DOCUMENT RESUME

ED 294 753 SE 049 151

AUTHOR	Hardiman, Pamela Thibodeau; And Others
TITLE	The Relation between Problem Categorization and Problem Solving among Novices and Experts.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	May 88
GRANT	BNS-8511069
NOTE	27p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Cognitive Development; *Cognitive Processes; *College Science; Higher Education; *Learning Processes; *Mechanics (Physics); *Physics; *Problem Solving; Science Education; Undergraduate Students

ABSTRACT

This investigation examined the relationship between problem solving ability and the criteria used to decide whether two classical mechanics problems could be solved similarly. The investigators began by comparing experts and novices on a similarity judgment task and found that experts predominantly relied on the problems' deep structure in deciding similarity of solution, although the presence of surface feature similarity had a clear adverse effect on performance. Novices relied predominantly on surface features, but were capable of using the problems' deep structure under certain conditions. In a second experiment, groups of novices who tended to employ different types of reasoning in making similarity judgments were compared. Compared to novices who relied predominantly on surface features, novices who made greater use of principles tended to categorize problems similarly to experts, as well as score higher in problem solving. These results suggest that principles play a fundamental role in the organization of conceptual and procedural knowledge for good problem solvers at all levels. (Author/CW)

THE RELATION BETWEEN PROBLEM CATEGORIZATION AND PROBLEM SOLVING AMONG NOVICES AND EXPERTS *

Pamela Thibodeau Hardiman, Robert Dufresne and Jose P. Mestre

Department of Physics and Astronomy

University of Massachusetts

Amherst, MA 01003

May, 1988

* Work supported by National Science Foundation grant BNS-8511069. The contents herein do not necessarily reflect the position, policy or endorsement of NSF.

Abstract

This investigation examined the relationship between problem solving ability and the criteria used to decide that two classical mechanics problems would be solved similarly. We began by comparing experts and novices on a similarity judgment task and found that experts predominantly rely on problems' deep structure in deciding similarity of solution, although the presence of surface feature similarity has a clear adverse effect on performance. Novices relied predominantly on surface features, but were capable of using problems' deep structure under certain conditions. In a second experiment, we compared groups of novices at the same level of experience who tended to employ different types of reasoning in making similarity judgments. Compared to novices who relied predominantly on surface features, novices who made greater use of principles tended to categorize problems similarly to experts, as well as score higher in problem solving. These results suggest that principles play a fundamental role in the organization of conceptual and procedural knowledge for good problem solvers at all levels.

What is the relationship between problem solving ability and the criteria one uses to decide whether or not two problems would be solved similarly? To date, attempts to answer this question have focused on investigating the end points of the spectrum of problem solving skill, namely experts and novices. For experts categorization of a problem as a type suggests possible solution strategies and can directly influence ability to generate a successful solution (Hayes & Simon, 1976; Hinsley, Hayes & Simon, 1978; Newell & Simon, 1972; Simon & Simon, 1978). Research in domains such as mathematics (Schoenfeld & Herrmann, 1982) and physics (Chi, Feltovich & Glaser, 1981) indicates that experts tend to focus on the deep structure of a problem (e.g., principles and concepts that could be used to solve the problems) to decide whether or not two problems would be solved similarly. These findings suggest that when attempting to solve a problem, experts first consider what principle(s) applies most appropriately to the situation, and then decide on a strategy or procedure that will be used to instantiate the principle (Larkin, 1983, 1981, Larkin, McDermott, Simon & Simon, 1980; Simon & Simon, 1978).

The picture is different for novices. When asked to categorize problems into types according to similarity of solution, novices tend to cue on surface features (e.g., problem jargon and descriptor terms) as the primary criterion of similarity (Chi, et al., 1981; Schoenfeld & Herrmann, 1981). When asked to state the general approach they would take to solve a problem, novices usually relate detailed information (e.g., equations and facts), rather than more general principles and concepts (Chi, et al., 1981). However, as problem solving skills develop, reliance on deep structure to categorize problems increases (Niegemann & Paar; 1986).

