Journal of Educational Psychology
1987, Vol. 79, No. 4, 347–362

0022-0663/87/$00.75

Effects of Schema Acquisition and Rule Automation on Mathematical Problem-Solving Transfer

Graham Cooper and John Sweller
School of Education
University of New South Wales
Sydney, New South Wales, Australia

We carried out a series of experiments in which we used algebra transformation and algebra word problems to investigate relations between schema acquisition and rule automation on learning and transfer. We hypothesized that schema acquisition would precede rule automation and that it would have a strong effect on problems similar to initial acquisition problems. We further hypothesized that rule automation would have its primary effect on transfer and that the use of worked examples could facilitate both transfer and performance on similar problems. Experiments 1 and 2 contained simple algebra transformation problems involving the changing of the subject of an equation. The results indicated that subjects whose training included a heavy emphasis on worked examples or an extended acquisition period were better able to solve both similar and transfer problems than were those subjects trained with conventional problems. In Experiment 3, the use of verbal protocols gave some support to the hypotheses. Experiment 4, using algebra word problems, yielded data supporting the hypotheses. We concluded that in mathematical problem solving, schemas and rule automation may facilitate problem solving on different categories of problems, that schema acquisition occurs before rule automation, and that the use of worked examples facilitates the development of both.

The transfer of problem-solving skill is an enigma. Simply solving a roblem may result in minimal change in subjects' performance on subsequent problems. Three factors may help to explain the variability of findings: awareness of problem relations, schema induction, and automation of problem-solving operators.

Several studies have provided evidence that a lack of awareness of problem relations may inhibit transfer. Reed, Ernst, and Banerji (1974) found that transfer from the jealous husbands to the missionaries–cannibals problem occurred only when subjects were explicitly informed that the two problems were related. Missionaries–cannibals to jealous husbands transfer was not obtained even with this information. Gick and Holyoak (1980, 1983) found that transfer to Duncker's radiation problem from an analogous military problem was low when subjects were not told that one problem was related to the other.

Knowledge that particular problem solutions are related is unlikely to be the only important factor influencing transfer. Only limited transfer is obtained in many problem contexts, even when problem solvers are aware of analogous relations between the problems. This seems to be the case especially in mathematical problem solving. Reed, Dempster, and Ettinger (1985) found limited transfer between algebra word problems, even though subjects were informed that their having solved previous problems should assist them in solving later ones. Even allowing subjects to look at solutions to previous problems while solving subsequent ones had no effect on transfer to similar problems. Sweller and Cooper (1985), using algebra transformation problems, found that although a heavy use of worked examples improved performance on very similar problems, there was no such effect when different problems requiring the same mathematical rules were presented. Schoenfeld (1983) obtained results indicating that even problem solvers who were familiar with relevant geometric procedures failed to apply them in a new context (although this result could also have been due to a lack of awareness of relations between problems).

Schema theory has been used to explain failure to transfer under conditions in which problem solvers are aware that problems are related. Gick and Holyoak (1983) tested the hypothesis that schema induction is an important facilitator of transfer. They defined a schema as the generalized description of two or more problems and their solutions. Gick and Holyoak's results suggested that when problem solvers generated a more effective schema, transfer was enhanced. Schemas, consisting of the overlapping elements of related problems, acted as mediators providing analogous solutions and thus allowing transfer.

Automation of problem-solving operators provides another explanatory mechanism. Kotovsky, Hayes, and Simon (1985) investigated the large differences in difficulty between several Tower of Hanoi isomorphs. They concluded that working memory load was an important differentiating factor and that

This research was supported by a grant from the Australian Research Grants Scheme to John Sweller.

We wish to thank both the New South Wales Department of Education and the Catholic Education Office for the northern region of Sydney for permission to carry out research in their schools. The help of the following mathematics masters is gratefully acknowledged: Christopher Farnsworth, St. Paul's Christian Brothers High School; Ronald Clarke, Mosman High School; Walter Mastus, Forest High School; John Mackie, Davidson High School; Tony Wells, John Potter, and Tony Locastro, all of Balgowlah Boys High School.

Correspondence concerning this article should be addressed to John Sweller, School of Education, University of New South Wales, P.O. Box 1, Kensington, NSW 2033, Australia.

this load could be reduced by automating the problem rules. Furthermore, automated rules provided an effective vehicle for transfer—subsequent problems were solved more rapidly because cognitive resources could be devoted to appropriate planning rather than to the details of legitimate moves.

Facilitating Transfer

In this article, we were concerned primarily with procedures that might facilitate schema induction and automation of operators in order to facilitate transfer. For those purposes, we define a schema as a construct that allows problem solvers to group problems into categories in which the problems in each category require similar solutions. This definition is in accordance with the Gick and Holyoak (1983) definition but places a greater emphasis on expert problem solvers' ability to place problems into appropriate categories (see Chi, Glaser, & Rees, 1982; Hinsley, Hayes, & Simon, 1977).

We hypothesized that extensive transfer would not occur until problem-solving operators had been automated, at least to some degree. Initially, a schema may categorize only a limited number of very similar problems. With the development of expertise, two processes may occur. First, category boundaries may expand with an increase in the number and variety of problems encompassed by a schema. Second, and simultaneously, problem-solving operators may become automated. As a consequence, an expert problem solver faced with an unfamiliar problem is more likely than a novice to be able to incorporate it into an existing schema because his or her existing schemas encompass a greater variety of problems. In addition, because problem-solving operators are automated, the problem solver has greater cognitive capacity available to deal with those aspects of the new problem that are unfamiliar.

When it is obvious that a problem can be incorporated into existing schemas, there is a tendency to refer to that problem as being similar to the other problems from which the schemas have been derived. It is less likely that it will be referred to as a transfer problem. Because there are no precise measures of problem similarity, it is not possible to precisely define the point at which a problem should be considered a transfer problem. Nevertheless, for operational purposes, if a problem solver or class of problem solvers has difficulty solving or cannot solve a problem after being able to solve previous problems requiring the same operators, it is clearly a transfer problem. If additional practice on the previous problems subsequently results in transfer, this is not likely to be due to an increase in the number of problems incorporated in the schemas, because no new types of problems are being solved. Transfer in this case is more likely to be due to automation of problem-solving operators.

There is evidence that the normal procedure used to enhance problem-solving skill—practice at solving many conventional problems—is relatively ineffective. This evidence has been obtained from both puzzle problems (e.g., Mawer & Sweller, 1982; Sweller, 1983; Sweller & Levine, 1982; Sweller, Mawer, & Howe, 1982) and mathematical problems (Owen & Sweller, 1985; Sweller & Cooper, 1985; Sweller, Mawer, & Ward, 1983). Results suggested that either practice on goal-free problems in which a nonspecific rather than a specific goal is used or a heavy use of worked examples is more effective in developing skill than are conventional problems. Lewis and Anderson (1985), also using both puzzle and mathematical problems, found that requiring subjects to hypothesize a solution strategy before problem solution, as opposed to simply solving problems, facilitated skill development.

These results have been explained by the suggestion that the means–ends strategy used by novice problem solvers on conventional problems both imposes a heavy cognitive load and directs attention away from those aspects of a problem that are important in learning. Goal-free problems, worked examples, or hypothesized solution strategies either disrupt, prevent, or are an adjunct to the use of means–ends analysis. Suitable organization of teaching materials may reduce cognitive load, redirect attention, or both, facilitating learning.

Procedures that enhance problem-solving skill should be equally effective whether or not the test problems require transfer. Nevertheless, this result has not always been obtained. As indicated above, Sweller and Cooper (1985), using algebra transformation problems, found that a heavy use of worked examples facilitated the development of problem-solving skill on closely related problems but not on transfer problems. According to current theoretical analysis, this may have been due to the fact that problem-solving operators (in the form of algebraic rules) were not sufficiently automated. This, in turn, may have been due to the complexity of the problems used by Sweller and Cooper (1985). The use of a relatively large number of algebraic rules was required in order to solve the problems. In the limited practice periods available, narrow schemas that allowed transfer to very similar problems may have been induced. Nevertheless, the practice periods may not have been sufficient to allow automation of the many rules required.

