DOCUMENT RESUME

ED 296 291 CS 009 198

AUTHOR	Chi, Michelene T. H.; And Others
TITLE	Self-Explanations: How Students Study and Use Examples in Learning To Solve Problems. Technical Report No. 9.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Psychological Sciences Div.
PUB DATE	Nov 87
CONTRACT	N0001-84-K-0542
NOTE	61p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Artificial Intelligence; *Cognitive Style; College Students; Higher Education; *Learning Strategies; Physics; *Problem Solving; *Protocol Analysis
IDENTIFIERS	Students as Subjects

ABSTRACT

A study examined in detail the initial encoding of worked-out examples of mechanics problems by "good" and "poor" students, and their subsequent reliance on examples during problem solving. The subjects, three males and five females, were selected from responses to a university campus advertisement. Six of them were working towards bachelor's degrees with varying majors. Two of the eight students had additional post-graduate training in psychology and none had a college physics course, although seven of the eight had taken physics in high school. Students with a range of grade-point averages were chosen so that learner differences could be examined. Subjects were given instruction in physics and required to demonstrate mastery of basic physical principles. Talk-aloud protocols were employed to examine the way the subjects learned and understood the examples, and then used their understanding to solve problems. The subjects were split into two groups post hoc using a median split on their problem solving success. Results indicated that "good" students tend to study examples and exercises by explaining and providing justifications for each action and relate their explanations to the principles and concepts in the text. "Poor" students do not often explain the examples or exercises to themselves, and when they do, their explanations do not seem to connect their understanding with the principles and concepts in the text. The results provide, at a gross level, empirical evidence to support existing artificial intelligence models of explanation-based generalizations. (Seven figures and seven tables of data are included and 27 references are appended.) (RS)

University of Pittsburgh

LEARNING RESEARCH AND DEVELOPMENT CENTER

Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems

Michelene T. H. Chi, Miriam Bassok

University of Pittsburgh

Matthew W. Lewis

Carnegie-Mellon University

Peter Reimann

University of Freiburg

Robert Glaser

University of Pittsburgh

November 1987

Technical Report No. 9

This research is funded by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-84-K-0542, to the first and the last authors. We are grateful for consultation provided by Frank Boyle, and Ted Rees on physics, and insightful comments on the manuscript by Kurt VanLehn.

Abstract

The present paper analyzes in detail (talk-aloud protocols) "Good" and "Poor" students' initial encoding of worked-out examples of mechanics problems, and their subsequent reliance on examples during problem solving. We find that "Good" students learn with understanding: they generate many explanations which refine and expand the conditions for the action parts of the example solutions, and relate these actions to principles in the text. These self-explanations are guided by accurate monitoring of their comprehension failures and successes. Such learning results in an example-independent knowledge and in a better understanding of the principles presented in the text. "Poor" students do not generate sufficient self-explanations, inaccurately monitor their learning and subsequently rely heavily on examples. The results are discussed relating these psychological findings to existing AI models of explanation-based generalizations.

SELF-EXPLANATIONS: HOW STUDENTS STUDY AND USE EXAMPLES IN LEARNING TO SOLVE PROBLEMS

Learning is a constructive process in which the student converts the words and examples from the teacher or the text, into usable skills, such as solving problems. This process of conversion is essentially a form of constructive self-instruction (Simon, 1979). Although the research on the quality of good teaching (such as those which attempt to identify the characteristics of a good Socratic tutor, Collins & Stevens, 1982), as well as research on the quality of a good text (such as those that manipulate the quality of elaborations, Reder & Anderson, 1980) may be informative, ultimately, learning rests on the learning skills that the students themselves bring to bear as they learn. The goal of this research is to understand the students' contribution to learning. To accomplish this goal, our approach is to study individual differences as the students learn.

The majority of problem solving research in the literature assumes that the declarative knowledge is first encoded from the text or from the teacher's words, then proceduralized into a skill. Most of the literature has concentrated on the conversion of already encoded knowledge into smooth, fast, skillful problem solving. This conversion process dominates, for example, Anderson's theory of skill acquisition (Anderson, 1987). In that theory, the process of conversion is achieved by using general weak methods which can convert declarative knowledge into domain specific procedures via the mechanism of compilation. Thus, in Anderson's theory, it is assumed that the effortful process lies in the conversion of the declarative knowledge into the procedural knowledge, whereas the encoding of the declarative knowledge is taken to be a straightforward storing:

In unanalyzed form our experiences in any domain, including instruction (if it is available), models of correct behavior [worked-out examples], successes and failures of our attempts, and so on... This means that we can easily get relevant knowledge into our system.... (p. 206).

We concur with the common assumption that learning a skill can be viewed as encoding of instruction followed by proceduralization of some kind. However, our research focuses on the encoding of instruction, because a leading conjecture is that individual difference is due to the differences in the encoded representation rather than differences in conversion of the encoded instructions to skill. Anderson (1987) also alluded to this difference in the encoded representation as a source of individual differences when he suggested that: "weak problem-solving methods like analogy can be much more effective if they operate on a rich representation of the knowledge..." (p. 206).

Our work focused on the knowledge that students acquire as they read from a text, prior to their attempts to apply that knowledge in the form of a procedural skill such as problem solving. We predict that learner differences arise primarily from the representation that the student has acquired. Thus, we believe that individual differences do not arise solely from the experiences the learner is exposed to in the form of instruction or text. Rather, we conjecture that learner differences may arise from differences in the ways students understand and learn from text, and moreover, these learner differences subsequently effect the use of a particular skill (solving problems).

Our interest in learner differences derives in part from an attempt to understand the acquisition of expertise. While it takes long periods of study and practice before one can become an expert in any domain. It is clear that extensive practice and study is a necessary but not sufficient criterion for becoming an expert. Thus, examining how students differ in the way they learn new materials may shed light on processes of skill acquisition that determine whether an individual will or will not achieve expertise.

The part of the text that we focus on is the worked-out examples. This choice was based on theoretical, empirical, as well as instructional reasons. Theoretically, there is some controversy in the literature concerning how generalizations are induced from examples (see Dietterich & Michalski, 1983, for a review of the AI literature, and Murphy & Medin, 1985, for a review of the psychological literature.) There are two views. A similarity-based approach claims that generalizations are developed by inducing a principle (or a set of common features) from multiple examples. Such a principle would embody the essential features shared by all the examples. On the other hand, an explanation-based approach to problem solving (Lewis, 1986; Mitchell, Keller & Kedar-Cabelli, 1986) claims that generalization can be obtained from a single or a few examples. Although many theories, including Anderson's ACT*, can only generalize from many examples, it is clear that students can often generalize from a single example (Elio & Anderson, 1983; Kieras & Bovair, 1986). Providing evidence on how generalizations can emerge from learning a few examples would give credence to an explanation-based approach to learning.

Empirically and instructionally, there is a dilemma in the literature as well. More and more empirical evidence is emerging showing the importance of examples in learning. Reder, Charney and Morgan (1986) for example, found that the most effective manuals for instructing students how to use a personal computer are those which contain examples. LeFevre and Dixon (1986) found that students actually prefer to use the example information and ignored the written instruction when learning a procedural task. Pirolli and Anderson (1985) also found that 18 of their 19 novices rely on analogies to examples in the early stages of learning to program recursion. Besides these laboratory findings. VanLehn (1986) also provided indirect evidence that examples are important in regular classroom learning. He found that 85% of the systematic errors in arithmetic, collected from several thousand students could be explained as deriving from some type of example-driven learning process.

On the other hand, although both students and instructional materials rely heavily on worked-out examples as an instrument for learning, the empirical work which directly examined the role of example solutions on problem solving found that students who have studied examples often cannot solve problems that require a very slight deviation from the example solution (Eylon & Helfman, 1982; Reed, Dempster, & Ettinger, 1985; Sweller & Cooper, 1985). The discrepancy between students' better performance from text materials that contain examples and the students' failure to generalize from examples, may be caused by the degree to which students understand the examples provided. Generally, in the empirical studies cited, no assessment is made about how well students understood the examples. As Pirolli and Anderson (1985) noted, although most of the students wrote new programs by analogy to example programs, the success depended on how well the students understood why the examples worked. Our work, in the domain of mechanics, will shed light on the degree to which students understand an example in relation to their ability to solve problems.

The method we use to study how students learn and understand an example is via the explanations they give while studying it. Such self-explanations are important, we propose, largely because examples typically contain a sequence of actions, without much explication of the rationale underlying the sequence of actions. Simon (1979), for example, noted that:

> Generally speaking, textbooks are much more explicit in enunciating the laws of mathematics or of nature than in saying anything about when these laws may be useful in solving problems. The actions of the productions needed to solve problems in the domain of the textbooks are laid out systematically, but they are not securely connected with the conditions that should evoke them (p. 92).

We can provide a concrete example of the inadequacy of worked-out examples by taking one example from the fifth chapter of the Halliday and Resnick text (1981), as shown in Figure 1. We can see the lack of specification of the explicit conditions under which the actions should be executed.