Although we can conclude that both surface features and deep stru (? are important, and perhaps competing, attributes used in judging word problem similarity, it may be inappropriate to construe the dichotomous use of features, i.e., that experts use deep structure exclusively and that novices use surface features exclusively. What is clear is that the extent to which problem solvers rely on each type of attribute seems to be related to problem solving ability. However, little is known about either how surface features and deep structure interact in generating a problem categorization, or how problem categorization is related to problem solving ability among novices. The two experiments we report here investigate these issues. In Experiment 1, we designed a similarity judgment task that allowed us to examine the relative contributions of surface features and deep structure in experts' and novices' categorization decisions in the domain of classical mechanics. The results of Experiment 1 suggested there may be individual differences in the categorization schemes used by novices. In Experiment 2, the similarity judgment task was refined in order to investigate these possible differences and to assess the relationship between categorization schemes and problem solving ability.

EXPERIMENT 1: THE INTERACTION OF SURFACE FEATURES AND DEEP STRUCTURE IN PROBLEM CATEGORIZATION

In order to study the influence and interaction of surface features and deep structure in categorization, we designed a similarity judgment task similar to those used in studies of object categorization (Rosch and Mervis, 1975; Mervis, 1980). In our task, a model problem and two comparison problems are presented, and the subject must decide which of the comparison problems would be solved most like the model problem. The comparison problems differ in which attributes match the model problem, making it possible to investigate systematically the interaction between surface feature, and deep structure attributes in subjects' similarity judgments. We think this type of task represents an advancement over the card sorting task commonly used in problem categorization experiments, in which a subject sorts word problems written on index cards into several piles which are then labeled by the subject to indicate the relationship among all the cards in a particular pile. This task requires that the subject develop a categorization scheme dealing with all problems simultaneously. In contrast, the similarity judgment task focuses the subjects' attention on specific problems, allowing problem attributes to be systematically varied. This simplifies the prediction of outcomes based on models of expert and novice performance, as well as the data analysis and interpretation of results.

In our similarity judgment task, a given comparison problem could match the model problem in surface features (S), deep structure (D), both surface features and deep structure (SD), or neither surface features nor deep structure (N). These comparison problems were paired in a way such that one and only one problem in the pair matched the model problem in deep structure. This led to four types of pairings, which we will refer to as "comparison types": 1) S-D, 2) S-SD, 3) N-D, and 4) N-SD. If it is the case that experts and novices rely primarily on different kinds of problem attributes in making similarity judgments, then the patterns of performance expected for experts and novices should differ. Assuming that experts base their categorization decisions solely on deep structure, then they should choose the comparison problem that matches the model problem in deep structure 100% of the time, and select D, SD, D, and SD respectively in the four comparison types. Novices who base their categorization decisions solely on surface features should choose the comparison problem that matches the model problem in surface features whenever it is possible to do so. Thus, they should choose S in the S-D comparison type and SD in the N-SD comparison type. When surface features do not allow a distinction to be made, as in the S-SD and the N-D items, either alternative should be equally likely. Hence, novices' choices should match the model problem in deep structure 0%, 50%, 50%, and 100% of the time for S-D, S-SD, N-D, and N-SD comparison types respectively, and 50% of the time overall.

Method

Subjects

The novice subjects were 45 undergraduate students at the University of Massachusetts who had completed the first semester physics course in classical mechanics and received a grade of B or better. The expert subjects were 8 Ph.D. physicists and 2 advanced physics graduate students who were nearing completion of the Ph.D. requirements. The novice subjects performed both a categorization task and a problem solving task, and were paid for their time. The expert subjects volunteered their time and only performed the categorization task.

Categorization Task

Each item on the categorization task was composed of three elementary mechanics problems similar in style and level of difficulty to problems in an introductory mechanics text (i.e., Resnick and Halliday, 1977). Each word problem was three to five lines long and contained neither pictures nor diagrams. For each item, one problem was designated as the model problem, while the other two were designated as comparison problems. Subjects were to indicate which of the two comparison problems "would be solved most similarly" to the model problem. A response was considered correct if the subject chose the comparison problem that matched the model problem in deep structure (i.e., the physical principle that would be applied to solve both problems was the same).

