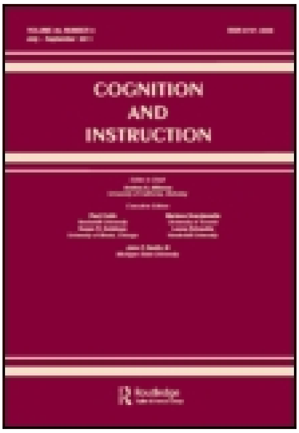

Cognition and Instruction
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hcgi20

The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra

John Sweller & Graham A. Cooper

Version of record first published: 14 Dec 2009.

COGNITION AND INSTRUCTION, 1985, *2* (1) 59-89

The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra

John Sweller and Graham A. Cooper
University of New South Wales
Sydney, Australia

The knowledge required to solve algebra manipulation problems and procedures designed to hasten knowledge acquisition were studied in a series of five experiments. It was hypothesized that, as occurs in other domains, algebra problem-solving skill requires a large number of schemas and that schema acquisition is retarded by conventional problem-solving search techniques. Experiment 1, using Year 9, Year 11, and university mathematics students, found that the more experienced students had a better cognitive representation of algebraic equations than less experienced students as measured by their ability to (a) recall equations, and (b) distinguish between perceptually similar equations on the basis of solution mode. Experiments 2 through 5 studied the use of worked examples as a means of facilitating the acquisition of knowledge needed for effective problem solving. It was found that not only did worked examples, as expected, require considerably less time to process than conventional problems, but that subsequent problems similar to the initial ones also were solved more rapidly. Furthermore, decreased solution time was accompanied by a decrease in the number of mathematical errors. Both of these findings were specific to problems identical in structure to the initial ones. It was concluded that for novice problem solvers, general algebra rules are reflected in only a limited number of schemas. Abstraction of general rules from schemas may occur only with considerable practice and exposure to a wider range of schemas.

In certain respects the teaching of mathematics and mathematically-based curriculum material is stereotyped. There are usually three steps followed: (1) Relevant information consisting of principles and relations, frequently in the form of equations, is introduced to students; (2) A relatively small number of

Requests for reprints should be sent to John Sweller, School of Education, University of New South Wales, P.O. Box 1, Kensington, New South Wales, Australia 2033.

example problems and solutions demonstrating the use of the new material is studied; and (3) A relatively large number of problems are presented for students to solve. Most time tends to be devoted to problem solving and it is the function of this activity in an educational context that will primarily concern us in this paper. Our knowledge of problem-solving mechanisms has expanded considerably in recent years and some of this new information may cast into doubt some of the heavy emphasis on problem solving in education.

Work on novice-expert distinctions has provided a major impetus in research on problem solving. DeGroot's (1966) work on chess furnishes us with the most important early contribution. He found that masters and novices did not differ on measures such as width or depth of search. Instead, major differences were found using a different measure—memory of realistic chess positions. If masters were shown a board configuration for 5 seconds that realistically could be encountered during a game, they were far superior than novices at reconstructing the configuration. This was due to enhanced knowledge rather than superior short-term memory per se, since both groups were equally poor at reconstructing random configurations. It can be concluded that during a game, due to previous experience, masters may both recognize most configurations met and know which move is most appropriate in that situation. We may describe them as having acquired *schemas.* In the present context, schemas are defined as mental constructs that allow patterns or configurations to be recognized as belonging to a previously learned category and which specify what moves are appropriate for that category. The presence or absence of these schemas may substantially explain the differences between levels of expertise.

Chi, Feltovich, and Glaser (1981) and Chi, Glaser, and Rees (1982) obtained results using physics problems that can be related to those of DeGroot. Both expert and novice physicists were presented with a series of physics problems and asked to categorize them. Experts used categories based on solution mode. For example, problems soluble by the conservation of energy principle were placed in the same category. In contrast, novices used surface features. Problems dealing with an inclined plane were classified as being similar even when different physics principles were required for solution.

These findings can be used to explain differences between experts and novices in problem-solving strategy. Larkin, McDermott, Simon, and Simon (1980 a, b) and Simon and Simon (1978) found that novices presented with a physics problem to solve relied heavily on general search strategies. An equation was chosen if it contained the goal term or a required subgoal. In contrast, experts, using previously acquired schemas, classified problems according to solution mode and then chose equations that were appropriate to the relevant category. This may be analogous to a chess master seeing a familiar board configuration and knowing from previous experience which moves should be made. Novices, not possessing the required schemas, must rely entirely on general search heuristics to guide their choice of equations.

If the major component of problem-solving skill is the acquisition of domain-specific knowledge in the form of schemas, how can these schemas best be acquired? As indicated at the beginning of this paper, the traditional answer appears to be practice on large numbers of problems. Other than the intuitive maxim "practice makes perfect" there appears to be no theoretical or empirical evidence that the solution of many conventional problems is the best way of acquiring problem-solving skill. In fact, some evidence appears to contradict this assumption.

Firstly, the voluminous research generated by the discovery learning movement (e.g., see Shulman & Keislar, 1966) provided some indication that search techniques retard knowledge acquisition rates. Secondly, using puzzle problems, employing a variety of problems under a variety of conditions, Mawer and Sweller (1982), Sweller (1983), Sweller and Levine (1982), and Sweller, Mawer, and Howe (1982) found that the conventional search strategy used by novices to obtain problem solution could be highly ineffective as a means of acquiring appropriate solution schemas. After an enormous amount of problem-solving practice, subjects could remain oblivious of a simple solution rule. Techniques which modified or eliminated problem-solving search such as the use of appropriately placed subgoals, elimination of a specific goal, or even direct information concerning previous problem solutions were required before a substantial degree of schema acquisition was observed. Using reduced goal specificity procedures, Sweller, Mawer, and Ward (1983) transferred some of these results to physics and geometry problem solving. They found that preventing solvers from using conventional search techniques enhanced the development of expertise.