For example, it is not clear in Statement 2 of this example-exercise why one should "consider the knot at the junction of the three strings to be the body." This is a critical piece of information because it implies that at this location (as opposed to the block), the sum of the forces is zero. Such lack of specification of the explicit conditions for actions occurs throughout the example. In Statement 6, how does the student know that $\vec{F}_A$, $\vec{F}_B$, and $\vec{F}_C$ are all the forces acting on the body, and that there are no others? Statement 7 is essentially a restatement of Newton's First Law,[1] but it requires chaining several inferences together, and translating them into an equation (for example, because the

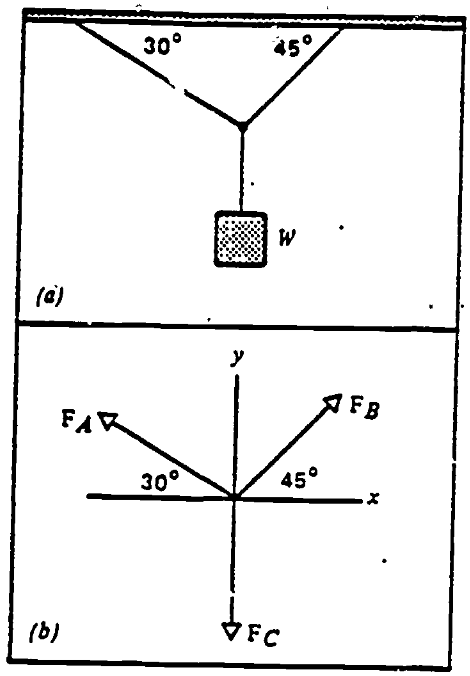


Figure 5-6 Example 5. *(a)* A block of weight *W* is suspended by strings. *(b)* A free-body diagram showing all the forces acting on the knot. The strings are assumed to be weightless.

1. Figure 5-6a shows an object of weight W hung by massless strings.
2. Consider the knot at the junction of the three strings to be "the body".
3. The body remains at rest under the action of the three forces shown Fig. 5-6.
4. Suppose we are given the magnitude of one of these forces.
5. How can we find the magnitude of the other forces?
6. F_A, F_B, and F_C are *all* the forces acting on the body.
7. Since the body is unaccelerated, $F_A + F_B + F_C = 0$
8. Choosing the x- and y-axes as shown, we can write this vector equation as three scalar equations:
9. $F_{Ax} + F_{Bx} = 0,$
10. $F_{Ay} + F_{By} + F_{Cy} = 0$
11. using Eq. 5-2. The third scalar equation for the z-axis is simply:
12. $F_{Az} = F_{Bz} = F_{Cz} = 0.$
13. That is, the vectors all lie in the x-y plane so that they have no z components.
14. From the figure we see that
15. $F_{Ax} = -F_A \cos 30^\circ = -0.866F_A,$
16. $F_{Ay} = F_A \sin 30^\circ = 0.500F_A,$
17. and
18. $F_{Bx} = F_B \cos 45^\circ = 0.707F_B,$
19. $F_{By} = F_B \sin 45^\circ = 0.707\, F_B.$

Figure 1: A strings example, taken directly from Halliday and Resnick (1981)

body is at rest, there are no external forces, therefore the sum of the forces on the body must equal to zero). Statement 8 is totally unexplained; why are the axes chosen as such? It is clear that the solution steps within an example are not explicit about the conditions under which the actions apply.

In order to learn with understanding, a student needs to overcome the incompleteness of an example by drawing conclusions and making inferences from the presented information (Wickelgren, 1974). To do so, a student needs to provide explanations for why a particular action is taken. Only then will the student be able to apply the acquired procedure to non-isomorphic problems that do not match exactly the conditions of the example solution. Thus, we suggest that a good student "understands" an example solution and will succeed in generalizing because he/she makes a conscious effort to explain and extrapolate the principle, and to ascertain the conditions of application of the solution steps beyond what is explicitly stated. Consequently, we ask in this paper how students can come to understand an example.

It is difficult to define what "understanding" means in the context of learning from examples. Operationally, our study was designed to assess understanding with three measures: solution of isomorphic problems, solution of far transfer problems, and elaborations generated during studying examples. The weakest method to detect understanding is to observe how successfully students solve very similar problems, since very similar problems can often be solved by a simple syntactic mapping of the example procedure to the to-be-solved problem. Another method is seeing if students can successfully use the principle involved in the example in a different and complex problem, (far transfer), because such problems prevent students from being able to solve them via a syntactic mapping. However, to a certain extent, studying far transfer reveals only that understanding exists and allows one to see what conditions facilitate it.

The method which permits the most direct assessment of understanding of an example is to examine the explicit elaborations that students provide while studying it. Elaborating is a mechanism of study that allows students to explain, infer, and explicate the conditions and consequences of each procedural step in the example. We postulate that elaborations can reveal students' understanding by showing whether or not they know (a) the conditions of application of the actions, (b) the consequences of actions, (c) the relationship of actions to goals, and (d) the relationship of goals and actions to natural laws and to other principles. This paper focuses primarily on the elaboration results.

In the actual procedure used, we invited students to explain to themselves what they understand after every line of a worked-out example. Often they made no comments at all. However, when they did generate a comment, this method allowed us to yoke the explanation to the statement line in the example so that we could interpret their protocols more easily.

Method

Subjects

Eight students (3 males and 5 females) were selected from responses to a campus advertisement. Six of them were working towards bachelor's degrees with varying majors. Two of the 8 students had additional post-graduate training in psychology. None of the students had a college physics course, although all of them (except one) had taken high school physics, with differential performance (reported grade) in that course. We intentionally chose students with a range of abilities in terms of grade-point average so that we could examine learner differences. Students were paid for their participation.

Procedure and Materials

In the present paper we focus on how students study three worked-out examples of problems dealing with the application of Newton's laws of motion and on how this initial learning relates to their subsequent problem solving. However, since the relevant chapter demands substantive background subject matter, the study of examples and the problem solving tasks are embedded within a longitudinal study in which students in our laboratory studied Newtonian Mechanics. Students learned the new material in a way they would normally do when studying on their own, in terms of the amount of time devoted to studying, the rate of self-pacing, and the use of studying habits, such as highlighting significant parts of the text or rereading. The laboratory learning differed from the way students would typically learn on their own in that all the learning took place in the laboratory, and that students gave talk-aloud protocols while studying examples and solving problems. The students spent between 8 to 29 hours to complete the study, spread over several weeks. Figure 2 presents a diagram summarizing the experimental procedure for the whole study. Basically, the study consisted of 2 major phases: knowledge acquisition and problem solving.

Knowledge Acquisition

During the first part of the knowledge-acquisition phase, subjects studied the necessary background subject-matter, covering the topics of measurement, vectors, and motion in one dimension. These materials are covered in the first three chapters of Halliday & Resnick. The fourth chapter on Motion in a Plane was eliminated because it did not have direct bearing on learning Chapter 5, the target chapter on Particle

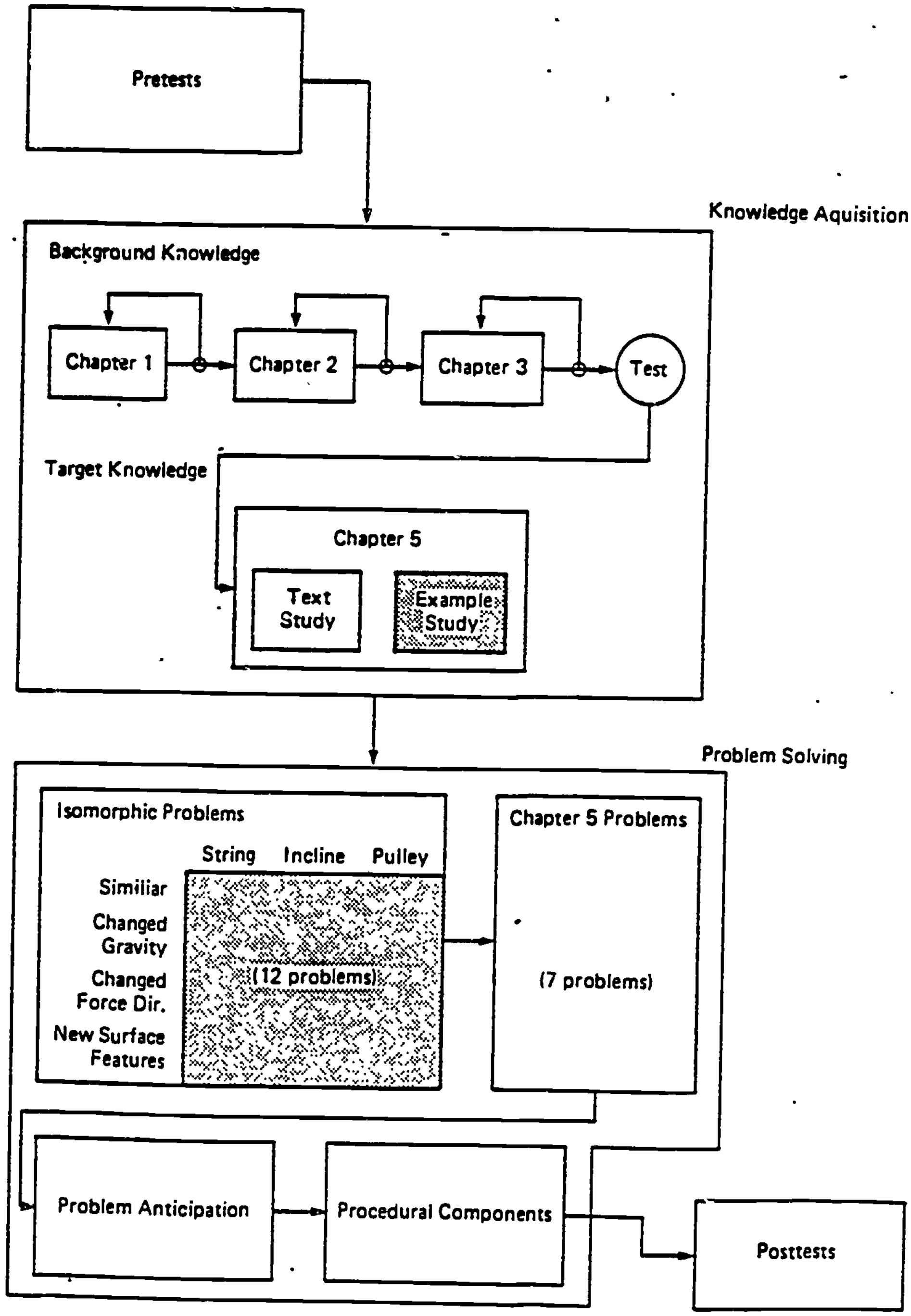


Figure 2: A diagram depicting the design of the study

Dynamics. Decisions of this kind, as well as designing of questions and problems, were made after consultations with physicists.