We tested this possibility in the current series of experiments by substantially reducing the number of algebraic rules required to solve the problems. We hypothesized that as the number of practice problems was increased, schema induction would initially allow improved performance on similar test problems. As the rules became automated, we predicted, this improvement would extend to transfer problems in which the same rules had to be used in a different context. Of primary importance was our prediction that both of these effects would be enhanced by a heavy use of worked examples. We were thus in a position to hypothesize that the use of worked examples can facilitate transfer, a result not previously obtained.

Experiment 1

Sweller and Cooper (1985) used algebra transformation problems in which students had to change the subject of an equation, for example, $c(a + d)/f = g$, solve for a. Equations such as this require the use of several algebraic rules for solution. Students, in effect, were being asked to learn that addition and subtraction are inverse operations that can be used to move a pronumeral from one side of an equation to another. A similar multiplication–division rule also had to be learned. Other equations that were presented required additional rules for solution. It may be reasonable to hypothesize that automation of the relevant algebraic rules did not occur because of the large number of rules and the limited time available. This in turn may have prevented transfer. Sufficient time may nevertheless have been available to allow schema acquisition to occur, thus improving performance on highly similar problems.

We designed Experiment 1 to increase the possibility of transfer effects under worked example conditions by including very simple algebra transformation problems in it that required far fewer algebraic rules than in the problems used by Sweller and Cooper (1985). Addition and subtraction of pronumerals were the only rules required. If students need to learn how to use only a limited number of problem-solving operators, automation and transfer may occur more readily.

Method

Subjects

The subjects were 24 eighth-grade students from the top mathematics class of a Sydney high school. This class was chosen after pilot studies had indicated that the students' level of accomplishment in algebra was sufficient to allow them to complete the problem sets used, but also sufficiently low to allow substantial improvement.

Procedure

All subjects were presented with an initial explanation sheet that contained a set of three worked examples (see Appendix A). The first worked example was intended to be an introduction to the concept of algebra manipulation through use of numerals. Both of the remaining worked examples were of problem types used in the experiment.

Subjects were asked if they had any questions, and questions were answered until subjects claimed they understood the example problems. This procedure was followed in order to ensure that the concepts, general procedure, and rules required to solve algebra transformation problems were understood by all subjects.

The explanation sheet was then removed, and an acquisition phase followed. Eight problems were presented. For solution, all problems required that a be made the subject of the equation. Subjects were instructed to work as rapidly and accurately as possible. If they obtained an incorrect answer, they attempted the problem again until they solved it correctly. A maximum of 5 min was allowed for each problem. If the correct solution had not been obtained during this period, then the correct worked solution was given by the experimenter. Subjects were then asked to try the next problem. All previous work and problems were covered from view before subjects attempted succeeding problems.

Four problem formats were used: $a + b = c$; $a - b = c$; $a + b - g = c$; $a - b + g = c$. Problems conforming to each of these formats were presented in the order shown above, with each format used twice before students progressed to the next problem type. All problems differed in terms of the variables used, except that the variable a always occurred in a position identical to that in the examples.

There were two groups of 12 subjects each. The conventional problem group was required simply to solve the eight problems by using pencil and paper. The worked example group was given the same problems, except that the first problem of each pair of identical format problems had the solution written out in a manner similar to that on the initial explanation sheet. Subjects were informed that they should study each worked example until they were sure they understood it, because the following problem would be similar.

Three test problems followed, each requiring a to be expressed in terms of the other variables. The first was $a - k = t$, and the second $a + c - n = s$. Both of these problems came from the domain of problem types that had been seen during the acquisition phase of the experiment. They are referred to as similar test problems. The third problem was $b + c - f = g + a - v$. This problem is of a different format from those seen during the acquisition phase. There were six variables instead of the three or four present in the acquisition problems, and the variable a was moved from its consistent position on the extreme left hand side of the equation to a position between variables on the right-hand side of the equal sign. Nevertheless, the third test problem could be solved by using the same algebraic manipulations used in the solution of the acquisition problems. Furthermore, like the second test problem, the transfer problem could be solved in two steps.

A maximum of 5 min was allowed for each of the test problems. Within the 5-min period, subjects were informed of their incorrect solutions and allowed to make additional attempts. If the correct solution was not obtained at the end of 5 min, subjects were asked to attempt the next problem. No solutions were given to unsolved problems, and all previous problems and work were unavailable to subjects. All subjects were tested individually.

Results and Discussion

Completion time and number of mathematical errors during both the acquisition and test periods are of interest. Medians are presented in Table 1. Because of the distorted (skewed) distributions, nonparametric tests were used. A Mann-Whitney U test indicated that the worked-example group required significantly less time during acquisition than the conventional problem group, $U(12,12) = 41$. (An alpha of .05 is used throughout this article.) Inspection of Table 1 (which separates the acquisition period into initial and repeat problem presentations) indicates that this difference is caused directly by the initial example problems, on which the worked-example group spent less time than the conventional group, $U(12,12) = 34$, and not by repeat presentations, which did not differ significantly, $U(12,12) = 67$. There was no difference between groups in mathematical errors on repeat presentations during acquisition, $U(12,12) = 69$. There was no opportunity for errors by the worked-example group on the initial problem presentations. The conventional-problem group made significantly more errors on the initial problems than on the repeat problems when a Wilcoxon matched-pairs signed-rank test was used, $T(9) = 1.5$. (The Wilcoxon matched-pairs signed-ranks test has an adjusted N according to the presence of any zero differences. Only nonzero differences influence the statistic. This accounts for the N of 9 above.)

There was no significant difference between groups on either of the similar test problems for time to solution; $U(12,12) = 69$ for the first similar test problem, $U(12,12) = 70$ for the second test problem. Similarly, there were no differences in mathematical errors; $U(12,12) = 60$ for the first similar test problem, $U(12,12) = 66$ for the second similar test problem. For the transfer problem, the worked-example group spent significantly less time on solution, $U(12,12) = 18$, and made significantly fewer mathematical errors, $U(12,12) = 25$.

Table 1
Median Seconds and Mathematical Errors in Experiment 1

	Acquisition phase		Test phase		
			Similar test problems		
Group	Initial presentation	Repeat presentation	First	Second	Transfer problem
Conventional					
Median seconds	392.3	103.0	15.0	27.5	300.0
Mathematical errors	4.0	0.0	0.0	0.0	2.0
Worked example					
Median seconds	120.9	94.4	14.5	25.2	157.4
Mathematical errors	NA	0.0	0.0	0.0	0.5

Note. The worked-example group scored no errors on the initial presentation of acquisition problems due to the presentation of worked examples. The acquisition scores represent median total seconds and errors over all relevant acquisition problems.

The worked-example subjects displayed superior performance on the transfer problem, despite their having spent less time in acquisition. In the 5 min available, 9 of the 12 worked-example subjects were able to solve the transfer problem, compared with none of the 12 in the conventional-problem group. Conversely, of the 12 subjects in the conventional-problem group, 9 made two or more mathematical errors on the transfer problem, whereas only one of the worked-example subjects made two or more errors. The worked-example group was clearly more proficient in solving the transfer problem.

We had predicted that heavier use of worked examples, by switching attention away from goal-directed search, should assist in both schema acquisition and rule automation. This in turn should result in superior performance, primarily due to schema acquisition, of the worked-example group when compared with the conventional-problem group on similar problems. Superior performance for this group was also predicted on the transfer problems, in this case due to rule automation. In fact, there were no differences between groups on the similar test problems. This result directly conflicts with Sweller and Cooper's (1985) results, which indicated that when subsequently tested on similar problems, subjects trained on worked examples were faster and made fewer errors than those trained conventionally.

The reduced complexity of schemas and the number of algebraic rules required to solve the current problems may have resulted in both groups being able to fully acquire the schemas necessary for solution of similar test problems. The absence of a significant difference could thus be explained by asymptotic effects. Both groups may have gained considerable proficiency at solving the acquisition problems, resulting in no measurable differences in performance on the similar test problems. The results provide some support for this theory. Median times for solution of the similar test problems were very low (see Table 1). Furthermore, only three subjects (one subject from the conventional-problem group and two from the worked-example group) displayed any mathematical errors at all on the similar test problems. These results are consistent with the idea that subjects from both groups had fully acquired the schemas appropriate to the acquisition problems.