A limited number of assumptions can account for these findings. The first assumption is that schemas that apply to highly specific problems must be acquired by experts. These schemas allow problem solvers to recognize a problem as belonging to a limited category that requires a particular solution mode. As an example, algebra problem solvers will learn to distinguish between $ab = c$ and $a + b = c$ (solve for a in both cases) as belonging to two distinct categories requiring different moves. A different schema must be acquired for each problem just as a different schema may be acquired by chess masters to handle differing board configurations. In any reasonably large problem domain there will be, as a consequence of limited category size, a large number of these schemas that need to be learned. The "difficulty" of problems in a particular domain may be heavily determined by the number of schemas that must be acquired.

We must also assume that the ease with which schemas can be acquired is inversely related to the amount of goal-directed problem-solving search required. Problem-solving search and schema acquisition may be unrelated, even mutually incompatible activities. Attention to one may, at least in part, preclude attention to the other. When solving a problem, a person using, for example, means-ends analysis, attempts to reduce differences between each problem state encountered and the goal state. To acquire schemas, problem state configurations and their associated moves must be distinguished and learned. There is little reason to suppose that a procedure designed to optimize goal-directed activities (problem-solving search) will also optimize learning to distinguish problem-state patterns and associated moves (schema acquisition).

One technique which might optimize schema acquisition would be to make far heavier use of worked examples than commonly occurs in teaching. Studying a worked example does not require extensive use of search techniques but does provide the information required for schema acquisition. Attention is shifted from the goal of a problem to relations between relevant problem states and relevant moves.

It must be noted of course, that worked examples are hardly a novelty in education. They are used extensively. Nevertheless, in most mathematics and mathematics-based courses, the emphasis tends to be placed on conventional problem-solving activity rather than on worked examples. If the above suggestions have validity, this emphasis is misplaced. The conditions facilitating schema acquisition are not met under normal educational conditions. Far greater use should be made of worked examples and far less use of problem solving.

Algebraic manipulation problems were used in the current series of experiments. There has been little work carried out on these problems (in contrast to algebra word problems). Lewis (1981) attempted and failed to find general novice-expert distinctions using algebra manipulation problems. Experiment 1 is partly intended to provide conditions that will allow the demonstration of novice-expert differences using these problems. Its major purpose is to provide evidence that the acquisition of a large number of limited, domain-specific schemas is an important component of expertise. Subsequent experiments test the effects of using example-studying procedures rather than problem solving on the acquisition of problem-solving expertise in general and schemas in particular.

EXPERIMENT 1

Most studies concerned with novice-expert distinctions have compared academics with high levels of expertise in the relevant area with undergraduate students. As previously indicated, the differing strategies used by experts and novices is a major finding. We cannot expect this conventional subject population to be useful when studying algebraic manipulation problems.

Lewis (1981) provides evidence for the difficulties associated with the use of undergraduates and academics as novices and experts, respectively. While differences were found in lengths of solutions and particular strategies applicable to specific problems, most differences appeared attenuated and the major conclusion was the surprising lack of difference. Under conventional educational conditions, high levels of expertise in algebraic manipulation may develop before students reach university undergraduate level. As a consequence, university students and mathematics professors may be almost equally expert at simple algebra. The use of students in different high school years may yield larger novice-expert distinctions.

Two measures of expertise were used in Experiment 1 – memory of algebra equations and occurrence or absence of *Einstellung*. Memory for algebra equations can be used to differentiate between experts and novices in a manner similar to that used by DeGroot (1966) to distinguish between masters and novices in chess. The previously outlined assumptions suggest that expertise requires the development of a large number of schemas that apply to narrow categories of problems. Some evidence for this suggestion would be obtained if experts have better recall of briefly seen equations than novices. These differences should not occur if symbols used in algebraic equations are presented randomly rather than in conventional form.

The presence or absence of *Einstellung* can also be used as an indicator of expertise. For example, Sweller, Mawer, and Ward (1983) found that problem solvers who had developed schemas to solve geometry problems tended to use those schemas to attempt to solve subsequent problems that appeared similar but in fact, were different. The attempted use of an inappropriate schema led to *Einstellung*. Under certain conditions (which are applied in the present experiment and are discussed later), this relation between *Einstellung* and expertise can be completely reversed with the failure to obtain the effect indicating expertise.

In order for *Einstellung* to occur, two defining conditions must be met: (1) problem solvers must develop an appropriate schema; (2) they must fail to detect essential differences between a category (or categories) of initial problems and a different category of critical *Einstellung* problems. If an experimental design involves training on categories of initial problems but not the critical *Einstellung* problem, then the occurrence of the effect indicates that schemas appropriate to the initial problems have developed. The development of such schemas provides evidence of a degree of expertise. If on the other hand, prior training has involved both initial and critical categories, schemas for both may have developed and problem solvers may be able to distinguish between initial and critical problems. Under these conditions, the nonoccurrence of *Einstellung* can be used as an indicator of expertise.

In summary, if prior training or experience has not included the critical *Einstellung* problems, the occurrence of the effect can be used as an indicator of expertise. If prior training or experience has included the critical problems, the *absence* of *Einstellung* can be used as an indicator of expertise. If the status of the critical problems is unambiguous, then the consequences of obtaining *Einstellung* are also unambiguous. (See Sweller & Gee, 1978, for a discussion of the properties of *Einstellung*.) Experiment 1 incorporates a design in which the extent to which problem solvers can distinguish between initial problems and critical *Einstellung* problems is used as a measure of expertise.

Method

Subjects. The subjects consisted of: (a) Twenty-two Year 9 students from the top two mathematics classes of a Sydney high school with six Year 9 mathematics classes—mathematics is a compulsory subject in Year 9; (b) Twenty-two Year 11 students undertaking an advanced level of mathematics—mathematics is not compulsory in Year 11 and more advanced or less advanced levels may be chosen; (c) Eighteen university students taking mathematics teacher education courses. These students had either completed a mathematics degree (5 students) or were currently enrolled in second- or third-year mathematics (13 students).

Procedure. All subjects were run individually and informed that they were to be given a series of memory and algebra questions. The memory questions were presented via a computer screen. There were eight algebraic statements. Each appeared on the screen for 5 seconds followed by 30 seconds in which subjects had to reproduce the statement as accurately as possible using a pencil and paper.