There were eight model problems, two dealing with energy principles, two dealing with momentum principles, two dealing with angular momentum principles and two dealing with Newton's Second Law or Kinematics. Each model problem appeared four times, once with each of the four comparison types. This yielded 32 items composed of one model problem and two comparison problems. The following is a sample model problem and the four comparison problems that were constructed to accompany it:

Model Problem

A 2.5 kg ball of radius 4 cm is traveling at 7 m/s on a rough horizontal surface, but not spinning. Some distance later, the ball is rolling without slipping at 5 m/s. How much work was done by friction?

S Comparison Problem

A 3 kg soccer ball of radius 15 cm is initially sliding at 10 m/s without spinning. The ball travels on a rough horizontal surface and eventually rolls without slipping. Find the ball's velocity.

D Comparison Problem

A small rock of mass 10 g falling vertically hits a very thick layer of snow and penetrates 2 meters before coming to rest. If the rock's speed was 25 m/s just prior to hitting the snow, find the average force exerted on the rock by the snow.

SD Comparison Problem

A 0.5 kg billiard ball of radius 2 cm rolls without slipping down an inclined plane. If the billiard ball is initially at rest, what is its speed after it has moved through a vertical distance of .5 m?

N Comparison Problem

A 2 kg projectile is fired with an initial velocity of 1500 m/sec at an angle of 30 degrees above the horizontal and height 100 m above the ground. Find the time needed for the projectile to reach the ground.

The experiment was run on IBM compatible PC's. The subject was told to read the model problem carefully and the two comparison problems that would appear below it, and decide which comparison problem would be solved most like the model problem. The items were presented in random order, with no limit imposed on response time. Most subjects completed the task within 45 minutes.

Problem Solving Task

In a separate hour-long session, the novice subjects were given a problem solving task containing seven classical mechanics problem. Four problems required the application of one principle for solution, whereas three problems required the application of two principles. Henceforth we will only discuss performance on the single-principle problems, since few subjects were able to solve the two-principle problems. The single-principle problems were all similar in style and level of difficulty to both the problems appearing in the textbook and the problems used in the categorization task. The principles involved in the four problems were: Newton's Second Law, Conservation of Energy, Conservation of Linear Momentum, and Conservation of Angular Momentum. Each problem was graded on a ten point scale by two physicists; whenever the score on a problem differed by two or more points, the solution was discussed and a score was determined by consensus. The total scores on the problem solving task ranged from 1 point to 40 points, with a mean of 21.2 points and standard deviation 11.42 points.

Results

The performances of the 45 novices and the 10 experts were compared in a 2 (Groups) by 4 (Comparison Types) by 8 (Model Problems) analysis of variance. In general, the experts were better able to determine whether two problems would be solved through application of the same principle, choosing the comparison problem that matched the model problem in deep structure 78% of the time. Novices chose the deep structure alternative 59% of the time, which was significantly less often than experts, $F(1,53) = 28.78$, $p<.0001$. As predicted by our assumptions about expert and novice performance, there was a difference in how the two groups responded to the four comparison types, as indicated by the Group x Comparison Type interaction, $F(3,159) = 10.00$, $p<.0001$. Therefore, we will discuss the influence of Comparison Type for experts and novices separately.

Experts

Comparison Type should have had no influence on experts' performance had they based their decisions about solution similarity strictly on deep structure (i.e., the principle involved). However, there was a significant main effect of Comparison Type for experts, $F(3,27) = 10.56$, $p=.0001$, indicating that the four Comparison Types were not of equal difficulty. The mean performances for the Comparison Types (see Table 1) suggest that surface features have an adverse influence on experts' categorization decisions. Although the differences among these means were not all significant, they do follow the trend predicted for novices, suggesting that experts were adversely affected by the same kinds of conditions that negatively influenced novices.

Performance on the N-SD items was significantly better than that on each of the other three Comparison Types ($p<.005$, Bonferroni familywise error rate: $EF/k = .05/6 = .008$). The mean performance of the experts was significantly higher, at $p<.01$, than that of the novices for each of the four Comparison Types, except for the N-SD type, where performance was quite high for both groups (see Table 1). Thus, although experts appear to focus on deep structure to a greater degree than novices, surface features do interfere with their performance.

