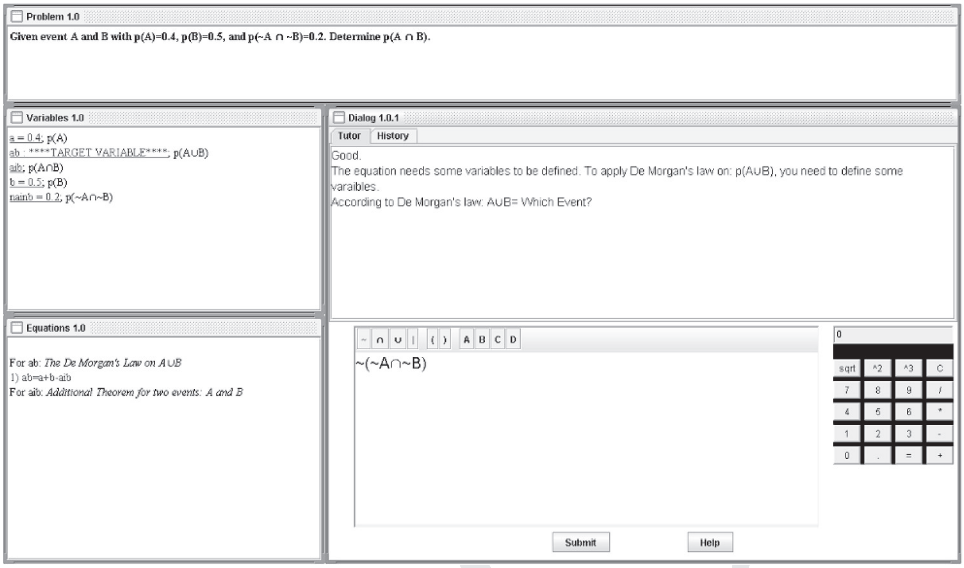

Figure 2.3. The Pyrenees screen. Clockwise from the top: windows display the problem statement, dialogue with the tutor, a workspace for composing and submitting an entry, equations and variables.

measure AFL. Each domain contained ten major principles.

Participants

Data were collected over four months during the fall of 2005 and the early spring of 2006. We recruited ninety-one college students. They were required to have basic knowledge of high-school algebra, but not to have taken college-level statistics or physics courses. Students were randomly assigned to the strategy and no-strategy groups. Each student took from two to three weeks to complete the study over multiple sessions. All the materials were online, and when students came back each time, they would continue from the point where they left off previously. Because of the winter break and length of the experiment, only forty-four participants completed the experiment. Two students were eliminated because of a perfect performance on the probability pre-test and a lack of time consistency, respectively. Of the remaining forty-two participants (59.5% female), twenty were assigned to the strategy group and twenty-two to the no-strategy group.

Three Intelligent Tutoring Systems

To control and measure the students' learning, three Intelligent Tutoring Systems were used: two for probability instruction, Pyrenees and Andes-probability, and one for physics, Andes-physics. Apart from the declarative knowledge, Andes-probability and Andes-physics were identical, and we will use Andes to refer to both of them. All the tutoring systems provided a screen that consisted of a problem statement window, a variable window for listing defined variables, an equation window, and a dialog window (see Figure 2.3).

Both Andes and Pyrenees are step-based tutoring systems (VanLehn, 2006). That is, the student can get feedback and hints on every step leading up to the final answer. Pyrenees explicitly taught the Target Variable Strategy and required the student to follow it during problem solving, so at any given time, only a few steps were acceptable to it, namely the steps consistent with the Target Variable Strategy. On the other hand, Andes provided no explicit strategic instruction, nor did it require students to follow any particular strategy. Students using Andes could input any entry, and Andes colored it green if it was correct and red if it was incorrect. Students could enter an equation that was the algebraic combination of several principle applications on Andes but not on Pyrenees.