Four of the statements consisted of random symbols, for example, $b =)dc$ $(cb(\,)e + a$. The remaining four statements consisted of algebraic equations. There were two categories with each represented by two equations that were identical except for the variables. Examples of the two categories of equations are: $(c + d)(a + e) = (c + d)(b)$, and $a + (ab/a) + ae/c = d$. The order of presentation was two equations followed by two random symbol sets alternated until the eight statements had been presented. There were two sets of eight statements and half of the subjects were presented with one set while the other half were presented with the alternate set. After presentation of the algebra problems (described later), each subject was given an identical memory test using the statement set not used initially.

The *Einstellung* test followed the memory test. A preliminary problem ($a + b - c = d$, solve for a) was presented in order to ensure that all subjects (especially Year 9 students) had sufficient relevant information and skills to be able to appropriately manipulate equations. If subjects had difficulty on this task they were shown how the equation $3 + 4 = 7$ may be transformed into $3 = 7 - 4$ by subtracting 4 from both sides. They were then *re*-presented the algebraic equation and asked to solve it. A total of 16 problems was presented in the test proper. The first two problems were identical to the last two. Comparisons of performance on these problems before and after the intervening 12 practice problems allowed us to detect the *Einstellung* effect. The practice problems were all similar in structure to one or the other of the two categories of equations used in the memory questions (assuming the equations had to be solved for a). These two types were randomly ordered but this order was identical for each subject. Each type occurred equally frequently. The *Einstellung* problems were produced from these two problem types by inserting an addition sign in such a way as to maintain the overall visual cues of the question while making the previously used solution inapplicable. The two *Einstellung* questions were: $(a + b)(c + d) = (a + b) + (g)$, and $a + \frac{(a + b)}{a} - ef/g = c$. In the first instance it is no longer valid to divide both sides by $(a + b)$ and in the second instance it is no longer valid to cancel out the a/a in the parentheses.

Problem solvers were allowed 90 seconds on each of the two *Einstellung* control problems and if unable to solve them, were not given the correct solution. If the solution had not been obtained within 90 seconds for each of the next four problems (which consisted of two problems from each category), the correct solution was given by the experimenter. No feedback was given for any of the remaining problems.

Results and Discussion

Memory tests. We will consider the results of the memory tests first. The basic measure used for both equations and random characters was the longest correct string. For each memory trial, the characters written by the subject were compared with the original. Any continuous sequence of characters that matched, irrespective of their location in the string, constituted a correct string. The number of characters in the longest such string provided the basic score obtained on each memory trial.[1]

[1]Modifications of this basic measure were required and these differed for the equations and the random characters. In the case of *equations,* three variations from the basic measure (longest correct string) were allowed providing mathematical meaning was retained. The variations were: (1) Optional parentheses were counted as part of the string if included by subjects but did not destroy the string if omitted. With respect to the equation $(b + e)(a + c) = (b + e)(d)$, a response of $(b + a)(d) = (b + e)\ d$ would result in a score of 8. The count begins at the last parenthesis immediately to the left of the equal sign and includes the letter d on the right-hand side despite there being no parentheses around the d. The inclusion of these parentheses would increase the count to 10. All other arithmetic symbols had to match exactly. (2) Letters did not have to match providing mathematical meaning was retained. In the above example, c could be substituted by any other letter which did not appear anywhere else in the subject's equation. Substituting x for c [giving $(b + e)(a + x) = (b + e)(d)$] would not change the mathematical meaning and could be counted. Substituting b for c [giving $(b + e)(a + b) = (b + e)(d)$], on the other hand, would change meaning and a string containing such a substitution would not be

The means for the random characters and equations on both the initial and final memory tests are presented in Table 1. Separate analysis of random characters and equations were carried out due to the differences in measuring the string lengths. For the equations, analysis of variance using orthogonal comparisons indicated a significant difference between Year 9 students and the remaining students, $F(1,59) = 16.0$.[2] There was no difference between Year 11 and university students, $F(1,59) = 1.2$. These results indicate that the Year 9 students have shorter correct strings than the other two groups. There was a significant improvement from the first to the second memory test, $F(1,59) = 88.3$. A significant interaction between test and year, $F(1,59) = 5.2$, indicates more rapid improvement by the Year 9 students compared to the remaining students. There is no interaction between the two memory test scores and Year 11 and university students, $F(1,59) = .2$.

An identical analysis of memory for random characters yielded only one significant effect. There was an increase in length of strings from the first to the second memory test, $F(1,59) = 4.2$. There was no significant difference between either Year 9 students and the remaining students, $F(1,59) = 1.5$, or Year 11 and university students, $F(1,59) = 3.6$. The interactions between tests and years were also nonsignificant.

In summary, these results indicate that the Year 9 students had a poorer memory of the equations than the other students. This difference was reduced after practice at solving relevant equations. Year 11 and university students were essentially identical and, in contrast to Year 9 students, did not improve greatly after practice using the equations. These differences were specific to equations rather than representing general memory characteristics. There were no differences in memory of random characters other than a small improvement from the first to the second memory test.

Einstellung test. *Einstellung* is indicated in this experiment when operations appropriate to the practice problems were inappropriately used on the *Einstellung* problems coming after the practice problems but were not applied to those problems when they came before the practice problems. Specifically, in the case of the equation $(a + b)(c + d) = (a + b) + (g)$, divid-

counted. Similarly, substitutions could be made for $b + e$ providing they were identical on both sides of the equation. If they were not identical, both could not be counted in a single string because of the alteration in mathematical meaning. (3) The order in which multiplied or added variables were entered did not affect the string count. The strings $(d)(e + b)$ and $(b + e)(d)$ were both considered correct.

With respect to the *random characters,* the concept of mathematical meaning could not be used. For this reason, the only variation to the basic rule (longest running string) was that letters, unlike arithmetic symbols, could be randomly substituted. Any letter could appear any number of times in any location. For a string to be deemed correct, the only essential factor was that arithmetic symbols identical to those presented originally were separated by a correct number of letters.

[2]The .05 level of significance is used throughout this paper.