For each of the three background chapters, students read through and studied the chapter. They were requested to record on a separate sheet of paper any questions which arose during their study. Each chapter also contained a note to stop and get the experimenter in order to give a verbal protocol of one worked-out example. This protocol was meant mainly as practice in giving protocols. When the students believed they were ready to be tested on the material in the chapter, they notified the experimenter, and were then required to produce correct answers to a set of Declarative, Qualitative and Quantitative questions, given in that order. Declarative questions were designed to assess the recall of critical facts from each chapter, e.g., "What is the difference between a scalar and a vector?" Some of these questions came from the textbook and some were designed by the experimenters. The Qualitative questions were designed to assess reasoning and inferences about the concepts in each of the chapters without reference to quantities, e.g., "Can an object have an eastward velocity while experiencing a westward acceleration?" The majority of these problems were taken from the textbook, several were generated by the experimenters. The Quantitative questions assessed procedural skills for quantitative problem solving, e.g., "Two bodies begin free fall from rest at the same height, 1.0 sec apart. How long after the first body begins to fall will the two bodies be 10 meters apart?" All of the Quantitative problems came from those presented in the text at the end of the chapters.

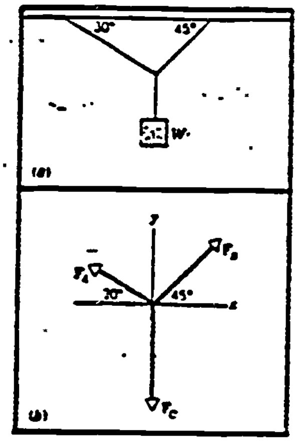
After solving each problem set, students returned the problems to the experimenter for grading. If no errors were made, they proceeded to the next set of questions. If an error was made, the questions were returned for correction. The incorrect answers were identified and students were given references to those sections of the text which addressed these questions. For incorrect Quantitative questions the correct final answer was also provided, analogous to the common practice of looking up the answer in the back of the book. Students then attempted to correct the answers and resubmitted them for grading. If there were still errors on the second try, students were provided with a worked-out solution for the incorrect answers. Students then studied the worked out solution, and after it had been removed, were asked to change their answers and/or to reproduce the correct solution from memory. If after these two tries the answer was still incorrect, the experimenter explained the worked-out solution and the student had to reproduce it. Thus, the first three chapters, consisting of the background subject matter, were studied until students could correctly solve a set of Declarative, Qualitative and Quantitative problems relevant to each chapter.

During the second part of the knowledge acquisition phase, students studied the target chapter on particle dynamics (Chapter 5 of Halliday & Resnick). Studying the target chapter proceeded exactly as for the other chapters, except that students were required to answer correctly the relevant Declarative and Qualitative questions before studying the three worked-out examples. This was done to assure that the students had acquired the relevant declarative knowledge from the text.

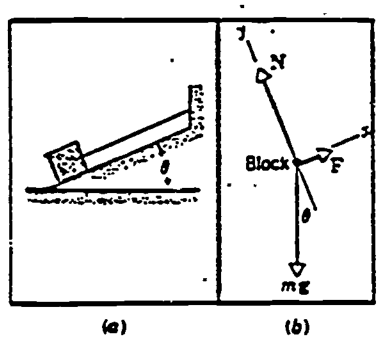
The study of examples was the major focus of the knowledge acquisition phase; students studied three worked-out examples, taken directly from the text. (See Figure 1 for an example of one of the worked-out solutions.) Each example solution represented a "type" of problem. There were three types: a strings, an inclined plane, and a pulley problem. (See Figure 3.)

Students talked out-loud while studying the examples, and their protocols were taped.

Strings



Incline



Pulley

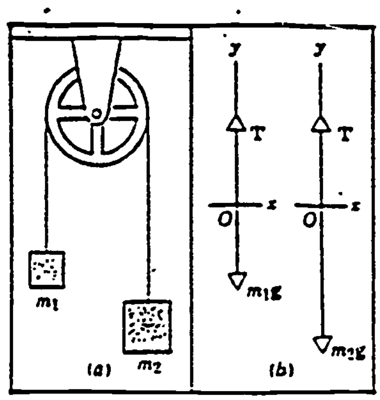


Figure 3: Diagram depictions of the three examples studied

Problem solving

During the problem solving phase, students solved two main sets of problems. The first set consisted of 12 "isomorphic problems", with four problems corresponding to each "type" of example (see the shaded areas in the problem solving phase of Figure 2). These problems were specially designed by us so that they varied in their degrees of similarity to the worked-out examples studied in the previous phase. Figure 4 shows one set of isomorphs corresponding to the "strings" example.

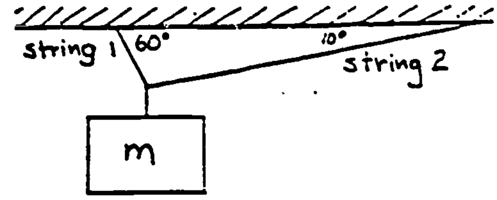
The second set consisted of seven "chapter" problems, which were problems taken directly from the end of the target chapter. In terms of our criteria for isomorphism to example problems, these problems can be considered "far transfer" problems. Students solved both sets of problems while giving talk-aloud protocols, and no feedback was provided. Only the problem solving protocols of the first set of three isomorph problems will be analyzed in detail for this paper.

Results

As indicated, this paper will report on the results of the elaboration protocols while students study the example solutions, presented in the target chapter, as well as how they use the examples in the problem solving protocols of the isomorphic problems. Individual differences will be reported by contrasting the performance of "good" and "poor" students. These two groups were defined post hoc, using a median split on the problem solving successes of our students on the 12 isomorphic and the 7 chapter problems. In scoring the problems, arithmetic slips were not considered as errors and partial solutions were credited proportionally. All problems received the same weight, so that the maximum score for the isomorphic problems was 12 and the maximum score for the chapter problems was 7. (All scoring, identification and classification of protocols

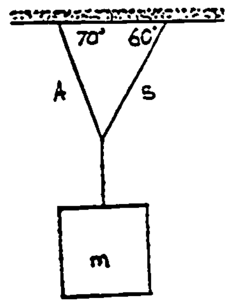
A block is hanging from three strings. If the tension in string 1 is 18 N, what is the mass of the block?

Similar



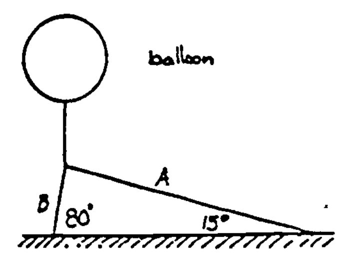
The block pictured is hanging from 3 strings. The mass of the block is 10 kg. If the acceleration due to gravity was reduced to 1/2 of its normal value, what would the tension in rope A be?

Changed Gravity



A balloon is being held down by three massless ropes. If the balloon is pulling up with a force of 300 N, what would the tension in rope A be?

Changed Force Direction



Three forces are holding a 800 kg block motionless on a frictionless surface. If force A is 50 N, what would force C be?

New Surface

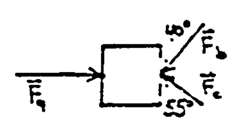


Figure 4: The set of four isomorphs to the strings example

were performed by two judges, with a mean interrater reliability across all the analysis ranging between 86% and 94%.)

There were four students in each group. The mean success of the Good students was 82% (96% for the isomorphic problems and 68% for the chapter problems). The mean success of the Poor students was 46% (62% for the isomorphic and 30% for the chapter problems).

Learning from Examples: Elaboration protocols

Amount of Elaborations Generated: Unit of Analyses.

Our first analysis began with a very gross count of the amount of elaborations students provided while studying the example solutions. An elaboration is a generic term we use to indicate any statement made that is not a first reading of an example line, nor is it any conversation carried on with the experimenter that does not refer to the subject matter of the example, such as "Do you have a calculator that I can use?", nor is it a response to the experimenter's requests to speak louder. Figure 5 is a transcript of one student's elaboration protocol. During the transcription, segmentation was made roughly at pause boundaries. We also assigned a line number to each line of the protocols. Counting each line that is not a reading or experimenter's comment as a line, this example shows a total of 13 lines. Notice that because we are not using sentence boundary as a line boundary necessarily, a short comment such as "Okay" is counted as one line, since it is inserted between different activities, such as reading.