Beside providing immediate feedback, both Pyrenees and Andes provided help when students asked. When an entry was incorrect, students could either fix it on their own or ask for *what's-wrong help*. When they did not know what to do next, they could ask for *next-step help*. Both Pyrenees and Andes gave the same *what's-wrong help*, but their *next-step help* differed. Because Pyrenees required students to follow the Target Variable Strategy, it knew exactly what step the student should be doing next so it gave specific hints. In Andes, on the other hand, students could enter correct equations in any order, and an equation was considered correct if it was true, regardless of whether it was useful for solving the problem. So Andes did not attempt to figure out the student's problem-solving plans or intentions. Instead, it picked a step that it would most like to do next and hinted at that step. Both *next-step help* and *what's-wrong help* were provided via a sequence of hints that gradually increased in specificity. The last hint in the sequence, called the *bottom-out hint*, told the student exactly what to do.

In summary, the strategy students were required to study and use the Target Variable Strategy to learn probability on Pyrenees and then learned physics on Andes-physics. The no-strategy students were not required to use any specific strategy and they learned both probability and physics on Andes.

Procedure

The study had four main parts: a background survey, probability instruction, Andes interface training, and physics instruction (see Table 2.1, left column). The background survey included high-school GPA, Math SAT, Verbal SAT score, and experience with algebra, probability, and physics, as well as other information.

The probability instruction had the five main phases: (1) pre-training, (2) pre-test, (3) watching a video introduction to the tutoring system, (4) training, and (5) post-test. The pre-training phase taught students individual probability principles, whereas the training phase taught students how to solve moderately complex problems involving multiple-principle applications. The experimental manipulation occurred only during the training phase. The pre-test and post-test used problems similar to the training problems. The pre-test was given *after* the pre-training and just before the training phase, so that the tests would more reliably measure differences in learning due to the manipulation, which occurred only during the training phase.

During phase 1 (pre-training), students studied the domain principles. For each principle, they read a general description, reviewed some examples, and solved some single- and multiple-principle problems. During this training, students were tutored by an answer-based tutoring system (VanLehn, 2006). For instance, suppose the system asked them to solve a problem: "Given that A and B are mutually exclusive events such that $p(A) = 0.3$ and $p(B) = 0.4$, what is the value of $p(A \cap B)$?" If the students enter the correct answer, 0, the system would show it in green; if it was incorrect, the system would show it in red and then ask the student to solve an isomorphic problem. They will not go onto the next domain principle or solve the next problem until they have solved such a problem correctly. If they failed three times, then the correct answer (which is similar in all three problems) was explained to them and they were allowed to move on. The purpose of pre-training was to familiarize students with the principles, not necessarily to master them. They were expected to continue learning about the principles during the training phase as it taught them how to apply several principles in combination order to solve moderately complex problems.

During phase 2, students took the pre-test. Students were not given feedback on

Table 2.1. Procedure for the Experiment

	Strategy Group	No-Strategy Group
Background Survey	Background survey	
Probability instruction	Probability pre-training	
	Probability pre-test	
	Pyrenees video	Andes-Probability video
	Probability Training on Pyrenees	Probability Training on Andes-Probability
	Probability post-test	
Andes interface training	Andes-Probability video	
	Solve a problem with Andes-Probability.	
Physics instruction	Physics pre-training	
	Physics pre-test	
	Andes-Physics video	
	Physics Training on Andes-Physics	
	Physics Post-test	

their answers or allowed to go back to earlier questions. (This was also true for the post-tests.)

During phase 3, students watched a video about how to use their tutoring system. In particular, the strategy group read a text description of the Target Variable Strategy, then watched a video on how use the Pyrenees. The no-strategy group watched a video on how to use the Andes.

During phase 4, students in both conditions solved the same twelve probability problems in the same order on either Pyrenees (strategy condition) or Andes (no-strategy condition). Each main domain principle was applied at least twice. All students could access the corresponding textbook and the strategy students could also access the description of the Target Variable Strategy. Finally, students took the post-test.