TABLE 1
Means of Longest Correctly Recalled String Length in Experiment 1

Group	N	Random Characters: Initial Presentation	Random Characters: Final Presentation	Equations: Initial Presentation	Equations: Final Presentation
Year 9	22	4.4	4.5	7.9	11.1
Year 11	22	4.5	5.6	10.3	12.1
University	18	4.4	4.6	10.9	13.0

ing both sides by $(a + b)$ indicated *Einstellung* while for the equation $a + \frac{(a + b)}{a} - ef/g = c$, cancelling a in the term $\frac{(a + b)}{a}$ similarly indicated the effect. Table 2 indicates the frequency with which the illegitimate operations were made on the initial and final presentations. Inspection of these frequencies indicates that on both problems most Year 9 students, many Year 11 students, but very few university students demonstrate *Einstellung*. Chi-square tests were used to test for differences in the frequency with which *Einstellung* was demonstrated on the two equations by the three age groups. (See criterion for *Einstellung* in previous discussion.) For the equation $a + \frac{(a + b)}{a} - ef/g = c$, $\chi^2(2) = 41.0$. For the equation $(a + b)(c + d) = (a + b) + (g)$, $\chi^2(2) = 25.9$.

In summary, increased expertise was associated with a superior memory of actual algebraic equations (but not random symbol strings) and an increased resistance to *Einstellung* effects in operating on equations. It may be of interest that the Year 11 students resembled the university students in memory of equations but were closer to the Year 9 students in accuracy of classification. We might expect students to first learn to identify equations and to subsequently learn which moves can be associated most appropriately with a particular equation. Our Year 11 students may have been transitional in that they may have adequately learned to identify the equations but not fully learned the associated algebraic manipulations.

These results suggest that more experienced problem solvers have a better representation of specific equation configurations and that their schemas link transformations more tightly to specific equation configurations. We may conclude that expertise in solving algebra manipulation problems requires knowledge of many specific schemas. This is in accord with results obtained in other problem domains.

EXPERIMENT 2

Experiment 1 provided evidence that expertise in solving algebra manipulation problems is, at least in part, schema based. The remaining experiments

TABLE 2

Number of Subjects Cancelling a in the term $\frac{a+b}{a}$ of Equation 1, and Cancelling $(a + b)$ in Equation 2 of Experiment 1

		Equation 1: $a + \frac{(a+b)}{a} - \frac{ef}{g} = c$		*Equation 2:* $(a+b)(c+d) = (a+b) + (g)$	
Group	*N*	*Initial Presentation*	*Final Presentation*	*Initial Presentation*	*Final Presentation*
Year 9	22	4	22	3	22
Year 11	22	2	17	3	12
University	18	0	1	0	4

described in this paper assume that schemas are an essential component of expertise and were designed to study some of the conditions we might expect, on theoretical grounds, to facilitate knowledge acquisition. If, as suggested previously, problem-solving search interferes with schema acquisition by directing attention away from patterns associated with given problem states, then procedures that direct attention more appropriately may be preferable. A heavy use of worked examples as a substitute for conventional problem solving may have the desired effect.

There are two possible advantages to the use of worked examples. Greeno (1980a) points out that domain-specific problem-solving strategies (equivalent to schemas in the present context) are not explicitly taught in a normal situation. They must be induced and this is not usually as efficient as step-by-step guidance in solution methods. Worked examples involve step-by-step guidance and this may provide one reason to expect the use of worked examples to be superior to conventional problem solving. In addition, detailed worked examples do not require the use of conventional problem solving search strategies. The cognitive processes involved in studying a worked example can be expected to be radically different from those required by goal-directed search. Rather than searching for operators that may reduce differences between the current problem state and the goal state, students studying a worked example may directly process the relation between given problem states and the moves required to transform these states into desired alternative states. Since schema acquisition requires (a) knowledge of problem states, (b) the operators that can be used when a given problem state has been attained, and (c) the consequences of using particular operators, we might expect schemas to be acquired more directly by a worked example approach as opposed to a conventional, goal-directed problem-solving search approach.

While the use of worked examples may potentially be more apt to result in schema acquisition than the solution of conventional problems, the two techniques may differ with respect to motivational factors. Problem solving is motivating probably because it requires activity (see Greeno, 1980a, for discussion). It is not possible to solve a problem without processing some of the information associated with it. While the information processed may not particularly assist in schema acquisition, it is possible to read a worked example and assimilate nothing if motivation is low. This problem was mitigated in Experiment 2 by alternating worked examples with structurally identical conventional problems. It was assumed that motivation, while reading a worked example, would be increased by the knowledge that a similar problem would need to be solved immediately afterwards.

In order to facilitate the acquisition of appropriate schemas, the equations used in Experiment 2 were simpler than those of Experiment 1. Since they were sufficiently simple to fall within the short-term memory span of the subjects, it was not appropriate to use memory tests as measures of expertise. Instead, time taken and number of errors made were the measures used. These permit an assessment of learning efficiency that is quite important in educational contexts.

Method

Subjects. The subjects were twenty Year 9 students from a second-level mathematics class of a Sydney high school. This class was chosen after extensive 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 a sheet of paper that contained a worked example of each of the two problem types used in the experiment. The sheet contained the following information:

There are two types of questions that you will be seeing. An example of each is given below.

1. For the equation $a = ag + b$, express a in terms of the other variables.

$$a = ag + b$$

$$a - ag = b$$

$$a(1 - g) = b$$

$$a = \frac{b}{1-g}$$

2. For the equation $\frac{b(a + c)}{e} = d$, express a in terms of the other variables.

$$\frac{b(a + c)}{e} = d$$

$$b(a + c) = ed$$

$$a + c = \frac{ed}{b}$$

$$a = \frac{ed}{b} - c$$

Subjects were asked if they had any questions and all questions were answered to the point where subjects claimed they understood the two example problems. This procedure was followed in order to ensure that the concepts, general procedures, and rules required to solve algebra transformation problems were understood by all subjects. An acquisition phase followed. Eight problems were presented. All problems had to be solved for *a*. Subjects were instructed to work as rapidly and as accurately as possible. If an incorrect answer was obtained, the problem had to be attempted again until it was solved correctly. Four of the eight problems were identical in format to the first problems presented on the worked example sheet. These problems differed in terms of the variables used, except that the variable *a* always occurred in a position identical to the example. The remaining four problems resembled the second problem on the worked example sheet. Two problems from the first category were presented followed by two problems from the second category. This double alternation procedure was followed until all eight problems had been presented.