Further evidence for the effects of schema acquisition can be seen most clearly by comparing performance on the similar test problems and on the transfer problem. Schemas acquired during acquisition can be used directly to solve the similar test problems but are probably of limited use for the transfer problem. This problem was the first problem seen by students in which the variable to be solved for was on the right-hand side of the equation and there was more than one variable on the other side. No opportunity to develop a generalized schema (see Gick & Holyoak, 1983) that incorporated both the acquisition and transfer problems was presented to subjects. Although schemas could be used to solve the first two test problems, students had little choice but to use a search strategy such as means–ends analysis when solving the transfer problem.

Rule automation, on the other hand, should not differentially affect performance on the two types of test problems (similar and transfer), because the two similar test problems and the transfer problem require similar algebraic transformations. In fact, the second test problem and the transfer problem require identical algebraic transformations. Consequently, differences between these two problems should provide some evidence of the influence of schema acquisition.

All 24 subjects spent more time on the transfer problem than on the two similar test problems combined. Furthermore, for both groups, significantly more errors, determined by our use of a Wilcoxon test, were made on the transfer problem than on the second similar test problem, $T(6) = 0$, for the worked-example group and $T(11) = 0$, for the conventional-problem group.

These results indicate clearly that the transfer problem is substantially more difficult than the second test problem. Because rule operators are identical between these problems, but schemas are not, we propose that the differential performance is due to differential applicability of schemas.

Evidence that supports the effects of automation is provided by the differences in time and errors between the two groups on the transfer problem. We propose that automation is the primary factor that facilitates solution of the transfer problem by the worked-example group compared with the conventional group. Schemas, as defined in this article and as is indicated above, should be of limited use in the solution of new problems that cannot be placed into previously learned categories. Automation of problem-solving operators, in this case the rules of algebra, should enhance performance on all problems to which the operators apply, whether or not the problem solver has acquired a schema.

The difference in error rates on the transfer task may provide a relatively direct indicator of the effect of rule automation. Fewer errors might be expected with the use of a more automated operator. This result was obtained. We therefore suggest that the enhanced performance by the worked-example group on the transfer problem is due to the fact that worked examples facilitated automation of problem-solving operators.

Experiment 2

We designed Experiment 2 both to replicate and to expand Experiment 1. Although the results of Experiment 1 indicated that the worked-example group outperformed the conventional-problem group on the transfer problem, there were no differences between groups on the similar test problems. If this lack of difference is due to an asymptotic effect, then one method of creating differences would be to reduce the length of the acquisition period. We tested this possibility in Experiment 2 by varying the length of the acquisition period. Second, in order to reduce the possibility of asymptotic effects, we sought a set of slightly more difficult algebra problems. The two problems we used, $ab + f = g$ and $b(a + f) = g$ (express a in terms of the other variables) were both solved by subtracting f and dividing by b. The simplest solution requires that procedures be used in a different order for each problem. Third, Experiment 2 incorporated the independent variable of student ability.

We predicted that differences on test problems between worked-example and conventional-problem groups would depend on interactions between (a) period of acquisition (long or short), (b) problem category (similar or transfer), and (c) student ability (high or low). With a short acquisition period and students of lower ability, a difference between worked-example and conventional-problem groups may only appear on similar problems, not transfer problems. Relevant schemas may have been acquired by worked-example subjects, but automation of the algebraic rules may not have occurred to a sufficient extent in the case of either group to allow transfer. As the period of acquisition or the ability of the students (or both) increases, differences may appear on the transfer task because the worked-example group should become familiar with the rules more rapidly. Simultaneously, differences on the similar problems may disappear as both worked-example and conventional-problem groups acquire the relevant schemas.

Method

Subjects

The subjects were 104 eighth-grade students from the top two mathematics classes of each of two Sydney high schools, each of which had nine eighth-grade mathematics classes. These classes were chosen after pilot studies indicated that these students' level of accomplishment in algebra was sufficient to allow them to complete the problem sets used, but also sufficiently low to allow substantial improvement.

Students who came from the top mathematics class for each school were classified as high-ability students, whereas those coming from the second mathematics classes were classified as low-ability students. Students had been placed into classes on the basis of their performance in mathematics during the previous year. Note that because there were nine classes in each school, students from the second class were actually very competent. For the purpose of this experiment, however, they were less able than those from the top class and so were termed low-ability students. Students from lower classes were less suitable as subjects, because fewer of them displayed measurable learning in the periods available for experimentation.

Procedure

The general procedure was similar to that used in Experiment 1, with the only major difference being in the problem sets used. A minor difference was that subjects were presented with a more detailed explanation sheet (see Appendix B), to provide them with all relevant information and to reduce the number of questions asked of the experimenter. Each new algebraic transformation appeared in red (underlined in Appendix B) on white paper with all other type being black. The purpose was to draw attention to each new algebraic transformation.

Acquisition problems were identical in format to those given in problems 2 and 3 on the explanation sheet. A paired-problem presentation procedure identical to that used in Experiment 1 was again followed during acquisition. The first pair of problems was identical in format to the second problem appearing on the explanation sheet. The second pair was identical in format to the third problem on the explanation sheet. This double alternation procedure was followed until all acquisition problems had been presented. Subjects whose training included a short acquisition procedure were given a total of 4 problems, whereas those whose training included a long acquisition procedure were given a total of 12 problems. Subjects in the conventional-problem groups were required to solve all problems. Subjects in the worked-example groups had the first problem of each pair of identical format problems appear as a worked example similar to the "simplified working" problem on the explanation sheet.

There were eight groups of 13 subjects each. They were part of a complete factorial experimental design based on the three independent variables of ability (high vs. low), practice (short vs. long acquisition), and teaching method (worked examples vs. conventional problem solving during acquisition).

Four test problems, identical for all groups, were presented after the acquisition procedure. The first two test problems were identical in format to the two types of acquisition problems. These are again referred to as the similar test problems. The third and fourth test problems were extensions of the acquisition problems. Test Problem 3 was $m(ac + b) = k$, and Test Problem 4 was $f(a + b) + w = g$ (express a in terms of the other variables). These are both three-step algebra transformation problems involving alternation of subtraction and division.

Test problems 3 and 4 are transfer problems. Knowledge of the operators used during acquisition is required for solution, but the problem formats differ from those in the acquisition problems.

Results

Table 2 provides median times and mathematical errors for each of the eight groups on each of the problem sets. Mann-Whitney U tests were performed for both time and errors on the initial and repeat acquisition phases, each similar test problem, and each of the transfer problems for each of the respective worked-example versus conventional-problem

Table 2
Median Seconds and Mathematical Errors in Experiment 2

Group	Acquisition phase: Initial presentation	Acquisition phase: Repeat presentation	Test phase: Similar test problems: First	Test phase: Similar test problems: Second	Test phase: Transfer problems: First	Test phase: Transfer problems: Second
Low ability, short, conventional						
Median seconds	397.9	74.2	57.2	34.4	300.0	300.0
Mathematical errors	2	0	0	0	2	1
Low ability, long, conventional						
Median seconds	370.6	147.8	22.0	19.7	300.0	300.0
Mathematical errors	3	0	0	0	2	2
High ability, short, conventional						
Median seconds	155.6	52.7	27.2	25.9	233.4	300.0
Mathematical errors	1	0	0	0	1	2
High ability, long, conventional						
Median seconds	206.8	128.9	19.6	18.7	127.6	75.6
Mathematical errors	1	0	0	0	0	0
Low ability, short, example						
Median seconds	103.6	52.8	20.9	26.9	300.0	200.3
Mathematical errors	NA	0	0	0	2	1
Low ability, long, example						
Median seconds	181.0	149.1	16.9	17.9	105.7	165.5
Mathematical errors	NA	0	0	0	0	2
High ability, short, example						
Median seconds	75.5	45.4	20.7	24.2	111.0	97.4
Mathematical errors	NA	0	0	0	0	1
High ability, long, example						
Median seconds	129.9	141.9	15.8	20.9	106.6	124.2
Mathematical errors	NA	0	0	0	1	1

Note. The worked-example groups scored no errors on the initial presentation of acquisition problems due to the presentation of worked examples. The acquisition scores represent median total seconds and errors over all relevant acquisition problems. Short = short acquisition; conventional = conventional problem; long = long acquisition; example = worked example.

pairs. This resulted in a large number of separate tests giving rise to an increased likelihood of Type I errors. Although the likelihood of those errors presents a difficulty, it may be relatively unimportant, because the pattern of expected effects was clearly based on theoretical considerations and on previous findings. Any Type I errors should be random. The large number of tests provides greater detail about the dynamics of learning, and we feel that this compensates for the increased probability of Type I errors.