After completing the probability instruction, the strategy students were taught how to use Andes without introducing any new domain knowledge. The user interface training, which was the same as the one given earlier to the no-strategy students, comprised a video and a probability problem to be solved. The problem was one of the twelve problems that the strategy students had solved earlier on Pyrenees. The pilot studies showed that solving one probability problem on Andes was enough for most students to become familiar with the Andes' user interface.

Finally, all students learned physics using exactly the same instruction. The instruction consisted of the same five phases as the probability instruction. That is, it consisted of (1) pre-training on individual principles, (2) pre-testing, (3) user interface training for Andes-Physics, (4) training on solving moderately complex problems using Andes, and (4) post-testing. To measure AFL, the strategy group and the no-strategy group received exactly the same physics instruction.

Table 2.2 shows the number of single- and multiple-principle problems in the experiment. In each post-test, five of the multiple-principle problems were isomorphic to training problems in phase 4. These functioned as near-transfer problems for assessing routine expertise. The other post-test problems (five for probability; eight for physics) were novel, nonisomorphic multiple-principle problems. Most of the multiple-principle problems had dead-end search paths so that the Target Variable Strategy could show an advantage in search efficiency. These functioned as far-transfer problems for sequestered assessment of adaptive expertise.

Table 2.2. Numbers of Problems for Each Phase of the Experiment

		Single-Principle Problems	Multiple-Principle Problems	Total Problems
Probability	Pre-Training	14	5	19
	Pre-test	10	4	14
	Training	0	12	12
	Post-test	10	10	20
Physics	Pre-Training	11	3	14
	Pre-test	9	5	14
	Training	0	8	8
	Post-test	5	13	18

To summarize, the main procedural difference between the two conditions was that during the probability instruction, the strategy students used Pyrenees while the no-strategy students used Andes. However, because familiarity with Andes might give the no-strategy students an unfair advantage during the physics instruction, where both groups used Andes, the strategy students were given a brief introduction to the Andes' user interface before the physics instruction.

Scoring Criteria

Three scoring rubrics were used: binary, partial credit, and one point per principle. Under the binary rubric, a solution was worth 1 point if it was completely correct and 0 otherwise. Under the partial-credit scoring rubric, the score was the proportion of correct principle applications evident in the student's solution – a student who correctly applied four of five possible principles would get a score of 0.8. One-point-per-principle scoring rubric gave a student a point for each correct principle application. Solutions were scored by a single grader who did not know which student or condition the solution came from. For comparison purposes, all of the scores were normalized to fall in the range of [0,1].

Results

Several measures showed that the incoming student competence was balanced across conditions: (1) there was no significant difference on the background survey between two conditions; (2) the two groups did not differ on the probability pre-test with respect to their scores on single-principle, multiple-principle, and overall problems across all three scoring rubrics; (3) during the probability pre-training, wherein students solved problems embedded in the textbook, the conditions did not differ on single-principle, multiple-principle, and overall scores. Thus, despite the high attrition, the conditions remained balanced in terms of incoming competence.

The two conditions did not differ on any of the training times: (1) the probability pre-training; (2) probability training on either Pyrenees or Andes-probability; (3) physics pre-training; and (4) physics training on Andes-physics. This is fortuitous as it implies that any difference in post-training test scores was due to the effectiveness of the instruction rather than differences in time-on-task.

The main outcome (dependent) variables are the students' scores on the probability post-test, the physics pre-training, the physics pre-test, and the physics post-test. We discuss each in turn.