Novices

Consistent with the assumption that novices cue on surface features when making similarity judgments, Comparison Type did influence novices' performance, $F(3,132) = 150.54$, $p<.0001$ (means are in Table 1). All pairwise comparisons differed significantly by $p<.0002$ ($EF/k = .05/6 = .008$). These results indicate that surface features play a major role in novices' categorization schemes, and directly influence the process by which they decide whether or not two problems would be solved similarly. For example, if one comparison problem matched the model problem in both surface features and deep structure, then the decision that they would be solved similarly was facilitated (71% correct for S-SD and N-SD comparison types, versus 46% correct for S-D and N-D comparison types). However, if a comparison problem matched the model problem only in surface features, then the decision that they would be solved similarly was adversely affected (40% correct for the S-D and S-SD pairings, versus 77% correct for the N-D and N-SD pairings).

Despite their attraction to surface features as a means of judging similarity of solution, novices as a group do not seem to rely solely on surface features. For S-D items, in which novices should have been most prone to ignoring deep structure, the 95% confidence interval (CI) was $20\%<M<32\%$, well above the predicted 0% correct had they used a strict surface feature categorization scheme. Further, in the N-D items where there was no distraction due to surface features the proportion of deep structure matches was significantly above the predicted 50% performance (95% CI = 63%<M<71%).

For both novices and experts, performance was influenced by Model Problem, $F(7,371)=8.46$, $p<.0001$. Items involving angular momentum were the most difficult, while those involving energy tended to be easier. In all but one of the 8 Model Problems, experts made more deep structure judgments than did novices, producing an interaction between Model Problem and Group, $F(7,371)=3.05$, $p=.0039$. There was also a significant interaction of Model Problem and Comparison Type, $F(21,1113)=13.33$, $p<.0001$, suggesting that the difficulty of making a decision based on deep structure in the various conditions is related to the context of the problem.

The extent to which novices appear to rely on deep structure when making categorization decisions is related to their problem solving ability as measured by the problem solving task. The correlation between total categorization score and the score on the problem solving skills test was .30, $F(1,43)=4.376$, $p=.0424$. Further, performance on the problem solving task supports the notion that novices were better able to select deep structure matches on the similarity judgment task in problem contexts they understood better. More specifically, subjects displayed a poor performance in both the problem solving task and the similarity judgment task on problems involving angular momentum, whereas they displayed a relatively good performance on problems involving energy.

Discussion

The findings of Experiment 1 are consistent with the existing literature in indicating that the schemes used by subjects to categorize classical mechanics problems are related to problem solving expertise in physics. This is most obviously reflected by the greater reliance of experts on deep structure similarity in making categorization decisions, and of novices on surface feature similarity. Experts were much more likely to judge that two problems would be solved similarly if they were similar in deep structure. In contrast, novices often indicated that problems with similar surface features would be solved similarly. However, the likelihood that both expert and novice subjects will select the deep structure alternative is influenced by what other problem attributes were present in the comparison problems. Among novices, and to some extent among experts, this performance pattern could be interpreted in terms of a threshold-type model (Smith, Shoben and Rips, 1974).

If the initial perception of similarity of one of the comparison problems to the model problem was high, a threshold model would predict that the subject would be inclined to make a response based on this overall impression of similarity, without conducting any further analysis. Hence, we note the relatively high rate of choosing the surface feature alternative in the S-D and N-SD comparison types where surface features were pitted against alternatives that had no obvious superficial similarity to the model problem. If neither comparison problem succeeded in crossing the threshold of similarity (as in the N-D items), subjects were forced to consider more carefully what would constitute similarity, and hence, might be more likely to consider principles.

Is it clearly beneficial for novices to consider principles in categorizing problems merely because experts appear to do so? Experiment 1 suggests the answer to this question may be yes, since there was a correlation of novices' problem solving and categorization scores. Novices as a group varied considerably in the degree to which they judged problems as similar that were matched in deep structure. Better novice problem solvers made more similarity judgments based on deep structure than did poorer novice problem solvers. Thus, in the domain of physics, the ability to categorize problems according to deep structure seems to be beneficial to problem solving.