There were two groups of 10 subjects each. The conventional problem group was simply required to solve the eight problems 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 the worked example 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. All previously solved problems and worked examples, including the two problems on the worked example sheet, were continuously available to subjects during this phase of the experiment.

Six test problems, identical for both groups, followed. These consisted of three problems each from the two categories of equations. They were identical in structure to the preceding problems. The two problem types were presented in alternating order. During this test phase of the experiment, subjects did not have access to previous work carried out during any phase of the experiment or to previously seen example problems. A maximum of 5 minutes 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 and the subject was asked to try the next problem. Within the 5-minute period subjects were informed of incorrect solutions and allowed additional attempts. These procedures were administered individually.

Results and Discussion

Completion time and number of mathematical errors during both the acquisition and test periods are of interest. Means are presented in Table 3. Due to the distorted 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, $U\,(10,10)$

TABLE 3
Mean Seconds and Errors Per Problem on Initial and Repeat Problem Presentation During Acquisition, and on Test Problems in Experiment 2

	Acquisition		
Group	*Initial Presentation*	*Repeat Presentation*	*Test*
Worked Example	27.5 (–)	50.5 (0.08)	105.4 (0.13)
Conventional Problem	73.3 (0.03)	42.0 (0.13)	127.1 (0.50)

Note: Mean errors appear in parentheses.

= 25. Inspection of Table 3 (which separates the acquisition period into initial and repeat problem presentations) indicates that this difference is caused directly by the initial example problems. There were very few errors during the acquisition period and the difference is not significant, U (10,10) = 48.5. There was no significant difference between groups in time spent on the test problems, U (10,10) = 47. With respect to errors, a Mann-Whitney U-test with a correction for ties indicated U (10,10) = 35, z = 1.28, p = .10.[3] Thus, the greater number of test errors for the conventional problem group may possibly represent a real difference.

We predict that heavier use of worked examples, by switching attention away from goal-directed search, should assist in schema acquisition which in turn should result in more rapid problem solution. We found no evidence for this. There is no difference between groups in time to solve the test problems, although there may be a real difference in number of errors in these problems. While these results do not support our prediction, they may nevertheless have practical significance. The worked example group spent less time on the acquisition problems with no discernible detriment to their subsequent problem-solving skill.

While our prediction was not supported, this may have been due to the simplicity of the materials used. The two equation formats and their solution may have been well-learned by both groups with little room for improvement. Experiment 3 was designed in part to rectify this possible problem.

EXPERIMENT 3

As was the case in Experiment 2, there were two groups—a conventional problem group and a worked example group. Five measures were taken in

[3]Throughout this paper, a correction for ties is only mentioned where its use may alter conclusions.

Experiment 3 to eliminate ceiling effects: (a) four rather than two categories of problems were used; (b) each problem type was seen twice rather than four times during the acquisition phase; (c) each problem type was presented once rather than three times during the test phase; (d) while working on any given problem during acquisition, no previously completed problems were accessible to problem solvers; and (e) the problems on the worked example sheet were merely similar rather than being identical in structure to those used in the remainder of the experiment.

Method

Subjects. The subjects were twenty-two Year 9 students from two Sydney high schools. As was the case in Experiment 2, students were chosen from classes that pilot studies indicated had had a sufficient background in algebra to allow them to complete the problem sets but not to exhibit a high degree of proficiency. Scope for improvement was thus available during acquisition.

Procedure. All subjects were presented with an initial example sheet containing three worked examples similar but not identical to the problem types to be presented subsequently. The three problems could be solved in 2, 3, and 5 steps respectively, with each step consisting of an algebraic manipulation required in the solution of the subsequently presented problems. The three worked examples, each solved for a, can be found in Table 4. All questions concerning these examples were answered. When subjects claimed that they understood the material, they were required to explain what the goal of the problem was (i.e., to make a the subject of the equation) and what mathematical operation was used to go from each line to the next. Assistance was given to any subject who had difficulty with these questions.

All subjects received identical acquisition problems in an identical order and were informed that they were to work on them as rapidly and as accurately as possible. The worked solution sheet was available at all times. Table 5 indicates the four types of questions used and the solutions that were shown to the worked example group. Each problem type was presented twice in succession. The variable a was the subject of the equation for each problem. This variable was placed in a location for repeated problems identical to its location in the preceding original (e.g., $a = b + ac$ and $a = x + ay$ are examples of an initial and repeated presentation problem). For the worked example group (11 subjects), the first presentation of each problem type had the solution given. Subjects were allowed to study each problem until they were satisfied that they understood it or until 5 minutes had elapsed. The conventional problem group (11 subjects) had to solve these problems themselves. A maximum of 5 minutes for all subjects was allowed on all conventional problems. If a problem had not been solved in this time frame, the solution was

TABLE 4
Problems on Initial Example Sheet in Experiment 2

$$\frac{ab}{e} = c$$
$$ab = ce$$
$$a = \frac{ce}{b}$$

$$a - ac + b = e$$
$$a - ac = e - b$$
$$a(1 - c) = e - b$$
$$a = \frac{e - b}{1 - c}$$

$$\frac{ab + ae + c}{d} = \frac{ag}{ah}$$
$$\frac{ab + ae + c}{d} = \frac{g}{h}$$
$$\frac{ab + ae}{d} = \frac{g}{h} - c$$
$$ab + ae = d\left(\frac{g}{h} - c\right)$$
$$a(b + e) = d\left(\frac{g}{h} - c\right)$$
$$a = \frac{d\left(\frac{g}{h} - c\right)}{b + e}$$

given by the experimenter. If any problem was completed with a mathematical error, it had to be repeated until it had been solved with no errors or until the 5-minute limit had elapsed. Previous solutions or worked examples other than those on the worked example sheet were not available at any stage.