First, on the probability post-test using the binary scoring rubric, the strategy students scored significantly higher than the no-strategy students: $t(40) = 3.765$; $p = 0.001$ (see Figure 2.4). Cohen's effect size (difference in mean post-test scores divided by pooled standard deviation) was 1.17, and is denoted "d" subsequently. Moreover, the

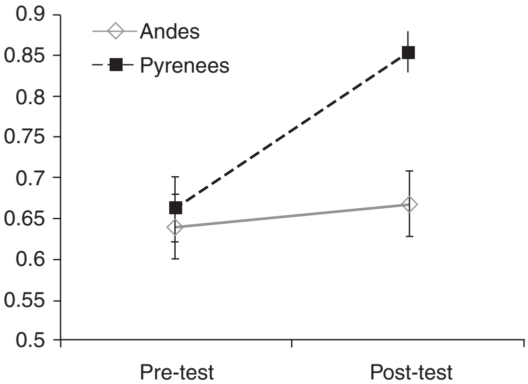

Figure 2.4. Results from the first learning period (probability task domain).

strategy students scored higher than the no-strategy students on both single-principle problems, $t(40)=3.960$; $p<0.001$; $d=1.24$, and multiple-principle ones, $t(40)=2.829$; $p=0.007$; $d=0.87$. The same pattern was found with using the partial-credit and one-point-per-principle scoring rubrics. Thus, the strategy students learned more probability than the no-strategy students.

Next we consider performance during the physics pre-training, reporting the binary scoring rubric only. The strategy students solved significantly more single-principle problems correctly in the first try than the no-strategy students: $t(40)=2.072$, $p=0.045$; $d=0.64$. No significant difference was found between the two groups on the number of multiple-principle problems solved correctly. This result could be due to an unlucky assignment of students to conditions, so that the strategy students happened to know more physics before the experiment began. Although we cannot rule this interpretation out, it does seem unlikely due to the use of random assignment and to the lack of difference in background questionnaire items that asked about the students' physics experience and grades. It is more likely that the strategy students were, even at this early stage, learning physics faster than the no-strategy students. This is consistent with the results presented next.

On the physics pre-test, the strategy students scored higher than the no-strategy students under the binary scoring rubric, $t(40)=2.217$, $p=0.032$, $d=0.69$. The same pattern was found with the other two scoring rubrics. On the single-principle problems, the two conditions did not differ significantly regardless of the scoring rubric, probably due to a ceiling effect. For example, under the binary scoring rubric, we have M = .93, SD = .097 (maximum is 1) for the strategy students and M = .86, SD = .11 for the no-strategy students. On the multiple-principle problems, the strategy students scored higher than the no-strategy students on the partial-credit rubric, $t(40)=2.913$, $p=0.0058$, $d=0.90$ and one-point-per-principle rubric $t(40)=.800$, $p=0.008$, $d=0.86$, but not on the binary rubric $t(40)=1.148$, $p=0.147$. This could be due to the inherently less sensitive nature of the binary rubric. Anyway, the overall physics pre-test scores indicate that the strategy students learned more effectively during the physics pre-training than the no-strategy students. It appears that the strategy training during probability accelerated the strategy group's future learning, that is, their physics learning.

On the last assessment, the physics post-test, the strategy students' score was much larger than the no-strategy student's scores under the binary scoring rubric, $t(40)=4.130$, $p<0.0002$, $d=1.28$ and the two other rubrics (see Figure 2.5). More specifically, the strategy students outperformed the no-strategy students on single-principle problems under the binary rubric, $t(40)=3.211$, $p=0.003$, $d=1.00$ and on the multiple-principle problems as well, $t(40)=3.395$, $p<0.001$, $d=1.23$. Similar pattern was also found under the other two rubrics. Thus, the strategy students scored much higher than the no-strategy students on the physics post-test scores as well as on the single- and multiple-principle problems under all scoring rubrics.