In Experiment 2, we will consider the issue of individual differences among novices more carefully. In doing so, we will need to clarify what types of reasoning lead to particular categorization responses, since the binary nature of the responses required in Experiment 1 did not make the subjects' reasoning explicit. For example, the assumption that novices who chose the deep structure alternative did so as a result of actually considering the problems' deep structure may not be valid. On the other hand, novices may have attempted to use deep structure more often than their actual performance indicates, but may not have been able to do so correctly. Therefore, in Experiment 2, we modified the similarity judgment task in an effort to make subjects' reasoning more explicit, as well as further explore the relationship between problem solving and problem categorization.

EXPERIMENT 2: CATEGORIZATION CRITERIA AND PROBLEM SOLVING ABILITY OF NOVICES

Study of the ends of the spectrum of problem solving skill, namely experts and novices, indicate there is a relationship between categorization criteria and problem solving ability. The relevance of this finding for understanding the development of expertise would increase if this relationship could also be demonstrated among novices at the same level of experience. Such a demonstration would indicate that skill acquisition is influenced from the beginning by the types of cues to which novices try to pay attention, and that the foundations for the acquisition of expertise are laid early in the learning process. Instruction which attempts to facilitate use of general principles may be more effective than instruction which ignores it.

In fact, the correlation between frequency of deep structure decisions and problem solving score in Experiment 1, as well as related research (such as Silver, 1979), suggests a more conclusive demonstration is possible. In order to demonstrate that categorization criteria and problem solving ability are related for novices at the same level of experience, we need a task which allows us to examine more directly the reasons subjects have for making categorization decisions. Simply inferring subjects' reasons for responses, as in Experiment 1, may be misleading -- for example, a misidentified principle may have led to an incorrect response, even though the novice was cuing on deep structure.

Therefore, we simplified the categorization task used in Experiment 1, such that there was only one comparison problem presented with the model problem in each item. This led to four comparison types: 1) S, 2) D, 3) SD, and 4) N. The comparison type nomenclature now denotes the problem attributes shared by the two problems in each item; for example, in the SD items both problems shared the same surface features and deep structure. Subjects were asked to decide whether or not the two problems would be solved similarly. They were also asked to give a reason for each response. The reasons given would provide the basis for separating subjects into groups according to the criteria they used to categorize problems. This would allow us to determine whether different patterns of responses are associated with different types of reasoning. Subjects using the appropriate deep structure reasoning should respond according to the following pattern on the four comparison types: S-No, D-Yes, SD-Yes, and N-No. Subjects using surface feature reasoning should respond similarly on the SD and N items, but in the opposite manner on the S and D items, resulting in 50% correct overall.

Method

Subjects

Forty-four undergraduate students at the University of Massachusetts who had completed a first semester physics course participated in this study. They performed the categorization task and a problem solving task, which included a mathematics proficiency component, and were paid. To have baseline data against which to compare the novice data, 7 expert Ph.D. physicists also performed the categorization task and were also paid for their participation.

Categorization Task

The word problems used in the categorization task were similar in type and difficulty to those used in Experiment 1. Two word problems were paired for each of the 32 items, one model problem and one comparison problem of type S, D, SD or N. There were eight model problems, each of which appeared four times, once with each of the four comparison problems. A response was considered "correct" if it was that expected (as defined earlier) when appropriate deep structure reasoning was used.

The task was presented in a booklet, with two problems per page. Subjects were instructed to decide whether or not the two problems would be solved similarly, and to respond by stating yes or no. They were then to provide a reason for their response. One hour was allowed for the task, and all subjects finished within this time limit.

Each of the reasons subjects gave was classified according to the following (non-mutually exclusive) characteristics: surface features, equation-based, physics terminology-based, and principles. Subjects were classified into three groups on the basis of the type of reasoning most frequently employed: 1) Surface Feature, 2) Principle, or 3) Mixed. Classification into either the Surface Feature group or the Principle group meant the subject had considered the same type of information on 17 or more of the 32 items. The members of the Mixed group employed a variety of reasoning strategies, none of which was used a majority of the time; they commonly employed equation-based or physics terminology-based reasoning on a large proportion of the items. In the novice group, 17 subjects were in the Surface Feature Group, 11 subjects were in the Principle Group, and 16 subjects were in the Mixed Group. All 7 expert subjects were primarily principle users.

Problem Solving Task

The problem solving task contained four problems, which were the same as single-principle problems given in Experiment 1. This task was a portion of a longer task which assessed both physics knowledge and mathematics proficiency. One hour was allotted for completion of the entire problem solving task.