Table 1 shows the average number of lines generated per example by the Good and Poor students, as well as the average amount of time it took them to study an example. These differences are contrasted with the average number of protocol lines generated and the amount of time taken to solve an isomorphic problem. The Good students generated a considerably greater number of elaboration lines than the Poor students (142 lines vs. 21 lines, $t(6) = 1.97$, $p < 0.05$). These elaboration activities necessarily led the Good students to spend more time on each example as well (13 min. vs. 7.4 min., $t(6) = 2.16$, $p < .05$).

	Read Line 6:		*Fa, Fb, and Fc are all the forces acting on the body.* (pause)
(I)	Elaboration:	Monitoring	1) Okay.
	Read Line 7.		*Since the ...* (pause)
	Experimenter:		(Okay, what are you thinking about?)
(II)	Elaboration:	Monitoring	2) I'm trying to think where Forces Fb and Fa 3) are going to get the thing.
(III)	Elaboration:	Explanation	4) They'd just be the force, the rest mass of 5) the thing holding it up would be the force. 6) It could, well, actually it'd be the force of weight.
(IV)	Elaboration:	Explanation	7) Cause being upheld by...it's the resistance 8) to weight W.
(V)	Elaboration:	Explanation	9) It would all be equal. 10) The change in the uh, negative Y will the...
(VI)	Elaboration:	Explanation	11) The force of negative Y will be equal to 12) the force in the positive Y.
(VII)	Elaboration:	Explanation	13) And they'll all equal out.

Figure 5: An example of a transcript of example-studying protocols (S101)

We rule out the possibility that Good students are simply more articulate or fluent, thus producing more elaborations, since, in contrast, the number of lines they generate while solving problems is approximately the same as the Poor students (141 vs. 122 lines per problem, see Table 1, bottom). This suggests that the students generate as many lines of protocols as they deem necessary in order to learn the example. Likewise, the

Table 1

Amount of Elaboration Generated While Studying Examples and Solving Problems

		GOOD	POOR
Example-Studying Protocols	No. of Lines	142	21
	No. of Min.	13	7
	No. of Statements or Single Ideas	51	18
Isomorphic Problem Solving Protocols	No. of Lines	141	122
	No. of Min.	13.8	14.3

Note. The numbers represent averages per student per example.

fact that the Good students took more time studying each example solution than the Poor students (13 min. vs. 7.4 min.) also reflects their choice to spend more time on each example, rather than a tendency to dwell unnecessarily on the examples, since the amount of time they spent solving each isomorphic problem is about the same as the time spent by the Poor students (13.8 min. vs. 14.3 min.).

The assumption we make, that elaborations are generated if some knowledge or inference is being processed or constructed in memory, is no different than assumptions made with reference to other types of dependent measures. For example, the assumption underlying protocol analyses is that the problem solver is saying what he is thinking about or dumping the content of his/her working memory. Thus, longer protocols simply refer to a greater degree of processing. Similarly, in research using eye movement protocols, the assumption is that the student is processing the location at which he/she is fixating (Just & Carpenter, 1976). Thus, longer fixations imply that the student is spending more time processing that location. Therefore, we view the greater amount of elaboration produced by the Good students as a natural consequence of wanting to understand the solution example better, rather than the possibility that they are more articulate and fluent.

It might be claimed that the reason that Good students learned more is just that they spent more time studying. However, this is not a very deep explanation, but merely a restatement of the correlation. The important question is what the Good students did while they were studying. As will be seen shortly, there are such large qualitative differences in what students did while studying examples that the simple, shallow, explanation--that studying twice as long makes one learn twice as much--is quite implausible because it cannot explain the other differences in students' learning behavior.

However, although we claim that longer protocols do not necessarily imply verbosity, using lines as a unit of analysis does not seem adequate for capturing what is going on. That is, an analysis at the level of lines is too fine-grained to characterize the nature of these comments. Instead, many lines seem to refer to a single idea, and thus we further collapsed the lines into units expressing a single idea. For example, the 13 protocol lines in Fig. 5 have been recoded into 7 elaboration "statements" (I-VII). Thus, basically the boundaries were placed on the basis of different ideas, pauses, or different sort of activities including reading or experimenter's interjections.

Using this sort of parsing, we see in Table 1 that Good students' protocols have been reduced from 142 lines to 51 statements and Poor students' protocols remain roughly the same (that is, 18 statements vs. 21 lines). Thus, in a sense we are penalizing the Good students for being articulate. What the data actually shows is that the Good students often elaborate quite a bit on a single idea either because they realize that they do not understand it or they want to provide a more complete explanation (see next sections), or the ideas are just harder to explain succinctly. Lewis and Mack (1982) have also noticed that learners often spontaneously offer explanations of why things happened the way they did. Even with this unit of analysis, the Good students elaborate significantly more frequently than the Poor students (51 vs. 18, $t(6) = 1.98$, $p < .05$). Henceforth, all analyses will be based on the number of single ideas or statements.

Kinds of Elaborations Generated.

Elaboration statements can be classified into three types. (See Fig. 6 for a breakdown of all the classifications to be discussed.)

An elaboration is considered to be an explanation if it says anything substantive about the physics discussed in the example statement. The following comment would be considered an explanation:

> *"Ummm, this would make sense, because since they're connected by a string that doesn't stretch."*

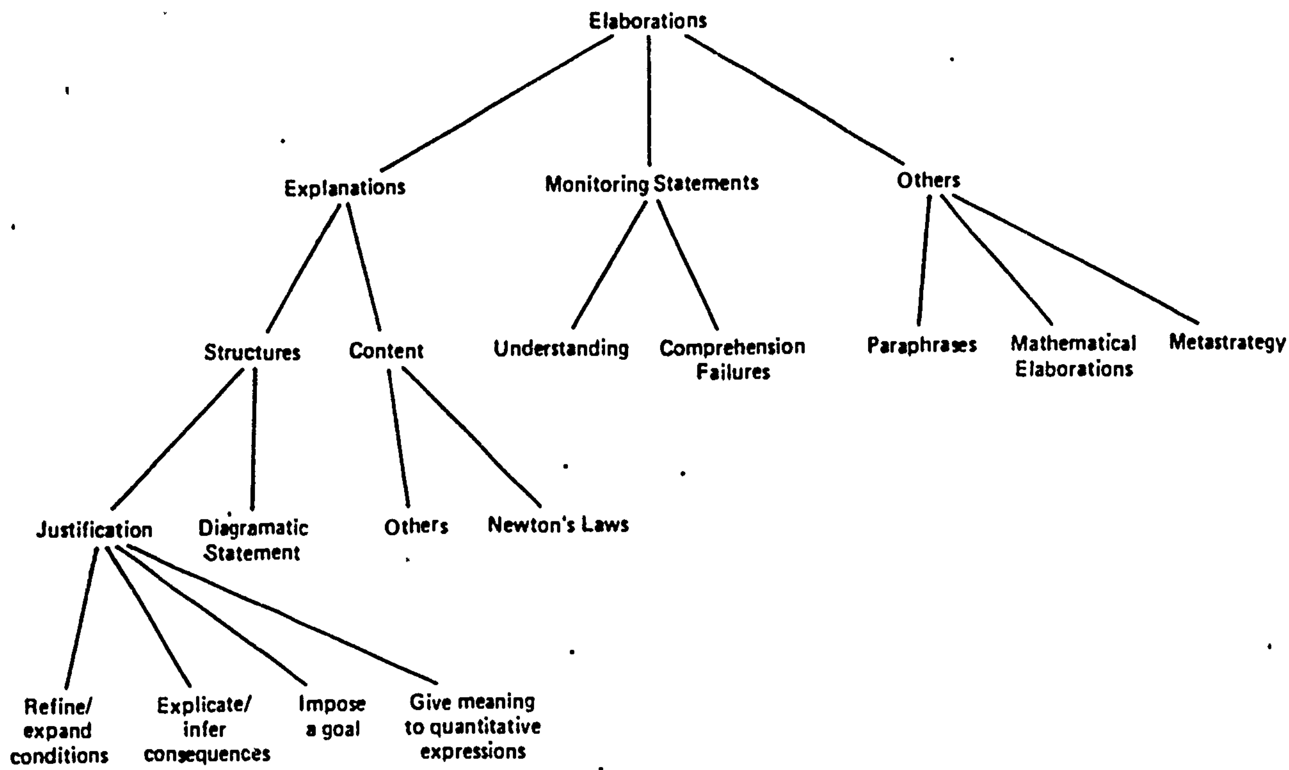


Figure 6: A classification scheme illustrating he decomposition of the analyses of the protocols.

or

"If the string's going to be stretched, the earth's going to be moved, and the surface of the incline is going to be depressed."

Other examples can be seen in Figure 5.

An elaboration is considered a <u>monitoring statement</u> if it refers to states of comprehension. For example, remarks such as:

"I can see now how they did it"

or

"I was having trouble with F-mgsinθ = 0"

would be considered monitoring statements.

A third category includes several <u>other</u> types of statements, mostly paraphrasing, mathematical elaborations, and metastrategic statements. Paraphrasing would be comments that restate what the example line said. For example, a student who read the example line

Notice that the magnitude of T is always intermediate between the weight of m_1 and the weight of m_2

remarked that

"Tension is always intermediate between the two...".

Such a statement would be coded as a paraphrase. Another example would be the remark

"There are no more forces"

after reading the example statement

Fa, Fb, and Fc are all the forces acting on the body.