Initial Acquisition Period

Times spent on initial (first of each pair) acquisition problems were significantly lower for each of the worked-example groups over the equivalent conventional-problem groups: $U(13,13) = 30.5$ for the low-ability, short-acquisition groups; $U(13,13) = 34$ for the low-ability, long-acquisition groups; $U(13,13) = 43$ for the high-ability, short-acquisition groups, and $U(13,13) = 24$ for the high-ability, long-acquisition groups. In contrast, there was only one significant difference between the worked-example and the equivalent conventional-problem group for repeat presentations (which were conventional problems in all cases). Worked examples resulted in faster solutions for the comparison between the low-ability, short-acquisition groups, $U(13,13) = 49.5$, but there were no other significant differences: $U(13,13) = 80$ for the low-ability, long-acquisition groups; $U(13,13) = 66.5$ for the high-ability, short-acquisition groups, and $U(13,13) = 50$ for the high-ability, long-acquisition groups. (Although the last statistic returns an apparently significant value on a one-tailed test, it must be rejected, because it is in the opposite direction to that predicted and does not reach significance on a two-tailed test.)

There were no differences between equivalent worked-example and conventional-problem groups in mathematical errors for repeat presentations: $U(13,13) = 74.5$ for the low-ability, short-acquisition groups; $U(13,13) = 63.5$ for the low-ability, long-acquisition groups; $U(13,13) = 78$ for the high-ability, short-acquisition groups; and $U(13,13) = 71.5$ for the high-ability, long-acquisition groups.

Each of the conventional-problem groups made significantly more errors on the initial presentations than on the repeat presentations during acquisition: Wilcoxon matched-pairs signed-ranks tests yielded $T(9) = 0$ for the low-ability, short-acquisition group; $T(10) = 0$ for the low-ability, long-acquisition group; $T(9) = 0$ for the high-ability, short-acquisition group, and $T(6) = 0$ for the high-ability, long-acquisition group. Worked examples precluded the worked-example groups from being able to make errors on the initial presentations.

In summary, the major finding concerns the relative performance of the conventional and worked-example groups. Worked-example groups spent less time on acquisition than did conventional groups.

Test Period

Our primary concern was the conditions under which worked examples facilitated performance on the similar and transfer test problems. We predicted that as the acquisition period or ability, or both, of the students increased, any differences that favored the worked-example groups on the similar test problems would be reduced. Although the worked-example groups would attain asymptotic performance more readily than the conventional groups because they would acquire schemas more rapidly, if the ability of the students or the period of acquisition was sufficiently substantial, conventional groups also would acquire the relevant schemas and perform at an equally high level. If rule automation was slower than schema acquisition, there would be minimal automation with either worked examples or conventional problems when lower ability students or reduced acquisition periods, or both, were used. Differences between worked example and conventional groups on the transfer problems would increase with increases in the acquisition period or ability of the students. This would continue until the worked-example groups began to be affected by asymptotic effects on transfer problems as well. At this point, differences between groups would begin to decrease. In summary, decreasing differences on the similar test problems, associated with initially increasing differences on the transfer problems, would provide some additional evidence for independent processes (schema acquisition and rule automation) governing skill.

Results for the similar test problems are considered first. For both of the similar test problems, the low-ability, short-acquisition, worked-example group required significantly less time for solution than the low-ability, short-acquisition, conventional-problem group: Problem 1, $U(13,13) = 24$; Problem 2, $U(13,13) = 33.5$; and made fewer errors: Problem 1, $U(13,13) = 58$, which is significant when a normal approximation corrected for ties is used, $Z = 1.84$; Problem 2, $U(13,13) = 60$, which is also significant when a normal approximation corrected for ties is used, $Z = 1.71$. (Corrections for ties are only quoted for cases in which conclusions are altered by their use.)

The differences in time to solution and mathematical errors on the similar problems with the low-ability, short-acquisition groups largely disappear when either the high-ability or long-acquisition groups are considered. For the low-ability, long-acquisition groups, the worked-example group was faster on the first similar test problem, $U(13,13) = 36$, but there was no difference for the second similar test problem, $U(13,13) = 70$. There were no differences in errors: Problem 1, $U(13,13) = 78$; Problem 2, $U(13,13) = 84.5$. There were no differences between the high-ability, short-acquisition groups in the time required for solution: $U(13,13) = 63$ for the first and $U(13,13) = 70$ for the second similar test problems. A similar lack of difference was found for errors: $U(13,13) = 71.5$ for the first and $U(13,13) = 71.5$ for the second similar test problem. There were no differences between the high-ability, long-acquisition groups in time to solution: $U(13,13) = 66$ for the first and $U(13,13) = 58$ for the second similar test problems. Again, a similar lack of difference was found with errors: $U(13,13) = 84.5$ for the first and $U(13,13) = 84.5$ for the second similar test problem.

The similar test problem results may be summarized as follows: Differences between worked examples and conventional problems were found on all measures for the low-ability, short-acquisition groups. These differences almost entirely disappeared for students with high-ability or long acquisition periods, or both.

With respect to the transfer problems, the low-ability, short-acquisition, worked-example group achieved a solution more rapidly than the low-ability, short-acquisition, conventional-problem group on the second transfer problem, $U(13,13) = 55.5$, yielding a normal approximation corrected for ties of $Z = 1.62$, but there was no difference between groups on the first transfer problem, $U(13,13) = 80.5$. Differences in errors were not significant: Problem 1, $U(13,13) = 74$, Problem 2, $U(13,13) = 64$. In contrast, the low-ability, long-acquisition, worked-example group was faster than the low-ability, long-acquisition, conventional-problem group for both transfer problems: Problem 1, $U(13,13) = 45$; Problem 2, $U(13,13) = 42$; and made fewer errors than the conventional-problem group on the first transfer problem, $U(13,13) = 52.5$, with a normal approximation corrected for ties of $Z = 1.71$. There was no difference between the two groups in errors on the second transfer problem, $U(13,13) = 61.5$. For the high-ability, short-acquisition groups there were no differences for times: Problem 1, $U(13,13) = 65.5$, Problem 2, $U(13,13) = 58.5$. There was no difference in errors between groups for the first transfer problem, $U(13,13) = 62$, though the worked-example group made fewer errors on the second transfer problem, $U(13,13) = 54.5$, with a normal approximation corrected for ties of $Z = 1.58$. Finally, for the high-ability, long-acquisition groups, there were no differences on either times: Problem 1, $U(13,13) = 82$, Problem 2, $U(13,13) = 72$; or on errors: Problem 1, $U(13,13\_= 69.5$, Problem 2, $U(13,13) = 80$. (It should be noted that with respect to one comparison, the use of medians in Table 2 may be misleading. The information in the table indicates that there is a large difference between the high-ability, short-acquisition groups in time on the second transfer problem. This difference is not significant. Seven of the 13 subjects within the high-ability, short-acquisition, conventional-problem group failed to solve the problem, whereas within the high-ability, short-acquisition, worked-example group, five subjects failed to solve the problem, and the two longest times for subjects' solutions were 243.2 and 97.4.)

In summary, on the transfer problems, little effect due to the use of worked examples was found with the low-ability, short-acquisition groups. Effects increased markedly as the length of the acquisition period increased and to a far lesser extent as the ability of the students increased. Effects disappeared entirely when high-ability and long-acquisition periods were combined.

As expected, there were clear differences on the test problems due to the length of the acquisition periods and student ability. The combined long-acquisition groups spent less total time on test problems (consisting of both similar test problems and both transfer problems combined), $U(52,52) = 892$, $Z = 2.99$, and made fewer errors, $U(52,52) = 1103.5$, $Z = 1.62$, than did the combined short-acquisition groups. Overall performance of combined high-ability groups was superior to that of combined low-ability groups for total time on test problems, $U(52,52) = 801$, $Z = 3.58$, and also for errors, $U(52,52) = 947$, $Z = 2.63$.