Test problems were presented after the acquisition phase. Four problems identical in format to those of Table 5 were used. Neither the problems of the acquisition phase nor the worked example problem sheet was available during the test phase. A maximum of 5 minutes per problem was again allowed before the correct solution was presented by the experimenter, and problems had to be reattempted if earlier attempts contained mathematical errors within the 5-minute period.

The experimenter noted the time to solve each problem or the time spent on each worked example during both the acquisition and test phases of the experiment.

Results

Table 6 contains the mean seconds and errors for acquisition and test problems. The difference between the two groups in time spent in acquisition is significant, $U\,(11,11) = 4$, replicating the results of Experiment 2. This difference is due almost entirely to the very limited time spent by the worked example group in studying their worked examples. The conventional problem group has taken far longer to solve the equivalent initial problems.

The difference between groups in total time to solve the test problems is significant, $U\,(11,11) = 27.5$. Despite a vast reduction in time spent on the

TABLE 5
Problem Solutions Presented to the Worked Example Group During the Acquisition Phase of Experiment 3

$$\begin{aligned} \frac{c\,(a + d)}{f} &= g \\ c\,(a + d) &= fg \\ a + d &= \frac{fg}{c} \\ a &= \frac{fg}{c} - d \end{aligned}$$

$$\begin{aligned} a &= d + ac \\ a - ac &= d \\ a\,(1 - c) &= d \\ a &= \frac{d}{1 - c} \end{aligned}$$

$$\begin{aligned} c\,(a + b) &= \frac{af}{a} \\ c\,(a + b) &= f \\ a + b &= \frac{f}{c} \\ a &= \frac{f}{c} - b \end{aligned}$$

$$\begin{aligned} \frac{af + e}{b} &= c \\ af + e &= bc \\ af &= bc - e \\ a &= \frac{bc - e}{f} \end{aligned}$$

TABLE 6
Mean Seconds and Errors Per Problem on Initial and Repeat Problem Presentation During Acquisition, and on Test Problems in Experiment 3

	Acquisition		
Group	*Initial Presentation*	*Repeat Presentation*	*Test*
Worked Example	32.0 (–)	53.2 (0.45)	43.6 (0.18)
Conventional Problem	185.5 (2.73)	59.5 (0.36)	78.1 (1.64)

Note: Mean errors appear in parentheses.

acquisition phase, the worked example group required less time to solve the test problems.

While the use of worked examples obviously benefitted problem solvers, a more detailed analysis is required to locate the precise reasons for the reductions in time demonstrated by the worked example group. We will begin by examining the mathematical errors made. A mathematical error is defined as any algebraic transformation that violates the rules of algebra. While our students had been taught the relevant algebraic rules both in class prior to the experiment and during the experiment while studying the worked example sheet, their familiarity with these rules was not high and therefore, we can expect mathematical errors. Furthermore, unlike Experiment 2 in which problem solvers had access to a worked example sheet containing structurally identical examples and to previously solved problems, these were not available in Experiment 3.

As can be seen from Table 6, the conventional problem group made a relatively large number of mathematical errors on the first presentation of each problem type. These were virtually eliminated on the subsequent presentation. However, errors reappeared under the test conditions when the immediately preceding problem was not structurally identical. The difference in errors between the conventional and worked example groups on the test is statistically significant, $U\,(11,11) = 30$, and may partially account for the difference between groups in time to solve the test problems.

There is another distinction between the groups that may also contribute to differences in solution times and in addition is of considerable theoretical interest. Two of the four problem types used contained expressions of the form $c\,(a + b)$. While this expression can be expanded to give $ca + cb$, the expansion is not needed to solve either problem type. If the expansion is carried out, the solution is still attainable but is less efficient, requiring an additional move. On the first presentation of the two relevant problem types during the acquisition phase, the conventional problem group carried out the expansion a mean of .55 times per problem. On the repeated presentation of these problems, this group used expansion a mean of .32 times per problem. In contrast, the worked example group used expansion a mean of .09 times per repeated problem. By assigning subjects a score of 0, 1, or 2 depending on whether they used expansion on 0, 1, or 2 problems, we found that this difference is significant, $U(11,11) = 40$, $z = 1.72$. On the two test problems allowing expansion, the worked example and conventional problem groups used the expansion move a mean of .09 and .42 times, respectively. This difference is also significant, $U(11,11) = 39$, $z = 1.82$. The conventional problem group often used expansion; the worked example group almost never used it in both the acquisition and test phases of the experiment.

Discussion

As expected, our results indicated that while problem solvers required less time to study worked examples than to solve the equivalent problems, they subsequently also required less time to solve conventional test problems. The improvement on the test problems appears partly to be due to fewer mathematical errors and partly due to the use of fewer mathematically correct but unnecessary transformations.

We have theorized that conventional problem-solving search retards the acquisition of schemas, which are an essential component of skill in problem solving. In a semantically rich domain such as algebra, acquisition of schemas may be inextricably linked with an increasing facility in the use of general algebraic rules. A schema, by our definition, consists of a construct incorporating both a pattern and a move associated with that pattern. Each move reflects a rule. It may be difficult or impossible to obtain a complete knowledge of that rule until a large number of schemas incorporating it have been acquired. Our finding of differential error rates between worked example and conventional problem subjects may reflect this link. The conventional problem group made many errors which would not be made by experts familiar with the relevant rules. In contrast, the worked example subjects may have acquired a limited number of schemas allowing them to solve specific problems with few errors. While this is unlikely to have allowed them to develop a strong representation of the relevant rules, it may be reasonable to suggest that it is a necessary first step. If a large number of schemas associated with, for example, multiplying out a denominator are acquired, then this rule will be used flawlessly and we will conclude that the rule has been fully abstracted and learned. In summary, if general rules (i.e., "principles") are intimately linked to schemas, as our definition suggests, then the ability to use algebraic rules without error may gradually increment in conjunction with the acquisition of schemas. Neither schemas nor knowledge of rules can substantially develop apart from the other because a representation of a rule that allows its errorless use may need to be abstracted from a large number of the schemas that incorporate it. For this reason, procedures such as a heavier emphasis on worked examples, which are designed to facilitate schema acquisition, may also facilitate knowledge of rules.