A mathematical elaboration would be a statement such as

"If this is the sine of 45 degrees, then this is the cosine of 45 degrees".

And a metastrategic statement concerns what the students are doing or planning to do, such as:

"I'm going to look at the diagram"
"I'm going to reread it"
"This is something to remember".

Table 2 shows the distribution of the three categories of statements for the Good and Poor students.

Looking at Table 2, one can see that although the Good students produced a significantly greater total number of elaborations (50.9) than the Poor students (17.5), they do not differ in the distribution of the proportion of each type of elaboration they provide. This similarity in the distribution suggests that the difference between the Good and Poor students may just be a quantitative one, so that a simple remedial treatment for the Poor students is simply to ask them to spend more time and effort at producing elaborations. However, closer examination of each of the categories suggests that the difference is not a straightforward quantitative one. Below, we analyze the differences in the structure and content of the Explanations, as well as differences in the Monitoring Statements.

Analyses of Explanations.

Of all the elaboration statements produced by the Good students, 31% are explanations relating to the physics content as compared to 24% for the Poor students. Even though both groups produced proportionately a similar amount of explanations, the

Table 2

Types of Elaborations

	GOOD		POOR	
	Prop.	Freq.*	Prop.	Freq.*
Physics Explanation	31%	15.9	24%	4.3
Monitoring Statements	39%	20.1	42%	7.3
Others (Include Paraphrase)	30%	14.9	34%	5.9
Total	100%	50.9	100%	17.5

*An average number of statements per student per example.

absolute number of these explanations is significantly different between the two groups (15.9 per example vs. 4.3, $F(1,6) = 7.25$, $p < .05$). Also, the correlation between the number of physics explanations and the subsequent success in solving the isomorphic problems is very high ($r = .76$, $p < .05$). Thus it seems that these explanations play an important role in learning from examples.

Structure of the Explanations. One way to analyze the explanations students provide upon studying example statements is to analyze the structure of their explanations, by which we mean the form or purpose of what they say. Since our conjecture in this research was that students often fail to understand example solutions because the example provides neither a clear specification for why each procedural step is taken nor an explication of the consequences of each step, this would imply that if students do understand an example solution well, they themselves must explicate the conditions and consequences of each solution step. Hence, we attempted to capture this characterization in the choice of the analysis categories. Analyzing the physics explanation statements, we could classify them into two major categories: justifications and diagrammatic statements (see Fig. 6 again). Diagrammatic statements are statements that basically describe in words what was depicted in the diagram. For example, a statement made by a student in reference to the diagram depicted in Figure 1 was:

> *"Okay, so three forces are on the two strings and from the string going down to the object."*

Justifications on the other hand, are explanations that either:

1) <u>Refine or expand the conditions of an action.</u> For example in response to Line 18 of the inclined plane example, which stated that

It is convenient to choose the x-axis of our reference frame to be along the incline and the y-axis to be normal to the incline

student S110 explained the conditions of such a choice by saying:

"and it is very, umm, wise to choose a reference frame that's parallel to the incline, parallel and normal to the incline, because that way, you'll have to split up mg, the other forces are already, component vectors for you."

2) Explicate or infer additional implications of an action. In response to the same line in the inclined plane example, student SP2 would say:

"So we can save 1 force...we save the F force. We save it so, that I don't have to calculate it by, with angles... So we can save two forces. Partitioning of two forces."

3) Impose a goal or purpose for an action. In response to the same Line 18, again, S110 said:

"Basically it looks like they are going to split up these three forces into their respective components."

4) Give meaning to a set of quantitative expressions. In response to Lines 13-16 of Fig. 1, which simply states two equations, student S110 stated:

"Umm, and looking along, they've done the **same thing** *for FsubB. They've separated it into its components, into its component vectors, and they are basically able to figure out umm, the tension in each of the strings and how much W weights providing that..."*

Another student (SP1) responded similarly by saying:

"Now they are going to do the **same thing** *with it to the Y".*

These latter two quotes show that not only can students understand in a sense the meaning and purpose of the quantitative expressions (which is a procedure for decomposing the forces), but they are also basically providing a goal for the set of actions. They refer to the goal of the equations as the "same thing".

Table 3 shows the distribution of the two kinds of Explanation statements. As can be seen, almost all of the Good students' explanations (98% or 15.5 statements) are of the justification kind, whereas a fairly high proportion (35%) of the Poor students' explanations are diagrammatic statements. (The interaction is significant at the .01 level, $F(1,6) = 20.55$.) The characteristic of the diagrammatic statements is that they do not necessarily add any new information. In some sense, they are paraphrasing into words what is shown pictorially, whereas the justifications are explanations of various sorts as discussed above. Diagrammatic statements were classified as "explanations" rather than included in the "other" category since they were not paraphrases of textual statements.

Table 3

Kinds of Explanation

	GOOD		POOR	
	Prop.	Freq.*	Prop.	Freq.*
Justifications	98%	15.5	65%	2.75
Diagrammatic Statements	2%	0.3	35%	1.5
Total	100%	15.8	100%	4.25

*An average number of statements per student per example.

Content of the Explanations. So far, we have just given a characterization of the structure of the explanation statements. A majority of the explanations justify actions stated in the text: they refine or expand the conditions of an action, explicate the consequences of an action, provide a goal for a set of actions, and explain the meaning of a set of quantitative expressions. However, there is an important difference in the Good and Poor students' explanations that is not easily captured by assessing the structure of what they say. This difference requires an analysis of the content of the explanations.

We can first give a descriptive analysis of the difference between the Good and Poor students' justifications by contrasting the following explanations generated after reading the equation from the Pulley example:

$$T - m_1 g = m_1 a$$

A poor student (S109) comments:

"OK, cause the acceleration is due to gravity."

This comment, at best, is an incomplete statement. It does not capture the interrelation between the force of tension and the force due to gravity. On the other hand, a Good student (SP1), in response to the same example line remarks:

"OK, so its basically a way of adding them together and seeing if there is anything left over. And if there is anything left over, it equals the force: mass times acceleration."

Such a comment is not only more complete, but it shows that the Good student is trying to understand the example by relating the example statements to explanations and principles stated in the text, and understanding how one example statement follows from the previous one(s). This is typical of Good students' explanations.

Our descriptive analyses of the nature and content of the explanations so far indicate that the Good students seem to generate explanations which relate to the principles stated in the text, as well as relating the consecutive example statements to each other (as when they impose a goal on a set of actions, or give meaning to a set of quantitative expressions). We can further capture the degree to which students' explanations reflect the principles learned from the text by judging the extent to which explanations can be said to be guided by these principles. For example, the following response to Line 3 of Figure 1:

> *"So that means that they have to cancel out, only the body wouldn't be at rest."*

would be judged to be guided by Newton's First Law that if there is no motion, then the sum of the forces must be zero. Thus, to increase the rigor of our analyses, we further restricted our analyses to only those explanations that are derived or inferred from Newton's Three Laws. About a third of the Good students' explanations (4.75 out of 15.8 statements) could be derived from or referred to Newton's Laws, whereas nearly none of the Poor students' explanations (0.7 out of 4.25 statements) related to any of Newton's Laws.

One might think that more of the Good students' explanations are derivable from the principles in the text than the Poor students' explanations because the Good students have understood the principles better prior to studying the examples. This interpretation can be ruled out by examining the content of the students' responses to the Declarative questions administered prior to the studying of the example-exercises. Recall that the students had to be able to answer a set of Declarative questions to the satisfaction of the experimenters before they could proceed to the example-studying phase of the study. One part of this declarative test asks the students to state in their own words Newton's Three Laws. We analyzed those responses using the same analysis used in the Chi, Glaser and Rees (1982) Study Six. In this analysis each of the Newton's Laws was decomposed into several subcomponents. For example, Newton's Second Law, F=ma, has been decomposed into four subcomponents:

1. Applies to one body
2. Involves all forces on the body
3. Net force is the vector sum of all the forces
4. F=ma, or the magnitude of F is ma, and the direction of a is the same as F.

Using these components as a scoring criterion, it can be seen in Table 4 that for the answers to the Declarative questions, both the Good and the Poor students encoded 5.5 components (out of a total of 12 possible components for the three Laws). Thus, before studying the examples, the two groups do not differ in the amount of declarative knowledge that they have encoded.

In order to compare the student's initial understanding of the Three Laws with their subsequent understanding as manifested in the explanations, we used the same component analysis to code those explanations which were previously judged to be derivable from Newton's Three laws. Within these subsets of explanations, the Good students verbalized 5.5 components of the Three Laws whereas the Poor students only stated 1.25 components ($t(6) = 2.14$, $p < .05$).

Table 4

Number of Components of Newton's Three Laws (Out of 12)

	GOOD	POOR
Declarative Test	5.5	5.5
Explanations	5.5	1.25
New Components	3.0	0.25

Note. The numbers represent averages per student.

What is most remarkable is that for the Good students, 3 out of the 5.5 components stated in the explanations are distinct from those previously mentioned in the answers to Declarative questions. In other words, while example-studying, the Good students gained 3 additional components of the Three Laws. For example, three out of four Good students inferred or understood the second component, that the force "involves all the forces on the body." This particular component of the Second Law was not mentioned initially by any of the Good students in the answers to the Declarative questions. However, self-explanations produced while studying the examples obviously instantiated this point. This suggests that the Good students can learn from the examples, perhaps because the examples instantiate components of the Laws that were not particularly salient from reading a declarative description of it in the text. Perhaps also, this is why the Good students feel so compelled to "explain" while studying, since they are learning and encoding new knowledge.