As in Experiment 1, the similar test problems were solved more rapidly and with fewer errors than the transfer problems. Subjects spent more time on the two combined transfer problems than on the two combined similar test problems, $T(103) = 0$. In fact, 88 subjects spent more time on each of the transfer problems than on the two similar test problems combined. Similarly, subjects made significantly more mathematical errors on the combined transfer problems than on the combined test problems $T(86) = 20$.

Discussion

Higher ability, longer acquisition periods, and worked examples during acquisition all facilitated learning on both similar test problems and transfer problems. The effects of worked examples and their interactions with ability levels and lengths of acquisition are of primary interest. Our general hypothesis suggested that relatively high performance on similar test problems is due primarily to schema acquisition, which occurs relatively quickly, and that transfer is due primarily to automation, which occurs relatively slowly. The results support this hypothesis. They indicate that under appropriate conditions, the relative advantage of worked examples over conventional problems occurs first on similar test problems and only subsequently on transfer problems. The differences in error rates on the transfer problems show that under suitable conditions, worked examples facilitated a higher level of rule automation.

In a comparison of worked examples with conventional problems, the low-ability, short-acquisition groups display significant differences on both of the similar test problems for both times and errors. For the transfer problems, however, the only significant difference is for time spent on the second transfer problem. No differences were found for time spent on the first transfer problem or for errors on either test problem. When the acquisition periods for the low-ability groups were extended, significant differences in errors on the similar test problems disappeared. We theorize that this is due to asymptotic performance that in turn may be due to a full acquisition of the appropriate schemas. In contrast, a reverse effect operated in the case of transfer problems. Significant differences on transfer problems, limited to errors using the short acquisition groups, occurred for each problem using time and for errors on the first transfer problem. We theorize that effects due to automation of problem-solving operators become apparent only after the extended practice given to the long-acquisition groups.

We further contrast these results with those obtained with the high-ability groups. These groups may have all acquired schemas rapidly, resulting in no significant differences at all on similar test problems, irrespective of the length of the acquisition period. The high-ability groups also may attain automation of problem-solving operators relatively quickly and as a consequence, the only significant difference is for errors on the second transfer problem with the short-acquisition groups. Even this effect disappears with the long-acquisition groups. With a long acquisition period, asymptotic effects may occur for both conventional-problem and worked-example groups with respect to both schema acquisition and automation.

The results show clearly that large differences favoring worked examples over conventional problems can be obtained for both similar and transfer problems. These differences are, nevertheless, subject to narrow limits determined by subjects' abilities, lengths of practice periods, and presumably, material used. The subtle nature of the limitations that cause these differences points to the importance of using suitable subjects, procedures, and materials. It must nevertheless be emphasized that when differences were obtained, they invariably favored the worked-example groups both on acquisition and test phases for both similar and transfer problems and furthermore, that they were sufficiently large to be obtainable with relatively small numbers of subjects.

Experiment 3

We designed Experiment 3 to provide direct evidence for the hypothesized interaction of cognitive load, means–ends analysis, schemas, and rule automation. Verbal protocols were obtained from subjects by means of the short-acquisition, conventional-problem and the long-acquisition, worked-example procedures of Experiment 2. These two groups were used because we expected that they would yield the largest differences.

Method

Subjects

The subjects were 6 eighth-grade students from the top mathematics class of a Sydney high school that had five eighth-grade mathematics classes. This class was chosen after pilot studies had indicated that these students' level of accomplishment in algebra was comparable to that of the low-ability group in Experiment 2.

Procedure

The procedure was identical to that used in Experiment 2, with the only difference being that verbal protocols were taken for each subject. This resulted in iteractions between the subject and the experimenter that were not present in Experiment 2. Three subjects followed a short-acquisition, conventional-problem procedure, whereas 3 followed a long-acquisition, worked-example problem procedure.

Results and Discussion

Because we had suggested that the use of means-ends analysis may be an impediment to the acquisition of skill in the form of schemas and rule automation, it was important to demonstrate that under appropriate circumstances, subjects do use this strategy when solving algebra transformation problems. A means–ends statement was defined as any statement that mentions a goal or subgoal in conjunction with statements indicating that attempts were being made to isolate the goal variable. These may be contrasted with statements that refer to previously learned problem states or their associated moves, or both. Such statements are characteristic of schemas. The number of explicit means–ends and schema-based statements for each subject appear in Table 3. Only statements that could be categorized unambiguously were included in Table 3.

It was hypothesized that novice problem solvers would use means–ends analysis on conventional problems. With respect to Problems 1 and 3 during acquisition, it can be seen from Table 3 that the conventional-problem group made several means–ends statements. (The worked-example group was presented with worked examples, which eliminated the use of this strategy.) The following are examples of means–ends statements by subjects in the conventional-problem group: On Problem 3 ($ac + e = g$, solve for a), Subject 1 of this group stated "I'm trying to get the a equal to —— (trailed off). I don't know how to get the a down" (meaning that he did not know how to isolate the a). Subject 2 stated, "If I can find out what a is by getting rid of c somehow. . . ." Subject 3, in response to the experimenter's saying, "What do you think you're concentrating on most at the moment for this question?" replied, "Probably the ac, how to delete the e so that I can get just ac." Although statements such as these were not made very often, they were sufficiently frequent to indicate that a means–ends strategy was unambiguously being used at times during the solution process. The strategy might be used at other times without its being revealed by protocols, because subjects inexperienced in verbalization techniques do not necessarily state all goals and subgoals.

For the conventional problems common to both groups during acquisition (Problems 2 and 4), there were very few means–ends statements by either group, but the worked-example group made a total of nine schema-based statements, compared with two schema-based statements for the conventional-problem group. Subject 5 (worked-example group), for example, stated immediately upon seeing problem 4 that he was "thinking about how the other one was done." The difference in the number of explicit schema-based statements suggests that schemas appear to have a clearer effect on the worked-example group than on the conventional-problem group when they solve these problems. Note, nevertheless, that on these problems the protocols consisted almost entirely of simple, rapid, step-by-step solutions of the problems, interspersed occasionally with statements indicating that a schema based on the previously seen problem was used. With respect to the extra problems that the worked-example group received, there was only one means–ends statement, compared with a total of 12 schema-based statements.

Note that the conventional-problem group made schema-based statements on the initial problem presentations as well as on the repeat problems. Because these schemas must have been acquired from the explanation sheet, they may be taken as evidence that schemas can be acquired very rapidly. For example, on the first problem Subject 1 stated, "[This problem is] a bit like the ones I just did" (referring to the explanation sheet).

With respect to the similar test problems, 5 of the 6 subjects solved the problems rapidly and smoothly, with no discernible evidence that they had used a means–ends strategy. Their statements consisted almost entirely of a description of the moves they made. This is consonant with a schema-based approach, though few explicit schema-based statements were made. The 6th subject, from the worked-example group, was unable to solve either of the similar test problems, although he appeared not to have difficulty during acquisition. Note that this was the only subject who made any means–ends statements while attempting the similar test problems.

The statements made on the similar test problems may be contrasted with those made on the transfer problems. For the transfer problems, there were 4 subjects who made at least one means–ends statement, with most such statements being made by subjects in the conventional-problem group. No explicit schema-based statements were made. Evidence that the smooth use of the mathematical principles on the similar test problems was not due to independent automation comes from the transfer problems. These required the same mathematical rules as the acquisition problems, but in a different context. Mathematical errors began to appear in moderate numbers. Associated with these errors were statements indicating uncertainty in the use of the mathematical rules. All 6 subjects made such statements. For example, 1 subject, in response to the question "What are you thinking?", replied,

Table 3
Number of Means–Ends and Schema-Based (in Parentheses) Statements Made per Subject in Experiment 3

	Acquisition				Test	
	First four problems		Last eight problems			
Group and subject	Initial	Repeat	Initial	Repeat	Similar test problems	Transfer problems
Conventional problem						
1	2 (1)	0 (1)	NA	NA	0 (0)	0 (0)
2	5 (0)	1 (0)	NA	NA	0 (2)	4 (0)
3	3 (1)	0 (1)	NA	NA	0 (0)	2 (0)
Worked example						
4	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)
5	0 (1)	0 (3)	0 (0)	0 (4)	0 (2)	0 (0)
6	0 (2)	2 (6)	0 (4)	1 (4)	3 (0)	1 (0)

Note. The conventional-problem group received only four problems within the acquisition period, whereas the worked-example group received twelve.