The results presented so far are consistent with two kinds of transfer: *Savings*: the strategy students had a head start over the no-strategy students when learning physics; and *AFL*: the strategy students learned physics faster than the no-strategy students. To determine whether it was a head start or a learning rate, we ran an ANCOVA on the physics post-test scores using the physics pre-test scores as a covariant. This yielded a post-test

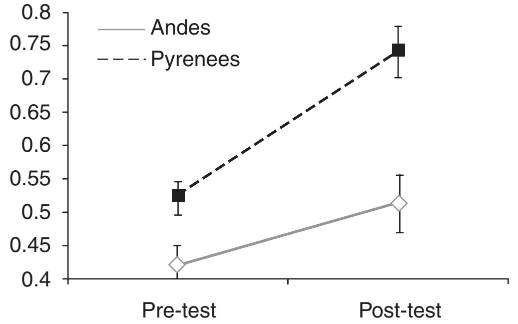

Figure 2.5. Results from the second learning period (physics task domain).

score for each student, adjusted for the difference in his/her physics pre-test score. With the binary scoring rubric, the strategy students had higher adjusted post-test score (M = 0.705, SD = 0.21) than the no-strategy students (M = 0.478, SD = 0.22). This difference was large and reliable $F(39) = 11.079$, $p = 0.002$, $d = 1.05$. A similar pattern held for the other rubrics: $F(39) = 6.155$, $p = 0.0175$, $d = 0.81$ for the partial-credit rubric and $F(39) = 5.290$, $p = 0.0269$, $d = 0.75$ for the one-point-per-principle rubric. This suggests that learning the Target Variable strategy in one task domain caused AFL of the second task domain. This is intuitively satisfying, as we chose task domains that had very little overlap, making a Savings-style transfer unlikely.

What Was Transferred?

The overarching goal of our study was to determine whether teaching students the Target Variable Strategy explicitly in one deductive task would improve their learning performance not only in that task domain, but also in a new one. We found evidence for learning in both task domains. First of all, consistent with our previous study of physics (VanLehn et al., 2004), the Target Variable Strategy improved students' learning significantly in the initial domain, probability. Because the Target Variable Strategy increased learning in two initial domains, it may be an effective strategy in other deductive domains as well. Second, teaching students the Target Variable Strategy in probability also significantly accelerated their learning during a second learning period, when they were not constrained to follow it.

Let us consider three hypotheses about *why* teaching the Target Variable Strategy improved learning in both domains. In particular, what did students learn during probability instruction that they transferred and used to improve their learning of physics? It will turn out that only the third hypothesis is supported by the data.

The first hypothesis, unsurprisingly, is that the strategy students learned the Target Variable Strategy and that they applied it during physics problems solving, which improved their search efficiency. The multiple-principle problems were constructed so that using the strategy would reduce the average number of steps required to solve the problems, and hence reduce both time and the likelihood of error. If the search-efficiency hypothesis is true, the strategy students should have performed better than the no-strategy students on the multiple-principle problems but not on the single-principle problems, where search is not required. However, the latter prediction was false (see Table 2.2). The strategy students outscored no-strategy students on single-principle post-test problems in both probability and physics.

The second hypothesis is that strategy students learned probability problem schemas better than no-strategy students, and that (somehow) they also learned physics problem schemas better than no-strategy students. A problem schema is a generalization formed as a by-product of problem solving by analogy that matches a whole problem and proposes a whole solution (VanLehn, 1989). Although students can construct problem schemas from only a few examples and are quite facile at using them to solve near-transfer problems, problem schemas often will not solve far-transfer problems (Gerjets, Scheiter, & Catrambone, 2004). Suppose the strategy students were better at learning problem schemas than the no-strategy students. This hypothesis would explain the strategy students' superior performance on near-transfer problems (i.e., test problems isomorphic to the training problems). However, this hypothesis predicts that on far-transfer problems, the two groups should perform equally poorly. The latter prediction is false. On the nonisomorphic multiple-principle problems (five in probability post-test and eight in physics post-test), the strategy students performed significantly better than the no-strategy students: $t(40) = 2.27$, $p = 0.029$ in probability post-test and $t(40) = 3.803$, $p < 0.0005$ in physics post-test.

The third hypothesis is that the strategy group learned both probability and physics principles better than the no-strategy group. The knowledge that was transferred could not be the principles themselves, because probability and physics share no principles. What may have transferred was the meta-knowledge that principles were the best thing to focus attention on when learning a new domain. That is, instead of trying to recall problem schemas, solved problems, equations, definitions of terms, and so forth, one should just focus on learning the principles.