The Poor students, in contrast, hardly added any new components to what they already knew (0.25 new components, see Table 4). The difference in the number of new components added is significantly different between the Good and the Poor students ($t(6) = 3.67$, $p<.01$). None of the Poor students articulated, for example, the second component of Newton's Second Law, in either their explanations or their declarative answers. Thus, the Poor students did not add much new knowledge to what they already knew about the Three Laws, and further, did not use what they did know.

Summary. Our analyses of explanations show that not only do the Good students produce a significantly greater number of explanations, but their explanations consist almost entirely of justifications which offer inferences about the conditions, the consequences, the goals, and the meaning of various mathematical actions described in the example. On the other hand, over a third of the Poor students' explanations are diagrammatic statements which do not add any inferences to what is already described. Furthermore, the explanations that the Good students provide are largely guided by the principles, concepts, and definitions introduced in the text.

We found that both the Good and the Poor students had basically the same understanding of these Three Laws initially, in terms of the number of components of the Three Laws that they understood, as cited in their answers to Declarative questions. (According to Greeno & Riley, 1987, explicit verbalization is the strictest criterion for assessing understanding of a principle). The self-explanations provided by the Good students while studying the examples made their understanding of the principles more complete (8.5 out of 12 components). For example, after their self-explanations, the Good students understood 3 out of the 4 components of Newton's Second Law (the only component missing is the first one), whereas prior to that, they only understood one of the four components (the last one). In contrast, since the Poor students did not provide many explanations, they did not advance their understanding of the Laws after studying the examples (5.75 out of 12 components). For example, for the Second Law, no new component was added to what they already knew (which was only the fourth component of the Second Law). Hence, their understanding of the Second Law after studying the example was just as incomplete as it was before studying the example. In fact, they did not use as many components in their explanations as they could have from their initial understanding.

Since all the students seem to have encoded the same number of components of the Three Laws initially, we are assured that the students did acquire the same amount of relevant knowledge from the chapter. In particular, the Good students did not have a better understanding of Newton's Laws prior to studying the examples. However, we further propose that by the time students attempt to solve problems, their representations of the principles and other declarative knowledge introduced in the text will differ depending on the degree to which their understanding of the principles is enhanced during their studying of examples. Such an interpretation is consistent with the findings showing that students prefer text materials that contain examples (Reder et al., 1986; LeFevre & Dixon, 1986). Thus, even though both the Good and the Poor students can begin problem solving by applying general weak methods (the process of proceduralization, Anderson, 1987), the Good students will achieve greater success in solving problems since an instantiation of the weak methods depends on the representations of their declarative knowledge, which is more complete. Thus, example-studying is a critical phase of skill acquisition.

It is interesting to compare these findings to our expert-novice results, in which the experts cited many more components in their summaries of Newton's Three Laws than the novices (Chi, et.al., 1982). In contrast, before studying the examples, our Good and Poor students did not differ in their declarative understanding. Both groups have relatively poor understanding. As a result of differential learning from examples, the two groups developed knowledge differences which are compatible with differences we found between experts and novices. This suggests that experts became that way not because they were better at encoding the declarative knowledge, but their greater understanding was probably gained from the way they studied examples and solved problems.

While we do not understand the mechanism underlying self-explanations at this point, it is clear that self-explanations not only construct better problem solving procedures, but they also help the students understand the underlying principles more completely. (More speculations about the mechanisms of self-explanations will be presented in the Discussion).

Monitoring Statements

So far, we have found that the Good students tend to explain an example to themselves (more so than the Poor students), and that these explanations are manifestations of learning that the Good students are undertaking while studying. The Poor students are not learning, at least not learning the components of the Three Laws, while studying the examples. Another possibility for why the Poor students are not elaborating is that they may not realize that they do not understand the material: another possibility is that they may actually think that they do understand the material, thus they need not elaborate. To see if these conjectures are viable, we have examined the kind of elaborations that have been categorized as monitoring statements. As we noted earlier in Table 2, around 40% of the elaboration statements are of this kind (39% for the Good students and 42% for the Poor students).

Monitoring statements can be further broken down into those that indicate that the student understood what was presented in the example line (such as "Okay," or "I can see now how they did it") and those that indicate that the student failed to understand (usually questions raised about the example line, such as *"Why is mgsinθ negative?"*). By examining the proportion of each type of monitoring statements, (see table 5), it seems that the Poor students detect comprehension failures less frequently than the Good students. Thus, the Poor students, within each example, generated an average of only 1.1 statements indicating that they failed to comprehend a line; whereas the Good students generated an average of 9.3 such comments ($t(6) = 2.32$, $p<.05$). Common sense would have predicted that the Poor students should have detected comprehension failures more frequently, rather than less frequently, since they were less successful at solving problems than the Good students.

Why is it important to be able to detect comprehension failures? We surmise that it is important to be able to detect comprehension failures in order for students to know that they ought to do something to try and understand. One way to assess this is to look at the frequency with which a detection of comprehension failure is followed by explanations. Indeed, for both the Good and Poor students, detections of comprehension

Table 5

Monitoring Statements

	GOOD		POOR	
	Prop.	Freq.*	Prop.	Freq.*
Comprehension Failure	46%	9.3	15%	1.1
Understand	53%	10.8	85%	6.2
Total	100%	20.1	100%	7.3

*An average number of statements per student per example.

failures do initiate explanations, although more so for the Good than for the Poor students, both proportionately and on an absolute basis. Eighty-five percent (or 8.0 out of 9.3 statements) of the Good students' and 60% (or 0.6 out of 1.1 statements) of the Poor students' detection of comprehension failure is followed by explanation, ($F(2,12) = 12.5$, $p < .01$). These results suggest that one crucial advantage of the Good students is in their ability to spontaneously identify the loci of their comprehension failures, which in turn initiate the necessary inferencing process.

Not only do the Poor students often fail to realize that they do not understand, but on the few occasions that they do think they do not understand, these occur strictly at loci of quantitative expressions, whereas only half of the Good students' detection of comprehension failure arise from such loci. The other half reside in places where we believe important physics principles and concepts are being explained. Our favorite example is the third line in Fig. 1,

Consider the knot at the junction of the three strings to be the body.

At this location, 3 of the 4 Good students indicated comprehension failure, whereas <u>none</u> of the Poor students did. (This is a critical piece of information because it tells the student that the <u>knot</u> and not the mass of the block, should be the center of the reference frame at which all the forces have to sum to zero.)

There is also a qualitative difference in the kind of questions that detection of comprehension failure raises. For example, Poor students, when detecting their failure to comprehend, often state their lack of understanding in a general way, such as:

"Well, what should you do here?"

or restate the equation that they do not understand, such as:

"I was having trouble with F-mgsinθ=0".

On the other hand, the questions that the Good students raise are specific inquiries about the physical situation described in the example. For example,

"I'm wondering whether there would be acceleration due to gravity?'

or

"Why the force has to change?"

Thus, it appears that the specific questions that the Good students pose can be answered by engaging in self-explanations, since the processes of self-explanations are essentially the processes of inferring the conditions and consequences of actions, inducing goals, and so on. On the other hand, the general lack of understanding as posed by the questions of the Poor students cannot be resolved by engaging in self-explanations.

In summary, the analyses of the monitoring statements essentially show that the Good students realize that they do not understand more often than the Poor students. The Poor students, in fact, seldom detect comprehension failures. When they do, it always occurs at loci of quantitative expressions. The ability to detect comprehension failures is important because such states (of incomprehension) tend to initiate explanation. When the Poor students do detect comprehension failures, their sense of it is vague and general, whereas Good students ask very specific questions about what they don't understand. These specific questions can potentially be resolved by engaging in self-explanation. In short, the Poor students seem oblivious to the fact that they do not understand, in part because they only have a superficial understanding of what they read. In the next section, we will discuss how these different patterns of learning dictate how examples are used during problem solving.

Problem Solving Protocols and References to Examples

We now turn to our analyses of student's problem solving protocols with reference to the use of examples for the 12 isomorphic problems (4 per each of the 3 example problems in the text). A global analysis is to examine the extent to which Good and Poor problem solvers use examples. References to examples were identified in three ways: (a) the student could make explicit remarks such as, "Now, I'm going to look at the example."; (b) the student could be looking at the example, rereading some of the example lines; and (c) the student could respond that she/he is looking at the example if the experimenter probed and asked what she/he was doing.

Both the Good and Poor students use examples frequently. The Good students referred to the examples in 9 out of the 12 isomorphic problems, and the Poor students referred to examples in solving 10 out of the 12 problems. Thus, at this global level, of whether or not students reference examples, our finding is consistent with those of Pirolli and Anderson (1985), as well as with the common intuition that students do refer to examples in learning to solve problems. However, a more detailed analysis shows that they reference examples in a very different way.

Reference to Example Episodes

At a more detailed level, we analyzed the number of "episodes" in which students refer to examples within each problem solving protocol. Only the first set of three isomorphic problems (one per example) are used for this analysis. We analyzed only the first set because we were interested in the students' initial use and reliance on the studied example; subsequent problem solving could have benefited from using procedures that were learned while solving the first set.