Each of the four physics problems was graded on a ten point scale. Scores on these four problems ranged from 0 points to 34 points, with a mean of 12.4 points and standard deviation of 10.28 points. The math proficiency portion covered topics in algebra, graphing, vectors, trigonometry and geometry. Scores on the math portion ranged from 14 to 40 out of 40 possible points, with a mean of 29.5 points and standard deviation of 6.4 points.

Results

The performances of the 44 novice and 7 expert physicists on the categorization task were compared in a 2 (Groups) x 4 (Comparison Types) x 8 (Model Problems) ANOVA. As in Experiment 1, experts made more correct decisions on the basis of matching deep structure (95%) than did the novices as a group (62%), $F(1,48)=70.27$, $p<.0001$, or any of the three novice subgroups (Surface Feature Group - 56%, Mixed Group - 63%, Principle Group - 69%), $F(1,22)=113.35$, $p<.0001$, $F(1,21)=131.26$, $p<.0001$), and $F(1,16)=73.63$, $p<.0001$. However, on the average, novices made many more decisions that were correct on the basis of deep structure than one might expect; the performance of the novice subjects, at 62% correct, was significantly higher than the 50% correct predicted for novices if we assume they employ only surface features in categorization decisions, 95% CI= 59%<M<63%.

Comparison Type. For novices, but not experts, categorization performance was influenced by Comparison Type, as indicated by a 3 (Novice Groups) x 4 (Comparison Types) x 8 (Model problems) ANOVA, $F(3,129)=132.71$, $p<.0001$, and by a 4 (Comparison Types) x 8 (Model Problems) ANOVA for experts, $F(3,18)=1.30$, $p=.3061$. Novice performance on each Comparison Type differs from that of every other type at a level of $p<.001$ (EF/k=.05/6=.008). As can be seen in Table 2, subjects experienced the most difficulty in correctly rejecting S comparison problems as being appropriate matches to the model problem. This result, in combination with the high rate of correct acceptance of the SD comparison problems, supports our findings from Experiment 1 by indicating the relevance that novices attach to surface features in making decisions about solution similarity. In both experiments, the presence of surface feature similarity depressed the rate of deep structure decisions when it was uncorrelated with deep structure similarity (as in the S-D and S-SD items in Experiment 1, and the S items in this experiment) and increased the rate when it was correlated (as in the N-SD items in Experiment 1, and SD items in this experiment).

The importance of surface feature similarity for novices does not mean deep structure is not considered, as we suggested in Experiment 1. In

Experiment 2, novices were much better at correctly accepting D comparisons than expected (50% actual versus 0% predicted). When there was no competition from surface features, subjects were much more capable of making correct decisions involving deep structure. In cases where there was neither surface feature nor deep structure similarity (i.e., N comparisons), subjects were reasonably good at assessing the lack of any similarity and making a correct rejection (90% correct).

Reasoning Employed. As expected, experts nearly universally (93% of the time) provided reasons for their judgments of similarity that were based on physics principles. The principles involved were identified correctly 98% of the time. Clearly, experts reason primarily on the basis of deep structure, as their responses in both Experiments 1 and 2 indicate, and do so appropriately.

Novices differ from experts and from each other in the degree to which they utilize principles in their reasoning. On the average, members of the Principle Group mentioned principles 70% of the time, members of the Mixed Group 23% of the time, and members of the Surface Feature Group 6% of the time, so there were major differences among the groups in proclivity to employ principles in reasoning. These principles were identified correctly by the three groups 60%, 61%, and 62% of the time, respectively. Thus, when novices chose to utilize principles in an explanation, there was no difference among the three groups in the rates of correct identification, although there were marked differences in the frequency of using principles among the three groups.

The tendency to employ principles in reasoning is related to overall success in categorization. The Surface Feature Group, with 56% correct, performed significantly lower than both the Mixed Group, with 63% correct, F(1,31)=6.73, p=.0143, and the Principle Group, with 69% correct, F(1,26)=15.64, p=.0005. The three groups also tended to have difficulty with different types of problems, as indicated by an interaction of Group and Comparison Type, F(6,123)=3.08, p=.0077.