Although this may seem implausible, examining the details of the strategy versus no-strategy training suggests otherwise. In Pyrenees, the Target Variable Strategy requires students to apply one principle at a time by selecting the desired principle from a menu, whereas Andes did not force them to apply one principle at a time, and seldom had students select principles from a menu. In Pyrenees, when students made a mistake, they got principle-specific feedback, but such feedback was rarely given by Andes because the no-strategy policy made it difficult for Andes to determine what principle the student was trying (and failing) to apply. Essentially, Pyrenees taught students over and over that if they could recall and apply the principles accurately, the Target Variable Strategy would take care of the rest of the problem solving. In particular, there was no use doing the usual schema building: recalling a solved problem, forming an analogy to this problem, mapping over the old problem's solution, using it to solve this problem, and forming a general schema that paired a general problem description with a generalization of the solution. This focus on principles during probability instruction may have convinced the strategy students to approach physics differently than they otherwise would – that is, to forgo schema building and to concentrate instead to memorizing, understanding, and becoming fluent in applying physics principles.

Anyway, the principle-learning hypothesis predicts that on all problems, the strategy students should have performed better than the no-strategy students. This is exactly what occurred, which suggests that the main effect of teaching the strategy was to get students to focus on learning the domain principles in both probability and physics.

Unfortunately, because the strategy students outperformed the no-strategy students on every measure, the findings are consistent with any hypothesis that predicts a general increase in ability. For instance, using Pyrenees could have increased the strategy students' self-efficacy, interest in science, or even intelligence.

As a further test, we coded the Andes-physics log files for the solution strategies being used by students. Although the details are presented elsewhere (M. Chi & VanLehn, 2007), we found that no-strategy students tended to use the same mixture of problem-solving strategies on both easy and difficulty problems, and about 55 percent of the mixture was consistent with the Target Variable Strategy. On the other hand, the strategy students tended to use the Target Variable Strategy on hard problems and a step-skipping strategy on easy problems. This suggests that the strategy students had begun to learn the physics principles well enough that they can combine two principle applications in working memory, as physics experts often do (Priest & Lindsay, 1992). However, on difficult problems, they fell back on the scaffolding afforded by the (well-known to them) Target Variable Strategy. This is consistent with the hypothesis that strategy students transferred a predilection to study the principles, which caused them to master the details of each principle more rapidly, which in turn allowed them to start applying principles in working memory instead of on paper, at least on easy problems.

To summarize, we have shown that teaching students an explicit problem-solving strategy improved their performance in the initial domain; more importantly, it seems to have caused AFL of a new domain. Because the improvement occurred with all types of problems, it may be due to the strategy instruction convincing students to focus on learning domain principles, as opposed to problem schemas.

Discussion

These results are unusually good. It is rare to observe interdomain transfer at all, and especially one with such a large effect size. AFL is also uncommon. Lastly, getting students to stop focusing on problems and start focusing on principles is probably rare. Despite all the good news, there are limitations to this study that need to be considered.

First, the initial step in both physics and probability problems is to idealize the given situation and view it in terms of point masses, connectors and contacts between them (physics), or events conditioned on other events (probabilities). In this study, the idealizations were obvious for most problems. When a problem's idealization was not obvious, the problem told students the key idealizations (e.g., what the events are). Problems that "give away" the idealizations are typical of introductory courses on physics and probability, so the idealization phase was tutored by neither Andes nor Pyrenees.