We can classify example usage into three categories: (a) Reading, (b) Copy and Map, and (c) Compare and Check. A Read episode is simply when students reread verbatim one or more example lines. A reading episode is marked by a reference to the example, reading some consecutive lines (with or without comments inserted), and terminating when the student goes back to solving the problem. Copy and Map episodes are instances where the student copies from the example either an equation, labels from the diagram, the free body diagram itself, or the axes. The following quote for student S105 would be coded as mapping:

> *"OK, so choosing the axis from the diagram, it would be better to tilt it 30 degrees."*

Compare and Check episodes are those in which the student turned to the example to check a specific subprocedure or a result, either before or after an independent attempt on the solution. For example, one of the students (SP1) forgot what the units for weight were after resolving two vectors into four components and referred back to the example with the comment

> *"In the example problem, how did they refer to weight?"*.

Table 6 shows the mean number of episodes of each kind for the problem solving protocols of the first set of isomorphic problems. Although we showed earlier that at a global level, both the Good and Poor students use examples equally often, this more detailed analysis shows that within a problem solving protocol, the Good students do refer less often to examples than the Poor students (2.7 episodes per problem vs. 6.6 episodes per problem), although the difference is only marginally significant ($F(1,6) = 3.46$, $p < .10$). (It is difficult to obtain dramatic differences statistically because the grain-size of this particular unit of analysis--the number of episodes--is rather large.) The less frequent references to the example suggest that the Good students probably have extracted more out of the examples while studying it (as evidenced by the greater amount of elaboration), so that they may not need to refer to it as often during problem solving.

This interpretation is supported further by comparing the different kinds of reference to example episodes. The distribution of the different episodes is significantly different for the two groups ($F(3,18) = 3.81$, $p < .05$ for the interaction), predominantly because there is a substantial difference in the number of Reading episodes between the Good and Poor students (0.6 vs. 4.1, HSD $p < .05$, Tukey test). That is, consistent with our interpretation, because the Poor students did not get much out of the examples when they were studying it, they now need to re-read them. (See Table 6). There are no differences in the number of Map and Check episodes.

We further suggest that the Poor students need to reread the examples not only because they did not get much out of it while studying, but that they are using the examples to "find a solution". This can be seen by the number of lines that students read within each Reading episode. The Poor students read, on average, a larger number of example lines per episode than the Good students (13.0 vs. 1.6 lines, $F(3,18) = 11.1$, $p < .01$). This difference also suggests that the function of the rereading for the Good and Poor students is different. The Good students seem to reread one line in the example as a way to locate certain relevant information they need in order to check and compare their solutions, whereas the Poor students reread several lines until they encounter an equation that they can map. Once they come to the relevant information, the Good and Poor students use the same number of lines for mapping (1.3 and 1.5 lines respectively) and for checking (0.5 and 0.3 respectively).

Table 6

Use of Example during Problem Solving

No. of Episodes Per Problem		GOOD	POOR
	Read	0.6	4.1
	Map	1.1	2.1
	Check	1.0	0.2
	Total	2.7	6.6
Within Each Read Episode:			
	No. of Lines Read:	1.6	13.0
	Loci of Entry:	Equation or FBD	first line

Note. The numbers represent averages per student per problem.

The fact that Good students reread a specific line in the example while solving a problem suggests that they consult the examples after they have a plan or formulated an idea on how to solve the problem, whereas the Poor students use the examples as a way to find a solution that can be copied. Another way to support this conjecture is to look at the location at which students begin reading the example lines. All the Poor students, for their first encounters with problem solving (during the first isomorphic problem), started rereading the example from the very first line; whereas none of the Good students started reading from the beginning. The Good students' first interactions with an example usually consisted of referring to an equation or to a free-body diagram.

The contrast in our interpretation can also be substantiated by the goals the students state while referring to the examples. Good students usually enter examples with a very specific goal. One student (S101), for example, said while referring to an example line:

> *"I'm looking at the formula here, trying to see how you solve for one (Force 1) given the angle"*,

whereas Poor student refer to examples with a general global goal, such as:

> *"What do they do?"*

then proceed to read the first line.

Learning From an Example During Problem Solving

Since the Poor students spend a considerable amount of effort re-reading the examples while solving problems, it may be that some students prefer to learn in this context, while others prefer to study without a specific problem solving goal. To evaluate this conjecture, we looked for explanations generated during reference to example episodes within the problem solving protocols of the first set of isomorphic problems for each of the three examples. Table 7 shows the number of each kind of explanations, using the same taxonomy we used earlier (see Table 3).

Table 7

__Frequency of Explanations Generated While Referring to Examples in the Context of Problem solving__

	GOOD*	POOR
Justification	2.0	2.5
Diagrammatic Statement	0.25	0.25
Total	2.25	2.75

Note. The numbers represent averages per student per problem.

*Only one Good student provided explanations.

First of all, even though the Good students had as many explanations as the Poor students, (2.25 vs. 2.75), all the explanations of the Good students were generated by one student (S110) who had generated the least amount of explanations while studying examples. Many of the explanations provided by the Good students (as shown in Table 3) were generated by only three out of the four Good students. Hence, this particular

Good student (S110) does fit our conjecture that some students prefer to explain while studying and others prefer to explain while solving problems. However, the amount of explanations provided by the Poor students in the problem solving context is nowhere near the level that the Good students generated while studying examples (Compare Table 7 with Table 3). If we combine the results of the explanation data generated during problem solving together with those generated during studying, (i.e., Tables 3 and 7), the resulting table will have exactly the same pattern as Table 3.

Discussion

Our research queried the extent to which individual difference in learning to solve problems is attributable to difference in the way knowledge is encoded from the example-exercises. We found, in general, that Good students (those who have greater success at solving problems) tend to study example-exercises in a text by explaining and providing justifications for each action (i.e., their explanations refine and expand the conditions of an action, explicate the consequences of an action, provide a goal for a set of actions, relate the consequences of one action to another, and explain the meaning of a set of quantitative expressions). More importantly, their explanations relate the actions to the principles and concepts in the text, which in turn further enhance their understanding of the principles. Essentially, Good students read the example with understanding. Poor students do not often explain the example-exercises to themselves. When they do, their explanations do not seem to connect with their understanding of the principles and concepts in the text.

Good students can also accurately monitor their comprehension failures and successes, while studying examples. This is consistent with what is generally found in the literature (see Brown, Bransford, Ferrara & Campione, 1983). They seem to detect accurately when they do and do not understand. Such accuracy in detection is extremely important because it leads to an attempt to try and understand the material. Poor students, on the other hand, seem less accurate at detecting comprehension failures. When they do, these occur where mathematical expressions are being manipulated, rather than at places where conceptual principles are being instantiated.

Finally, we found that the Good students used the examples in a very different way from the Poor students. In general, Good students, during problem solving, used the examples for a specific reference, whereas Poor students reread them as if to search for a solution. This interpretation is consistent with the following set of behaviors. First, Good students refer to the examples less frequently than the Poor students, within each solution attempt. Second, when they do refer to the examples, Good students read only one or two lines, whereas Poor students read around 13 lines. The Good students read only one or two lines because they have a specific goal in mind. Once they found what they needed from the examples, both the Good and Poor students can adequately map the example and check it for accuracy. These differences in the ways Good and Poor students use examples suggest that the Good (but not the Poor) students have understood the examples when they studied it, as a consequence of the explanations they generated, and now can use the examples as a reference.

Our results stand in contrast to the mixed results often obtained in research examining the role of elaboration in problem solving. It has sometimes been found that elaboration generally do not facilitate problem solving (Reder et al. 1986; Reed et al., 1985). The difference between these studies and the present one is that we examine the elaborations and explanations students spontaneously produced, whereas these other studies provided the elaborations for the students to read. The difficulty of interpreting results of a study which provides the elaboration is that when it fails to facilitate problem solving, one cannot know whether the elaborations themselves are inadequate, or whether providing elaborations is futile.

Although the task in which we observed learning was how students studied example-exercises, we believe that our results generalize to how good and poor students learn many kinds of materials, such as declarative text (Bransford, Stein, Vye, Franks, Auble, Mezynski, & Perfetto, 1982), diagrams, tables, charts, and so on, as well as solving other types of problems. In fact, our results are reminiscent of an earlier study by Gagne and Smith (1962) which showed that asking students to verbalize while they solve a 3-disc puzzle-type problem correlated significantly with success at solving problems. Although Gagne and Smith did not report individual differences, they were able to categorize the protocols into 4 types:

1) those which oriented toward single moves in the solution of the problems, with explanations such as:

> *"only possible move"*
> *"Just to try it"*
> *"don't know"*

2) those which anticipated to the extent of two moves, with comments such as:

> *"To get at the larger disc"*
> *"to free up a space"*

3) reasons which anticipated sequences of moves, such as:

"move as with a three-disc sequence"
"if disc is odd-numbered, move to circle B"

4) and reasons which explained the principles, such as:

"move odd-numbered disc in the clockwise direction"
"move even-numbered discs in the counterclockwise direction".

In light of our present analysis, we would expect the poor solvers not only to produce fewer explanations (i.e., less articulation), but also to produce explanations that are not necessarily justifications (as those in Type 1). We would expect good solvers to produce both more explanations and also to produce explanations which basically state subgoals, (as those in Type 2), which induce a goal from a sequence of actions (as those in Type 3), and which relate actions to principles (as those in Type 4). Unfortunately, Gagne and Smith did not break their students down into good and poor solvers. In other words, we do not think that articulating an explanation per se is the critical factor (as suggested by Gagne and Smith); rather, *what* the students articulated is the most important factor.