"How you divide by c." The lack of appropriate schemas was also apparent. Statements such as "do something with the m ... there must be something" were common. Subjects were clearly engaging in search behavior, using poorly automated operators. The absence of highly automated rules is not surprising. These subjects were comparable to the low-ability subjects in Experiment 2. Although they demonstrated some learning of rules, they did not show signs of an asymptotic effect with respect to rule automation.

These results are summarized as follows: During acquisition, there was a clear difference between the two groups, in that the conventional-problem group appeared to use a means–ends strategy far more readily than did the worked-example group. This difference resulted from the initial presentations, in which the worked-example group was presented with a worked example. The worked-example group also made a higher proportion of schema-based statements during acquisition than did the conventional-problem group. Most subjects appeared to have used a schema-based strategy on the similar test problems, which were solved readily in all cases except for one, in which means–ends analysis appears to have been used and in which the similar test problems remained unsolved. Subjects reverted to a search strategy on the transfer problems, and all of them showed evidence of being unsure of how to use the appropriate mathematical rules. Except during the acquisition phase, there were no discernible qualitative differences between the two groups (although it is possible that the conventional-problem group was more dependent on means–ends analysis when solving the transfer problems than was the worked-example group). If the distinct strategies used during acquisition have a subsequent influence on the test problems either with respect to schema acquisition or to automation, the influence is presumably quantitative rather than qualitative and is more likely to show up in quantitatively oriented studies such as Experiments 1 and 2.

Experiment 4

Algebra transformation problems were used in the previous experiments. Algebra word problems were used in Experiment 4. We designed the experiment to investigate whether transfer effects can be enhanced by increased practice on initial problems in an area of mathematics different from that used in the previous experiments. Although we used worked examples, we did not design Experiment 4 to test again the relative effectiveness of conventional problems versus worked examples, as we did in the previous experiments. The results of the previous experiments suggested that in order for transfer to be obtained, a certain degree of rule automation is required. This suggests that some previous failures to obtain transfer may have been due to the use of procedures that failed to allow a sufficiently high degree of rule automation. With this in mind, we designed Experiment 4 to test our theoretical constructs by attempting to obtain transfer using problems of a type that other investigators have previously found to be intractable. We attempted to do this by using worked examples and comparing short- versus long-acquisition periods.

Reed, Dempster, and Ettinger (1985) investigated how well subjects could use a single example of an algebra word problem to solve unrelated, similar, and transfer problems. Subjects attempted to solve the problem and were then given the correct solution as a worked example. Reed et al. concluded that although the initial training aided in performance on subsequent similar test problems compared with unrelated problems, there was little evidence of subsequent higher performance on transfer problems. In transfer problems, identical problem operators needed to be used in a dissimilar context. (It might be noted that Reed et al. labeled similar problems as *equivalent* and transfer problems as *similar*). Limited transfer that was obtained in one experiment was ascribed to subjects' use of the formula of the worked example and their filling in the slots. (The worked example was available during the test phase).

With an interpretation of these results in terms of our current analysis, it may be argued that subjects who were presented similar problems during acquisition and test phases were able to use their newly acquired schemas to solve the test problems. This assumes, as we assumed above, that schemas can be acquired relatively rapidly. Although schemas may have been acquired by subjects during the acquisition period, these schemas would not be useful on transfer or unrelated problems. Furthermore, the period of acquisition may have been insufficient for rule automation to have occurred. Consequently, when they were presented with a transfer problem, the best that subjects could do was to substitute numerical values into the formula of the worked example for cases in which this was possible. If this was not possible because of the problem structure, or if the worked example was not available during the test phase, then even this limited performance improvement disappeared. We hypothesized that an extended acquisition period including a greater number of worked examples would facilitate automation and transfer in a manner similar to that in Experiment 2. We tested this prediction in Experiment 4 with the distance problems used by Reed et al. (1985).

Method

Subjects

The subjects were 26 twelth-grade students from the top mathematics classes of both of two Sydney high schools, each of which had 5 twelth-grade mathematics classes. The students were at a higher level of mathematics study than were average students. These classes were chosen after pilot studies indicated that the students' level of accomplishment in algebra word problems was sufficient to allow tlem to complete the problem sets used in the experiment, but also sufficiently low to allow substantial improvement.

Procedure

All subjects were presented with an initial explanation sheet that contained a single worked example of a simple distance–speed–time problem. The sheet contained the information that is presented in Appendix C.

Subjects were asked if they had any questions, and questions were answered until subjects claimed that they understood the example problem. This procedure was followed in order to ensure that all subjects were familiar with the use of the formula $D = vt$. The explanation sheet was then removed and an acquisition phase followed, in which a procedure similar to that used in the worked-example conditions of Experiments 1 and 2, in which worked examples and conventional problems were presented alternately, was used. Some procedural alterations were required. Worked examples and conventional problems were of a slightly different format (see Appendix D). This difference lay in the wording of the problems. Although the wording of worked examples differed from that of the conventional problems, the work required for solution was identical, except for a single extra step in the worked examples. Within the worked-example problems, the difference between time 1 (t_1) and time 2 (t_2) needed to be calculated, whereas for the conventional problems, the difference between t_1 and t_2 was given explicitly. Thus the conventional problems were slightly easier than the worked examples. For both worked examples and conventional problems, subjects were required to find the time it took a car to overtake another car, in a distance–speed–time problem. All worked examples were identical in format, with the difference only in the numerical values used (see Appendix D). Similarly, all conventional problems were identical in format, with a difference only in the numerical values used (see Appendix D).

All subjects were required to study each worked example for 2 min before going on to the following conventional problem, which they were told was similar to the worked example. Subjects were instructed to work as rapidly and accurately as possible on the conventional problems. If they obtained an incorrect answer, they had to attempt the problem again until they solved it correctly. A maximum of 6 min was allowed for each conventional problem. If the subjects had not obtained the correct solution during this period, they were asked to look at the next worked example. (They were not presented with solutions to unsolved problems.) All previous work and problems were covered from view before subjects attempted the succeeding problems. If subjects were unable to solve both of the first two conventional problems presented to them, the subjects were dropped from the experiment. One subject from each group was dropped from the experiment for this reason. Two groups of 12 subjects each remained. The short-criterion group was required to continue in the acquisiton phase until each subject in the group solved one conventional problem. Subjects in the long-criterion group were required to continue in the acquisition phase until each had solved five out of six conventional problems consecutively.

Once subjects had reached their assigned criterion, they were presented with a single transfer problem (see Appendix D) and allowed 15 min in which to attempt to solve it. If subjects presented incorrect answers within this time limit, they were informed that their answer was wrong and were asked to attempt the problem again. Although both the worked examples and conventional problems required subjects to find the time it took a car to overtake another car, the transfer problem required subjects to find the velocities of two vehicles in a similar context.

Results and Discussion

Completion times and number and type of mathematical errors during both the acquisition and test periods are of interest. Mathematical errors were either functional or translation errors. A translation error occurred when an equation was written with substitutions that did not correspond to or follow from the problem statement. Functional errors occurred when incorrect mathematical values were obtained from a calculation (e.g., $2 \times 3 = 7$), or when an invalid algebraic manipulation was made (e.g., $ab = c$, so $a = c - b$). Functional errors were often a combination of these two types of errors, and could not be clearly categorized as either calculation or algebraic errors. Because there were only three functional errors (one made by a short-acquisition subject on the transfer problem, one made by a long-acquisition subject on the transfer problem, and one made by a long-acquisition subject on the fourth conventional problem within the acquisition phase), all analyses were conducted on the basis of total errors. Note also that problem-solving errors in which subjects made irrelevant but mathematically correct moves or failed to use all information given within problems were not counted as mathematical errors.