Second, experts in physics can plan solutions mentally without writing principle applications down on paper. This allows them to sort problems by their solution principles (M. T. H. Chi, Feltovich, & Glaser, 1981), articulate a principled basic approach (M. T. H. Chi, Glaser, & Rees, 1982), and determine when two problems' solutions use the same principles (Hardiman, Dufresne, & Mestre, 1989). Even though the strategy group in this experiment exhibited AFL, and that is a manifestation of adaptive expertise, they were not physics experts. Indeed, they were selected to have had no university physics and were given only a few hours' training, so they are not even close to being physics experts. Nonetheless, according to one theory of the development of physics expertise (VanLehn & van de Sande, 2009), they were on their way to becoming experts. It would be interesting to see how adaptive they eventually become.

Third, it may seem that these results are limited to task domains that have both equation-based principles and problem schemas, and that students tend to ignore the principles and focus on acquiring the problem schemas. These properties are key to explaining the AFL results. When the strategy group used Pyrenees to master the probability principles, they also learned that mastering principles was better than learning schemas. Thus, whereas the no-strategy group skimmed the pre-training discussions of physics principles so that they could get to the examples and problems, the strategy group probably paid a great deal of attention to the initial presentation of the physics principles. When no-strategy students got stuck during physics problem solving, they probably hunted for an example that was similar to their problem so that they could map its solution over, whereas the strategy students started using the Target Variable Strategy and their list of principles (available in the help system and the textbook). This manipulation worked, one might say, only because the task domains happened to have both principles and problem schemas.

Hatano and Inagaki (1986) were well aware that adaptive expertise only makes sense when a task domain has both problem schemas (which they called procedures) and conceptual knowledge. Although they sometimes referred to conceptual knowledge as principles, their major examples of conceptual knowledge were mental models, such as the schematic of the phasor bank mentioned earlier. Regardless of the epistemology type of conceptual knowledge, there must exist at least *something extra* beyond procedural skills (schemas) so that the adaptive experts can have it and the routine experts can lack it. Moreover, this extra knowledge must be able to derive or justify the procedures because that allows the adaptive expert to modify and adapt the procedures. In short, our requirement that a task domain have both schemas and principles that derive them is rather close to Hatano and Inagaki's requirement that the task domain have both procedures and a certain kind of conceptual knowledge.

This suggests that the instructional method of Pyrenees could be generalized. The basic idea is to explicate the conceptual knowledge in the declarative (pre-training) materials; then, during problem-solving practice, both compel students to use conceptual knowledge to solve problems and block their use schemas.

Mathematics educators may roll their eyes and exclaim that they have already tried that, and it seldom works. For instance, the principles of the base-10 numeration system plus the principles of set theory can be used to derive the multicolumn subtraction procedure (VanLehn & Brown, 1980). Many instructors and mathematics researchers have tried with little success to get students to understand this conceptual knowledge, often by using Dienes blocks and other concrete base-10 notations. However, the derivation of the written procedure from the principles is extremely long and complex (VanLehn & Brown, 1980). It is well beyond the capabilities of elementary school children. Thus, instructors often end up simply teaching an intermediate procedure (e.g., subtraction using Dienes blocks) that is almost as opaque as the written procedure, even though it is "halfway" between the first principles and the written procedure.

Although the third limitation (that a task domain have both schemas and principles) is not very constraining because many task domains probably have both schemas and principles, we now have a fourth limitation that probably *is* quite constraining. For novices, solving problems with first principles must be feasible and not too much more difficult than solving problems with schemas (procedures). For instance, if novice students must spend twice as much time using principles as schemas to solve problems, then they will probably stop using principles as soon as the tutoring system stops nagging them, and they are not likely to pay much attention to principles when they are taught another task domain. In our task domains, the strategy students, who we assume were using principles, took the same amount of time to solve problems as the no-strategy students, who we assume were using schemas and analogies to past problems. Thus, our task domains meet the criterion that principle-based problem solving is efficient. Multicolumn subtraction does not meet this criterion, so explicitly teaching its principles and blocking its schemas has not worked well.