Why do self-explanations help understanding and problem solving? To put it in another way, what is learned from self-explanations? We have several tentative conjectures, which may not be mutually exclusive. One interpretation is that self-explanations consist of the creation of inference rules that are instantiations of the principles and definitions introduced in the text. These inference rules are specific to the example presented. (They may be generalized later during episodes of learning from problem solving.) For example, we can collapse related explanations into inference rules. They are related usually if they refer to the same conditions. For example, the last 5 explanations shown in Fig. 5 (III - VII), can be converted to three inference rules (shown in Fig. 7).

Rule 1: IF there is a body, and it has weight

THEN the weight will act as a force.

Rule 2: IF there are forces upholding a body

THEN they are resistant to the body's weight and

they are equal to he orce of weight.

Rule 3: IF (there is the force of the weight of the body

and there are resistance forces...)

THEN they will all equal out.

Figure 7: Coding of lines 6-13 of Figure 5 into three inference rules.

These inference rules may be taken as instantiations of the definition of a force (Rule 1), application of Newton's Third Law (Rule 2), and instantiation of Newton's Second Law (Rule 3). In fact, Larkin and Simon (1987) found it necessary to model problem solving by the use of similar kinds of inference rules. The present study provides evidence to support such a claim. In fact, our data further support Larkin and Simon's (1987) conjecture that "students may well be unable to solve problems in part because they learn principles, and do <u>not</u> translate them into inference rules" (p. 75). As noted earlier, we found no differences in the abilities of the Good and Poor students to write down the declarative definitions and principles introduced in the text part of the chapter. However, unless students translate the declarative knowledge into specific procedural inference rules, they cannot use them to solve problems.

Another interpretation of self-explanations, based on the analyses of the explanations which related to Newton's Laws, is the idea that students, while reading the text, have only encoded a subset of the components of the Laws. That is, some features of the Laws are not noticed as important, until one sees that they are emphasized in the example. Through self-explanations, the students are relating these new components of the Laws to components they already know.

A third way to look at the role of self-explanations is to say that the declarative Laws have been encoded entirely, but that some components are more accessible than others, and thus more easily articulated explicitly (as in the case of answering Declarative questions about them). On the other hand, self-explanations during studying examples may make other latent or implicit components more accessible. Our data cannot discriminate between the states of knowledge in which this information was available but inaccessible versus not available at all.

Another possible mechanism underlying self-explanations may be that self-explanations produce a qualitative constraint network which represents knowledge of the solution steps. For instance, a system of trigonometry constraints could link together a general equation, such as $F_{ax} + F_{bx} = 0$, the diagram of forces, and a specific equation, $-F_a\cos30^{\circ} + F_b\cos45^{\circ} = 0$. When a similar problem is encountered, say one with different angles, qualitative propagation through the constraint network can yield a plan for the quantitative solution. More importantly, the constraints can propagate information in both forwards and backwards directions, so that the above constraint system could be used to generate a plan for solving for one angle given the values of the other angle and the net force. Such qualitative constraint networks may be the kind of inference structure generated by self-explanations, at least initially (VanLehn, personal communication).

Finally, models do exist in the AI literature which can generalize from a single training example (Lewis, 1986; Mitchell, Keller & Kedar-Cabelli, 1986). Such generalizations are acquired by constructing a proof (explanation) of how an example is an instance of the concept which it exemplifies, then, within the formal theory of the task domain, the explanation is generalized to cover a larger class of examples. Our data, at a gross level, provide empirical evidence to suggest that such explanation-based learning may indeed be the way students generalize from single examples. However, we need a more detailed look at these models to see the extent to which they can accommodate our data. In Lewis' (1986) model (EXPL), the system basically contains a set of heuristics for detecting the relation between the action taken and the outcome of an action. For example, in learning text editing, suppose a student is observing a demonstration of the following action and its outcome: the screen draws a box around an object, and shortly after, the object in the box disappears. An IDENTITY heuristic will be able to conjecture that the relation between the action of drawing a box around an object causes the object to disappear. Thus, the function of explanations is to apply these heuristics to come up with a hypothesis of the relation between the action and the outcome. Notice that these heuristics are domain-independent. In our data, understanding an action and its outcome requires conjecturing not only a relation between the action and its outcome, but also to explain how that action-outcome pair is an instantiation of a principle which is part of the domain theory. Thus, the greatest difference between what our students can do and what EXPL can do is that EXPL cannot capture the role of the domain knowledge. In fact, the to-be-learned domain (text-editing) is very syntactic in nature, and does not require domain knowledge in order to understand the example actions.

In Mitchell et al.'s (1986) system (EBS), however, domain knowledge is provided. The contrast between Mitchell's theory and our data is that Mitchell's system has "too much" domain knowledge in the following sense: when an example has to be proved to be an instance of a goal concept, the system already has a clear and a complete definition of this goal concept. Thus, the function of explanation is to produce a proof that the example is an instance of the goal concept. This may require that some features of the example be inferred so that the features of the example will match those defined in the goal concept. If we assume that the knowledge that permits these inferences to be made on the example concept is analogous to the domain knowledge of physics, then this particular aspect of the model does seem to simulate parts of our students' behavior. However, the key difference between the knowledge possessed by the EBS system and our students is that EBS has all the knowledge of the goal concept, whereas our students do not. For example, our students do not have a complete understanding of Newton's Laws. Furthermore, the function of students' explanations seems to be both the enhancement of their initial understanding and generation of new understanding.

In sum, while our data does seem to support these two AI models in a general way, detailed analyses point out that these models are incomplete in describing actual learning, mainly because students' initial learning from text is more limited than the EBS model, for example, and yet their subsequent learning from the examples is more sophisticated.

We conclude by once again quoting Simon (1979), with regard to the characteristics of the learner:

> It is left largely to the student (by examining worked-out examples, by working problems, or <u>in some other way</u>) to generate the productions, the condition-action pairs, that are required to solve problems (p. 92, emphasis added).

We believe that the "some other way" that students learn is via generating and completing explanations. Hence, we have provided evidence for Simon's insight that "students learn both by being taught and by self-instruction" (p. 87), whereby self-instruction is mediated by self-explanations.

References

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review, 94(2)*, 192-210.

Bransford, J. R., Stein, B. S., Vye, N. J., Franks, J. J., Auble, P. M., Mezynski, K. J., & Perfetto, G. A. (1982). Differences in approaches to learning: An overview. *Journal of Experimental Psychology: General, 111(4)*, 390-398.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Cognitive development* (Vol. III of P. H. Mussen, Ed., *Handbook of child psychology*, pp. 77-166). New York: Wiley.

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. 7-76). Hillsdale, NJ: Erlbaum.

Collins, A., & Stevens, A. L. (1982). Goals and strategies of inquiry teachers. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2, pp. 65-119). Hillsdale, NJ: Erlbaum.

Dietterich, T., & Michalski, R. S. (1983). A comparative review of selected methods for learning from examples. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine Learning: An artificial intelligence approach*. Los Altos, CA: Morgan-Kaufmann.

Elio, R., & Anderson, J. R. (1983). Effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 113*, 541-555.

Eylon, B., & Helfman, J. (1982, February). *Analogical and deductive problem-solving in physics*. Paper presented at the AERA meeting, New York.

Gagne, R. M., & Smith, E. C., Jr. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology, 63(1)*, 12-18.

Greeno, J., & Riley, M. S. (1987). Processes and development of understanding. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding*. Hillsdale, NJ: Erlbaum.

Halliday, D., & Resnick, R. (198_). *Fundamentals of Physics*. New York: John Wiley & Sons.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 227-299.

Kieras, D. E., & Bovair. S. (1986). The acquisition of procedures from text: A production system analysis of transfer of training. *Journal of Memory and Language, 25*, 507-524.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is sometimes worth ten thousand words. *Cognitive Science, 11*, 65-99.

LeFevre, J., & Dixon, P. (1986). Do written instructions need examples? *Cognition and*

Instruction, 3, 1-30.

Lewis, C. (1986). *Why and how to learn why: Analysis-based generalization of procedures* (Tech. Rep. No. CS-CU-347-86). Boulder: University of Colorado, Department of Computer Science.

Lewis, C. H., & Mack. R. L. (1982). Learning to use a text processing system: Evidence from "thinking aloud" protocols. In *Proceedings of the conference of human factors in computer systems*. New York: ACM.

Mitchell, T. M., Keller, R. M., & Kedar-Cabelli, S. T. (1986). Explanation-based generalization: A unifying view. *Machine Learning, 1*, 48-80.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology, 39(2)*, 240-272.

Reder, L. M., & Anderson, J. R. (1980). A comparison of texts and their summaries: Memorial consequences. *Journal of Verbal Learning and Verbal Behavior, 19*, 121-134.

Reder, L. M., Charney, D. H., & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. *Memory & Cognition, 14(1)*, 64-78.

Reed, S. K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11(1)*, 106-125.

Simon, H. A. (1979). Problem solving and education. In D. Tuma & F. Reif (Eds.), *Problem solving and education: Issues in teaching and research*. Hillsdale, NJ: Erlbaum.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2(1)*, 59-89.

VanLehn, K. (1986). Arithmetic procedures are induced from examples. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Erlbaum.

Wickelgren, W. A. (1974). *How to solve problems: Elements of a theory of problems and problem solving*. San Francisco: Freeman.

Notes

[1]Newton's First Law states that "Every body persists in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by forces impressed on it.