Our finding that the conventional problem group made greater use of expansion is also of theoretical interest. Expansion was not demonstrated on the initially presented worked example sheet and was not required in order to solve any of the problems. At best it led to an inefficient solution. Under these circumstances, the use of expansion presumably reflects a schema previously acquired in the classroom. Furthermore, it is a schema, which in this instance, has resulted in *Einstellung*. According to our definition, *Einstellung* occurs when a previously acquired schema is inappropriately used because a problem is incorrectly perceived as belonging to a familiar category that requires the use of that particular schema. This definition is applicable to several of the students in the conventional problem group who, having previously learned that expansion can be used to solve some problems, used it unnecessarily. Presumably, this is due to their inability to distinguish between problems that require expansion and those that do not. The conventional problem group appears to have used a mixture of (a) previously acquired schemas when these were available and apparently applicable, and (b) problem-solving search when schemas were not available. Schema acquisition by this group during the acquisition phase seems to have been minimal. In contrast, the worked example group, having readily acquired appropriate schemas while studying the worked examples, did not attempt to use unnecessary techniques such as expansion.

EXPERIMENT 4

The decrease in the number of mathematical errors made by worked example groups in Experiment 3 suggests the possibility that transfer of skill might be generalizable to problems not closely related to the initial problems. This can only occur, of course, if the decrease in mathematical errors reflects an increased knowledge of mathematical principles. We suggested in our interpretation of Experiment 3 that the decreased error scores by worked example subjects may indicate that they have gained general proficiency in the use of algebraic rules. If this knowledge is independent of the specific schemas acquired on the initial problems, then we might expect transfer to new problems. If on the other hand, as suggested by the results of Experiment 1, expertise is closely tied to highly specific problem-state configurations, then we might expect that increased skill in solving one type of problem will not transfer strongly to other types. In the previous experiments, the test problems were identical to the initial problems except for alterations in the variables.

Experiment 4 tests the extent to which general rules are abstracted by including transfer problems, which differed in structure from the initial problems, but required similar algebraic manipulations.

Method

Subjects. Subjects consisted of forty Year 8 students from a Sydney high school. Pilot studies indicated that these students were comparable in performance on algebra manipulation problems to the Year 9 students from schools used in the previous experiments. In general, Year 9 students at this particular school could solve the problems rapidly and without errors. This left little scope for additional learning.

Procedure. The general procedure was similar to that used in Experiment 3. A minor difference was that subjects were reminded before studying each worked example that they should ensure that it was fully understood before covering it from view. In addition, after incorrectly completing a problem subjects had to cover the attempt from view before reattempting the problem.

Two kinds of test problems were used. Four problems were identical to those used in Experiment 3 and were thus similar to the acquisition problems. The other set of four problems was dissimilar to those in the acquisition set. Table 7 contains these latter problems. As can be seen in Table 7, the algebraic operations required are essentially identical to those of the initial problems (see Table 5) but the problem structures differ, resulting in differences in the order of required moves. These problems will be referred to as test-dissimilar problems while those of Table 5 will be referred to as test-similar problems.

Four groups of 10 subjects each were used in Experiment 4. Two worked example groups were treated identically during the acquisition phase as were the two conventional problem groups. As previously indicated, this treatment was very similar to that used in Experiment 3. Following acquisition, all groups were presented with the similar and dissimilar sets of test problems. The group names used below indicate whether worked examples or conventional problems were used during acquisition and whether the dissimilar test problems preceded or followed the similar test problems. The worked example-similar-dissimilar group was presented the worked example initial problems (alternated with conventional problems—see Experiments 2 and 3), followed by the similar test problems, and then by the dissimilar test problems. The worked example-dissimilar-similar group was identical except it had the order of the two sets of test problems reversed, with the dissimilar problems preceding the similar set. The same relation was true for the conventional problem-similar-dissimilar group (the similar problems preceded

TABLE 7
Test Problems Dissimilar to Acquisition Problems in Experiment 4

$$(a + b)\,e = \frac{afg}{a}$$
$$(a + b)\,e = fg$$
$$a + b = \frac{fg}{e}$$
$$a = \frac{fg}{e} - b$$

$$\frac{a + b}{k} - d = c$$
$$\frac{a + b}{k} = c + d$$
$$a + b = k\,(c + d)$$
$$a = k\,(c + d) - b$$

$$\frac{a\,(a + f)}{a} + b = g$$
$$a + f + b = g$$
$$a + f = g - b$$
$$a = g - b - f$$

$$ab - a - g = e$$
$$ab - a = e + g$$
$$a\,(b - 1) = e + g$$
$$a = \frac{e + g}{b - 1}$$

the dissimilar set) and conventional problem-dissimilar-similar groups (the dissimilar problems preceded the similar set).

Results

Table 8 indicates mean scores for the four groups on each of the problem sets. For purposes of analysis, the scores of the two conventional problem groups on the acquisition problems were combined (because they were treated identically during acquisition), as were the scores of the two worked example groups. In time to complete the acquisition problems, there was again very little overlap between the twenty subjects given conventional problems and the twenty presented with worked examples. Using a Mann-

Whitney U-test, U (20,20) = 11. Far more time was required by the conventional problem group.

The conventional problem-similar-dissimilar and worked example-similar-dissimilar groups may be used to check whether the test results of Experiment 3 have been replicated. The worked example group required significantly less time on the similar test problems than did the conventional problem group, U (10,10) = 27. No difference between these two groups was obtained on the dissimilar test problems which followed, U (10,10) = 41. With respect to the conventional problem-dissimilar-similar group and the worked example-dissimilar-similar group, there was no significant difference on either the dissimilar test problems, U (10,10) = 40 or the similar test problems, U (10,10) = 38. It might also be noted that over all four groups, using a Wilcoxon signed rank test, significantly more time was spent on the dissimilar test problems than on the similar test problems, T (20) = 129, z = 3.78.