Median times and errors are presented in Table 4. There was no difference in time between the short-criterion group and the long-criterion group on the first conventional problem, $U(12, 12) = 65$, or in number of errors $U(12, 12) = 70.5$. All subjects in the long-criterion group spent more time in solving their first conventional problem than their last (which for 10 subjects was their fifth conventional problem in acquisition and for the remaining 2 subjects was their sixth). Using a Wilcoxon test, we found no difference in number of errors, $T(3) = 0$. We were unable to make such a comparison for the short-criterion group, because subjects solved only one conventional problem. All subjects in the long-acquisition group solved their last conventional problem faster than all subjects in the short-acquisition group solved their last problem (which for 11 subjects was the only conventional problem seen). Note that no errors at all were made by the long-acquisition group on the last conventional problem, indicating that subjects had become highly proficient at solving the conventional problems. Furthermore, all 12 subjects within the long-criterion group solved their five conventional problems consecutively. Although 2 subjects failed to solve the first conventional problem, once they had solved a conventional problem, they had no further failures on the remaining four problems. Furthermore, they made no translation errors at all within the acquisition phase, other than those made on the first conventional problem.

Table 4
Median Seconds and Errors in Experiment 4

Group	Acquisition phase: First conventional problem solved	Acquisition phase: Last conventional problem solved	Transfer
Short criterion			
Median seconds	169.7	142.8	620.3
Errors	0.0	0.0	1.0
Long criterion			
Median seconds	156.6	44.4	397.8
Errors	0.0	0.0	0.5

Note. For the short criterion group there was only one subject who received more than one conventional problem within the acquisition phase.

With respect to the transfer problem, the long-criterion group spent significantly less time on solution than did the short-criterion group, $U(12, 12) = 34$, although there was no difference between groups in number of errors, $U(12, 12) = 54$. Note that three of the subjects in the short-criterion group failed to find numerical values for either time 1, time 2, or both, on the transfer problem. These values were essential for solution, although failure to find these values was not classified as a mathematical error. Rather, it was an error in problem-solving strategy that could arise because of cognitive overload. Although the necessary information for finding the times was given, subjects were unable to process such information, presumably because of their preoccupation with the velocities of the two vehicles. No subject from the long-acquisition group made such an error in attempting to solve this problem, which contributed to the difference in solution time between the two groups.

Eleven subjects in the short-acquisition group were able to solve the first conventional problem within the 6 min allowed. In contrast, only 2 subjects were able to solve the transfer problem within 6 min. This yielded a significant effect when a Fisher exact probability test was used. A similarly significant effect was obtained for the long-acquisition group, with 10 of the 12 subjects solving the first conventional problem within 6 min and 5 solving the transfer problem within 6 min. (A comparison of errors is invalid due to the differences in time allowed on the conventional problems and the transfer problem.) These results indicate that the transfer problem was substantially more difficult than the first conventional problem. Because rule operators between these problems were identical, the differences may have been due to schema induction based on the first worked example that immediately preceded the first conventional problem and was structurally very similar. These schemas should not have been applicable on the transfer problem, which was structurally dissimilar, leaving students with little choice but to rely solely on problem-solving search techniques such as means–ends analysis.

On the basis of our previous experiments, we predicted that an extended presentation of worked examples would assist in rule automation. This, in turn, should have resulted in superior performance on the transfer problem when subjects from the long-acquisition group were compared with subjects given an abbreviated acquisition period. The results of Experiment 4 support this hypothesis. They also support the idea that rule automation is an important factor in transfer performance and that rule automation develops relatively slowly. Single presentations of a problem may be insufficient for the development of an appreciable degree of automation, and this may account for previous failures to obtain transfer effects.

It may be suggested that the results of Experiment 4 can be explained by a lack of schemas, rather than a lack of automation, on the part of subjects in the short-acquisition group. Gick and Holyoak (1983) suggested that schemas are not likely to develop through a single presentation of a problem format but are likely to develop with the presentation of two or more problems with multiple formats. Although this explanation is highly plausible in appropriate contexts, in the current experiment (and the preceding ones), the high degree of similarity between the acquisition problems reduces its likelihood.

General Discussion

Our results have theoretical and educational implications. From a theoretical perspective, we suggest that schema acquisition and automation of problem operators may be the major determinants of skilled problem-solving performance. Schemas appear to heavily influence performance on problems sufficiently similar to those solved previously that are recognized to be within the orbit of an available schema. It might be expected that automation of problem-solving operators would also facilitate performance on this type of problem. Our experiments were not designed to enable us to address this question, but there may be grounds based on theory for assuming that this is not an important factor. Schemas, by definition, incorporate problem operators, which may make independent automation of operators irrelevant under schema-based problem solving. An operator that is used automatically when embedded in a schema may not be used automatically when used in less familiar circumstances. Automation under these special circumstances may be due entirely to the incorporation of the operator in a schema and may not occur independently of the schema. Circumstantial evidence for this possibility comes from the reduced number of errors that result from the use of mathematical rules on the similar problems compared with the transfer problems.

Automation may be an essential ingredient for successful performance on transfer problems. These problems are sufficiently different from previous ones to make schema-generated moves ineffective. Problem solvers have no choice other than to use conventional search techniques such as means–ends analysis. If, as suggested elsewhere (Owen & Sweller, 1985), means–ends analysis imposes a heavy cognitive load, automation of operators should be highly beneficial because it reduces cognitive load. Problem solvers need only be concerned with the problem space, not with the characteristics of the operators, because these are relatively well known. In contrast, it may not be surprising that problem solvers who must simultaneously use means–ends analysis and search for the correct usage of an operator may frequently have difficulty with or fail to solve the problem.

Our major finding concerns the conditions that facilitate transfer through automation. There are two points to note. First, automation appears to occur relatively slowly. Consequently, when complex problem-solving operators are used, it should not be expected that appreciable transfer will be demonstrated after the solution of one or two problems. This analysis may throw light on previous failures to obtain transfer. Transfer studies rarely give subjects extensive experience on acquisition problems before presenting the critical transfer tasks. Except under conditions in which problem-solving operators are very simple and already automated (which often occurs when puzzle problems are used), additional practice may be needed to allow automation and transfer.

The second point to note concerns the conditions under which practice is most effective. Problem-solving search strategies such as means–ends analysis that impose a heavy cognitive load and direct attention away from those aspects of a problem critical to learning may be relatively poor techniques for situations in which learning is a major goal. Automation may occur more rapidly if alternative techniques, such as a heavy emphasis on worked examples, are used. In effect, we are suggesting that under appropriate circumstances, observing the use of an operator can be more effective than successfully using it oneself.

Although automation may occur more slowly than schema acquisition, paradoxically, the major differences found between novices and experts in some areas (e.g., Chase & Simon, 1973, using chess), are more likely to be due to differential schema acquisition than to automation. There are not likely to be substantial differences between weekend chess players and grand masters in automation of legitimate moves. There are more likely to be differences in the number and sophistication of their schemas. Although individual schemas may be acquired more rapidly than automation, the large number of schemas characteristic of expertise in a complex domain may explain the enormous periods of time required to achieve high levels of skill.

It may be useful to distinguish three levels of expertise. Initially, a problem solver may be able to solve readily a limited number of problems using a limited number of schemas. General search techniques with nonautomated operators must be used on all other problems. Under these circumstances, solution may be slow and error prone. Subjects in our poorest performing groups may use techniques that fall into this category. With additional expertise, subjects may automate their problem-solving operators to a greater degree, which would allow improved performance on new (transfer) problems. Subjects in our better performing groups may fall into this category. Over a far longer period of time than that covered by our experiments, the major effect of additional experience may be to extend the range and sophistication of acquired schemas. In very complex domains, this process may continue over an effectively indefinite period.

In an educational context, the results suggest that learning to solve algebra transformation problems can best be done not by solving large numbers of conventional problems, but rather by placing a stronger emphasis on worked examples. This contrasts with current teaching techniques. Topics are normally introduced by teachers and in textbooks with an outline of the new material followed by a few example problems. Students are then usually required to solve a relatively large number of conventional mathematics problems. When learning occurs, it may be relatively slow and clumsy. In realistic teaching situations in which time on task is at a premium, the use of worked-example procedures as described within the current set of experiments may prove to be effective. Any fears that an increased use of worked examples may stifle flexibility are not supported by our data. Increased use of worked examples is capable of increasing rather than decreasing transfer performance.

References

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

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

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.

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

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

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*, 248–294.

Lewis, M., & Anderson, J. (1985). Discrimination of operator schemata in problem solving: Learning from examples. *Cognitive Psychology, 17*, 26–65.

Mawer, R., & Sweller, J. (1982). The effects of subgoal density and location on learning during problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 252–259.

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

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

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology, 6*, 436–450.