As can be seen in Table 2, the relative difficulty of the four Comparison Types was the same for all three groups. The performances on the SD and N items did not differ significantly among the groups, which is consistent with the 100% correct predicted performance for both surface feature and principle users. It was on the S and D items that the differences among the groups appeared. The Surface Feature Group performed lower than the Principle Group on both S and D items, t(26)=3.10, p=.0046 and t(26)=3.23, p=.0033. We had predicted surface feature users would make correct responses 0% of the time on these two Comparison Types, while principle users would be correct 100% of the time. Thus, the performances of these groups were in the predicted directions.

Clearly, although the performance of the Principle Group was much better than that of the average novice, their performance was far from that predicted for one who relies on principles alone. Two factors contribute to this outcome: 1) principles were not used in every problem analysis, and 2) the principles identified were often inappropriate. Of the 70% of the time that members of the Principle Group used principles, 38% of the time they identified principles incorrectly. Hence, what may be more important than the appropriateness of a principle in the development of expertise is the frequency with which one attempts to apply principles to a problem analysis.

The data argue that attempted principle use is related to problem solving ability. Mean performances on the problem solving task were 14%, 32%, and 57% correct for the Surface Feature, Mixed, and Principle Groups respectively. The three groups were significantly different on this measure, $F(2,41)=16.19$, $p<.0001$, and each group differed from the other at a level of $p<.008$. The correlation between the frequency of attempts of an individual to reason by principle and the problem solving test score was .63, $F(1,42)=27.657$, $p<.0001$. One might argue that a third factor, such as intelligence, is responsible for this relationship. However, even when level of mathematics proficiency (which we take as an index of intelligence) was held constant, the correlation was still significant, $r=.505$, $Z=3.516$, $p<.0004$. Novices who attempt to categorize problems using principles tend to be better problem solvers.

Discussion

Experiment 2 demonstrates that the relationship between use of principles in categorization and problem solving skills is not only an appropriate characteristic for making distinctions between experts and novices, but is also appropriate for distinguishing among "good" and "poor" novice physics students with similar educational experiences. Novices who attempt to analyze mechanics problems using principles make more correct judgments concerning solution similarity and are better problem solvers. Note that these novices were often incorrect in identifying the principle needed to solve a problem, but that the principle-based approach to problem categorization generally appears to have a value beyond the successfulness of the attempt to classify a problem.

Why is principle use so highly correlated with problem solving ability? We believe that storing information about types of physics problems in terms of general principles, as opposed to equations and surface features which novices generally employ (Chi, et al., 1981; Mestre & Gerace, 1986), is a much more efficient form of representing physics knowledge. The effort required to organize physics knowledge in terms of broad categories is probably initially much greater than that required to organize the knowledge in terms of equations. However, the effort involved in maintaining categorical information is much less than that needed to maintain and search through a large equation data base. Therefore, our findings indicate that three months after finishing their mechanics course, principle users could solve problems more effectively than non-principle users. Although this study cannot address the causal relation between principle use and problem solving skill, it suggests that pedagogy in physics might be more effective if attempts were made to convey information in a manner conducive to organization by principles, a view supported by other research as well (Eylon & Reif, 1984; Heller & Reif, 1984).

General Discussion

In these two studies, we attempted to characterize how and when novice physics students and expert physicists use surface features and deep structure to determine that two problems would be solved similarly, and how categorization and problem solving skills are related. In agreement with other studies (Chi, et al., 1981; Schoenfeld & Herrmann, 1982), we found the presence of surface features to adversely influence the categorization decisions of novices, and of experts to some extent as well. Despite the apparent difficulty in ignoring the semblance of similarity conveyed by surface features, the conclusion that novices focus almost exclusively on surface feature similarity is unwarranted.

Two pieces of information argue against such a conclusion. First, in conditions where surface feature similarity was not available to "assist" in a decision, the performance of novices was much better than that expected if they had been relying solely on surface feature similarity. Second, when novices were asked to state why they believed two problems would be solved similarly, many of them responded with arguments based on principles, although equation-based reasoning was also fairly common.