To summarize, our instructional method is to teach principles explicitly and to require students to solve problems with principles instead of problem schemas. We speculate that it should lead to adaptive expertise when the following conditions are met:

1. The idealization phase of problem solving does not play a major role in this task domain.
2. The students are novices. It is not clear what happens after thousands of hours of training.
3. The task domain has principles or some other conceptual knowledge in addition to procedural knowledge, and the conceptual knowledge can be used to solve problems and modify procedures.
4. For novices, solving problems with conceptual knowledge is only a little more work than solving the problems with schemas.

In collaboration with Jared Freeman and Zachary Horn, these hypotheses are being tested in the task domain of naval force protection tactics.

Acknowledgments

This research was supported by several grants from NSF (EIA-0325054, PSLC-0354420, and DRL-0910221) and an STTR contract from ONR (N00014–09-M-0327).

References

Barnett, S. M., & Koslowski, B. (2002). Adaptive expertise: Effects of type of experience and the level of theoretical understanding it generates. *Thinking and Reasoning, 8(4)*, 237–267.

Bhaskar, R., & Simon, H. A. (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive Science, 1*, 193–215.

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, Vol. 24* (pp. 61–100). Washington, DC: American Educational Research Association.

Catrambone, R. (1994). Improving examples to improve transfer to novel problems. *Memory and Cognition, 22(5)*, 606–615.

Chi, M., & VanLehn, K. (2007). The impact of explicit strategy instruction on problem-solving behaviors across intelligent tutoring systems. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Conference of the Cognitive Science Society* (pp. 167–172). Mahway, NJ: Erlbaum.

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.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33–58.

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

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.

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

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis of findings. *Review of Educational Research, 66*, 99–136.

Haywood, H. C., & Tzuriel, D. (2002). Applications and challenges in dynamic assessment. *Peabody Journal of Education, 77(2)*, 40–63.

Holyoak, K. J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 301–336). New York: Cambridge University Press.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8*, 255–273.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopher, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-Ray Pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*. Mahway, NJ: Erlbaum.

Martin, T., Petrosino, A., Rivale, S., & Diller, K. R. (2006). The development of adaptive expertise in biotransport. *New Directions for Teaching and Learning, 108*, 35–47.

Pandy, M. G., Petrosino, A., Austin, B. A., & Barr, R. E. (2004). Assessing adaptive expertise in undergraduate biomechanics. *Journal of Engineering Education* (July), 1–12.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82(1)*, 33–40.

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

Reif, F., & Allen, S. (1992). Cognition for interpreting scientific concepts: A study of acceleration. *Cognition and Instruction, 9(1)*, 1–44.

Ross, B. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*, 629–639.

Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16(4)*, 475–522.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Charlotte, NC: Information Age Publishing.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22(2)*, 129–184.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction*, *24(2)*, 261–289.

VanderStoep, S. W., & Seifert, C. M. (1993). Learning "how" versus learning "when": Improving transfer of problem-solving principles. *The Journal of the Learning Sciences*, *3(1)*, 93–111.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 526–579). Cambridge, MA: MIT Press.

VanLehn, K. (2006). The behavior of tutoring systems. International Journal of Artificial Intelligence and Education, 16, 227–265.

VanLehn, K., Bhembe, D., Chi, M., Lynch, C., Schulze, K., Shelby, R., et al. (2004). Implicit vs. explicit learning of strategies in a non-procedural skill. In J. C. Lester, R. M. Vicari, & F. Paraguaca (Eds.), *Intelligent tutoring systems: 7th International Conference* (pp. 521–530). Berlin: Spring-Verlag.

VanLehn, K., & Brown, J. S. (1980). Planning nets: A representation for formalizing analogies and semantic models of procedural skills. In R. E. Snow, P. A. Frederico, & W. E. Montague (Eds.), *Aptitude learning and instruction: Cognitive process analyses* (pp. 95–138). Hillsdale, NJ: Lawrence Erlbaum Associates.

VanLehn, K., & van de Sande, B. (2009). Acquiring conceptual expertise from modeling: The case of elementary physics. In K. A. Ericsson (Ed.), *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 356–378). Cambridge: Cambridge University Press.