Schoenfeld, A. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive Science, 7*, 329–363.

Sweller, J. (1983). Control mechanisms in problem solving. *Memory and Cognition, 11*, 32–40.

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

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means–ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 463–474.

Sweller, J., Mawer, R., & Howe, W. (1982). Consequences of history-cued and means–ends strategies in problem solving. *American Journal of Psychology, 95*, 455–483.

Sweller, J., Mawer, R., & Ward, M. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112*, 634–656.

(*Appendix follows on next page*)

Appendix A

Explanation Sheet Used in Experiment 1

You will be solving algebra problems where you must express a in terms of the other variables. Examples are given below:

1.
$$\begin{aligned} a + 3 &= 5 \\ a &= 5 - 3 \\ a &= 2 \end{aligned}$$

2.
$$\begin{aligned} a + b &= c \\ a &= c - b \end{aligned}$$

3.
$$\begin{aligned} a + b - g &= s \\ a + b &= s + g \\ a &= s + g - b \end{aligned}$$

Appendix B

Explanation Sheet Used in Experiment 2

You will be solving algebra problems where you must express a in terms of the other variables. Examples are given below. [Note that each new move appeared in red, shown here underlined.]

1. Express a in terms of the other variables

$$\begin{array}{lll} & a + h & = u \\ \text{subtract } h \text{ from both sides} & a + h \underline{- h} & = u \underline{- h} \\ \text{cancel out } +h \text{ with } -h & a & = u - h \end{array}$$

Simplified working for this problem is

$$\begin{aligned} a + h &= u \\ a &= u - h \end{aligned}$$

2. Express a in terms of the other variables

$$\begin{array}{lll} & e(a + b) & = g \\ \text{divide both sides by } e & e(a + b)/\underline{e} & = g/\underline{e} \\ \text{cancel out } e \text{ top and bottom} & a + b & = g/e \\ \text{subtract } b \text{ from both sides} & a + b \underline{- b} & = g/e \underline{- b} \\ \text{cancel out } +b \text{ with } -b & a & = g/e - b \end{array}$$

Simplified working for this problem is

$$\begin{aligned} e(a + b) &= g \\ a + b &= g/e \\ a &= g/e - b \end{aligned}$$

3. Express a in terms of the other variables

$$\begin{array}{lll} & ab + v & = f \\ \text{subtract } v \text{ from both sides} & ab + v \underline{- v} & = f \underline{- v} \\ \text{cancel out } +v \text{ with } -v & ab & = f - v \\ \text{divide both sides by } b & (ab)/\underline{b} & = (f - v)/\underline{b} \\ \text{cancel out } b \text{ top and bottom} & a & = (f - v)/b \end{array}$$

Simplified working for this problem is

$$ab + v = f$$
$$ab = f - v$$
$$a = (f - v)/b$$

Appendix C

Explanation Sheet Presented in Experiment 4

Problem

A car travels at 30 kph for 2.5 hours. How far does it travel?

Answer

This problem is a distance–speed–time problem in which Distance = Speed × Time. This formula is often written as $D = v \times t$, where D = distance, v = speed, and t = time.

Speed = 30 kph

Time = 2.5 hours

Substituting gives the following:

$$\begin{aligned} \text{Distance} &= \text{Speed} \times \text{Time} \\ &= 30 \times 2.5 \\ &= 75 \text{ kilometers} \end{aligned}$$

Appendix D

Examples of the Problem Formats Used in Experiment 4

Example of a Worked-Example Format Used in Experiment 4

Problem

A car travelling at a speed of 20 kph left a certain place at 3:00 p.m. At 5:00 p.m. another car departed from the same place at 40 kph and travelled the same route. In how many hours will the second car overtake the first?

Answer

The problem is a distance–speed–time problem in which Distance = Speed × Time. Because both cars travel the same distance, the distance of the first car (D_1) equals the distance of the second car (D_2). Therefore

$$D_1 = D_2 \quad \text{or} \quad v_1 \times t_1 = v_2 \times t_2,$$

where v_1 = 20 kph, v_2 = 40 kph, and $t_1 = t_2 + 2$ hours. Substituting gives the following:

$$\begin{aligned} 20 \times (t_2 + 2) &= 40 \times t_2 \\ 20t_2 + 40 &= 40t_2 \\ 20t_2 &= 40 \\ t_2 &= 2 \text{ hours} \end{aligned}$$

Example of a Conventional Problem Used in Experiment 4

Problem

A car travels at the speed of 10 kph. Four hours later a second car leaves to overtake the first car, using the same route and going 30 kph. In how many hours will the second car overtake the first car?

Answer

(Note that subjects were not presented with solutions to conventional problems.) The problem is a distance–speed–time problem in which Distance = Speed × Time. Because both cars travel the same distance, the distance of the first car (D_1) equals the distance of the second car (D_2). Therefore

$$D_1 = D_2 \quad \text{or} \quad v_1 \times t_1 = v_2 \times t_2,$$

where v_1 = 10 kph, v_2 = 30 kph, and $t_1 = t_2 + 4$ hours. Substituting gives the following:

$$\begin{aligned} 10 \times (t_2 + 4) &= 30 \times t_2 \\ 10t_2 + 40 &= 30t_2 \\ 20t_2 &= 40 \\ t_2 &= 2 \text{ hours} \end{aligned}$$

Transfer Problem Used in Experiment 4

Problem

A car leaves 4 hours after a truck, but overtakes it by travelling 40 kph faster. If it took the car 2 hours to catch the truck, find how fast each vehicle was travelling.

Answer

The problem is a distance–speed–time problem in which Distance = Speed × Time. Because both the car and the truck travel the same distance, the distance of the truck (D_1) equals the distance of the car (D_2). Therefore

$$D_1 = D_2 \quad \text{or} \quad v_1 \times t_1 = v_2 \times t_2,$$

$$\text{where} \quad t_2 = 2, \qquad t_1 = t_2 + 4 = 2 + 4 = 6,$$

and $v_2 = v_1 + 40$.

Substituting gives the following:

$$v_1 \times 6 = (v_1 + 40) \times 2$$

$$6v_1 = 2v_1 + 80$$

$$4v_1 = 80$$

$$v_1 = 20.$$

Substituting in $v_2 = v_1 + 40$ yields $v_2 = 20 + 40$, thus $v_2 = 60$.

Received January 27, 1987
Revision received May 11, 1987
Accepted July 5, 1987 ■

Establishment of the Underrepresented Groups Project

The *Journal of Educational Psychology* is pleased to announce the establishment of the Underrepresented Groups Project (UGP). The UGP stems from the 1986 initiative taken collectively by the American Psychological Association's Publications and Communications Board, the P & C Board's ad hoc Committee on Increasing the Participation of Underrepresented Groups in the Publication Process, and the APA Board of Ethnic Minority Affairs. The Journal has agreed to take the lead in this endeavor.

The UGP has two interconnected purposes. First, the Journal will seek to encourage the publication of research dealing with the educational psychology of underrepresented ethnic minority populations (e.g., Asian American, Black, Mexican American, Native American, Puerto Rican). Second, the Journal will develop a mentoring process designed to foster the publication of research by junior scholars and those ethnic minority authors who may have limited experience in publishing.

The Journal invites scholars investigating the educational psychology of underrepresented groups to participate in the UGP by submitting pertinent manuscripts for publication consideration. The Journal focuses on original investigations and theoretical papers dealing with learning and cognition, particularly as they relate to issues of instruction, and with psychological development, relationships, and adjustment of the individual. If you are currently doing research in these areas with underrepresented populations, you are encouraged to submit manuscripts, as the Journal is a prominent scholarly periodical.

A special feature of the UGP is its mentoring program. Each mentee will be linked with a mentor who will provide substantive and/or methodological guidance *before* the manuscript is submitted to the Journal for regular review. Throughout the review process, the manuscript will go through the same rigorous evaluation as do all submitted papers. Currently, a number of distinguished scholars have offered to serve as mentors. The Journal will try to make a match between mentee and mentor and will monitor the process.

If you wish to have more information about the UGP, please write to

Journal of Educational Psychology
Richard R. Valencia, Associate Editor
School of Education
Stanford University
Stanford, CA 94305-3096

———Robert C. Calfee, *Editor*, and Richard R. Valencia, *Associate Editor*