Novices are not a uniform crowd. Some do rely primarily on surface feature similarity to categorize problems, while others attempt to reason fairly consistently by principles. It is not at all clear that a picture of the novice progressing from reliance on surface features to reliance on deep structure is accurate. Novices who are better problem solvers, and presumably the ones more likely to continue in the field, tend to apply principles more often when deciding whether or not two problems would be solved similarly. Surface features interfere with the decision, but are not the primary focus of attention for the good problem solvers. As the novice becomes more able to distinguish the critical attributes of problems, surface feature similarity has less influence on problem analysis.

These results suggest that our goal as educators should be to structure the information presented in the classroom in a way that assists the learner in organizing knowledge by principles. There is evidence that standard pedagogical practices do not incorporate this strategy (Collins, Brown & Newman, in press). We may not be able to ensure that every student views principles as fundamental, but the centrality of principles to both experts and the better novices suggests that this path is more likely to lead to eventual understanding.

REFERENCES

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

Collins, A., Brown J.S. & Newman S.E. (In press). Cognitive apprenticeship: Teaching the craft of reading writing and mathematics. In L.B. Resnick, (Ed.), Cognition and Instruction: Issues and Agendas, Hillsdale, NJ: Lawrence Erlbaum Associates.

Eylon, B.S. & Reif, F. (1984). Effect if knowledge organization on task performance. Cognition & Instruction, 1, 5-44.

Hayes, J.R. & Simon, H.A. (1976). The understanding process: Problem isomorphs. Cognitive Psychology, 8, 165-190.

Heller, J.I. & Reif, F. (1984). Prescribing effective human problem solving processes: Problem description in physics. Cognition & Instruction, 1, 177-216.

Hinsley, D.A., Hayes, J. R. & Simon, H.A. (1977). From words to equations: Meaning and representation in algebra word problems. In M.A. Just & P.A. Carpenter (Eds.), Cognitive Processes in Comprehension (pp. 89-106). Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkin, J.H. (1983). The role of problem representation in physics. In D. Gentner and A.L. Stevens (Eds.) Mental Models. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Larkin, J.H. (1981) Enriching formal knowledge: A model for learning to solve problems in physics. In J.R. Anderson (Ed.) Cognitive Skills and Their Acquisition. Hillsdale, N.J.: Lawrence Erlbaum Associates.

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

Mervis, C.B. (1980). Category structure and the development of categorization. In R.J. Shapiro, B.C. Bruce & W.F. Brewer (Eds.), Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence and Education (pp. 279-307). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mestre, J.P. & Gerace, W.J. (1986). Studying the problem solving behavior of experts and novices in physics via computer-based problem-analysis environments. Proceedings of the Eighth Annual Conference of the Cognitive Science Society (pp. 741-746). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A. & Simon, H.A. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

Niegemann, H. & Paar, I. (1986). Classification of algebra and physi... wo.d problems by subjects of different levels of expertise (Tech. Rep. #33). Saarbrucken, Germany: Universitat des Saarlandes, Fachrichtung Allgemeine Erziehungswissenschaft.

Resnick, R. & Halliday, D. (1977). Physics. New York: Wiley.

Rosch, E. & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 8, 382-439.

Schoenfeld, A.H. & Herrmann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory and Cognition, 8, 484-494.

Silver, E.A. (1979). Student perceptions of relatedness among mathematical verbal problems. Journal for Research in Mathematics Education, 10, 195-210.

Simon, D.P. & Simon, H.A. (1978). Individual differences in solving physics problems. In R.S. Sigler (Ed.), Children's Thinking: What Develops? Hillsdale, NJ: Lawrence Erlbaum Associates.

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 81, 214-41.

Table 1: Predicted and Observed Performance of Experts and Novices

	Experts		Novices	
Comparison Type	Predicted	Observed	Predicted	Observed
S-D	100%	66%	0%	26%
S-SD	100%	71%	50%	54%
N-D	100%	84%	50%	67%
N-SD	100%	91%	100%	87%
------	------	-----	------	-----
Total	100%	78%	50%	59%

Table 2: Performance of 3 groups on 4 comparison types

	Surface Features	Mixed	Principle
S	18%	23%	43%
D	39%	53%	64%
SD	76%	87%	78%
N	90%	90%	91%
------	-----	-----	-----
All	56%	63%	69%