An identical pattern of results was obtained by analyzing the algebraic error data (see Table 8 for means). A significant difference was obtained between the conventional problem-similar-dissimilar and worked example-similar-dissimilar groups on the similar test problems, U (10,10) = 28 and with a correction for ties, z = 1.73. These two groups did not differ significantly on the dissimilar test problems, U (10,10) = 48.5. The conventional

TABLE 8
Mean Seconds and Errors Per Problem in Initial and Repeat Problem Presentation During Acquisition, and on Similar and Dissimilar Test Problems in Experiment 4

	Acquisition		*Test*	
Group	*Initial Presentation*	*Repeat Presentation*	*Similar*	*Dissimilar*
Worked Example Similar-Dissimilar	48.5 (–)	62.6 (0.20)	56.3 (0.25)	103.6 (0.63)
Worked Example Dissimilar-Similar			63.8 (0.23)	136.2 (0.80)
Conventional Problem Similar-Dissimilar	207.6 (0.94)	59.8 (0.20)	81.1 (0.43)	87.9 (0.43)
Conventional Problem Dissimilar-Similar			87.3 (0.53)	159.9 (1.23)

Note: Mean errors appear in parentheses. Acquisition means are based on a combination of both worked example groups and a combination of both conventional problem groups.

problem-dissimilar-similar group and worked example-dissimilar-similar group differed on neither the dissimilar set of problems, $U\,(10,10) = 34.5$ nor on the similar set of problems, $U\,(10,10) = 49$. Over all groups, a Wilcoxon signed rank test significantly indicated that more errors were made on dissimilar test problems than on similar test problems, $T\,(20) = 79$, $z = 3.60$.

As was the case in Experiment 3, expansion is an unnecessary algebraic manipulation. It is mathematically correct but strategically inappropriate because time to solution is extended by its use. Table 9 indicates the mean number of times each group used expansion in the two acquisition problems allowing this step (see Table 5). As can be seen from Table 9, expansion was used frequently on these two problems by the conventional problem groups. While no comparison with the worked example groups on these problems is possible, a valid comparison can be made between the two worked example and two conventional problem groups on the repeated presentation of these problems. Assigning subjects scores of 0, 1, or 2 depending on the number of problems on which expansion was used, this difference is significant, $U\,(20,20) = 126$.

Table 9 also indicates the mean number of times each group used expansion on the test problems. The difference between the worked example-similar-dissimilar and conventional problem-similar-dissimilar groups on the similar test problems is not significant, $U\,(10,10) = 40$, thus failing to replicate the results of Experiment 3 (although differences are in the pre-

TABLE 9
Mean Frequency with which Expansion Was Used in Experiment 4

	Acquisition		*Test*	
Group	*Initial Presentation*	*Repeat Presentation*	*Similar*	*Dissimilar*
Worked Example Similar-Dissimilar	(–)	.05	.1	.05
Worked Example Dissimilar-Similar			.3	.05
Conventional Problem Similar-Dissimilar	.675	.325	.3	.1
Conventional Problem Dissimilar-Similar			.75	.35

Note: Acquisition mean frequencies are based on a combination of both worked example groups and a combination of both conventional problem groups.

dicted direction). On the dissimilar problems, expansion was rare for both groups (see Table 9). Identical comparisons on the worked example-dissimilar-similar and conventional problem-dissimilar-similar groups both yielded significant effects. On the dissimilar problems, $U\,(10,10) = 24.5$, while on the similar problems, $U\,(10,10) = 26.5$. It is not clear why a similar-dissimilar order should yield nonsignificant differences while the reverse sequence results in significance. It might be noted that considering both the acquisition and test phases of Experiments 3 and 4, the conventional problem groups used expansion significantly more frequently than did the worked example groups in all cases except when the comparison involved the conventional problem-similar-dissimilar and worked example-similar-dissimilar groups.

Discussion

The major general finding of interest in Experiment 4 is that while worked examples are of assistance to students when faced with similar problems, the advantage does not extend to dissimilar problems, which nevertheless require similar algebraic manipulations. Similar problems were solved rapidly and with fewer errors when they immediately followed upon acquisition problems that included worked examples. Differences were not obtained using dissimilar problems although there was a suggestion that reduced use of inappropriate expansion occurred on dissimilar problems after exposure to worked examples. It might be noted also that the advantage of worked examples on subsequent similar problems were reduced when dissimilar problems were interpolated between the initial and similar test problems. In total, the results strongly confirm the specificity of schemas used in knowledge-based problem solving but also suggest that under the conditions of Experiment 4 the schemas acquired were easily forgotten.

The error data are of particular theoretical interest. Firstly, the identical pattern of results obtained for both the error data and time to solution suggest that solution times are very heavily influenced by the number of errors and less heavily influenced by factors such as the use of expansion. Secondly, from a theoretical point of view, the error data suggest that novice problem solvers have considerable difficulty abstracting general rules from the specific schemas that incorporate them. Unless abstraction has occurred, attempts to use the rule under conditions where acquired schemas cannot be applied may result in errors. We might expect that the acquisition of an adequate cognitive representation of a rule involving considerable abstraction is a necessary step in attaining a high level of expertise. Until this level of expertise has been attained, general rules may only be associated with a limited number of schemas. Errors may be more likely to occur where these schemas cannot be used.

EXPERIMENT 5

While Experiments 3 and 4 have indicated that increased use of example problems can facilitate subsequent problem-solving performance, the largest differences occurred during the acquisition stage. The use of example problems requires far less time than conventional problem solving. If the time spent on acquisition was equalized between example problem and conventional problem groups, then the amount of practice given to an example problem group should increase substantially in terms of the number of problems presented. This should increase differences on subsequent problems. Since presentation of the test problems to the two groups in a dissimilar-similar sequence did not yield differences in time or errors on either the dissimilar or similar problems of Experiment 4, this presentation order was used again in Experiment 5. Rather than being equated with respect to number of problems initially presented, subjects were equated with respect to time spent on the initial problems.

Method

Subjects. The subjects were twenty-four Year 8 students from a Sydney high school with eight Year 8 mathematics classes. The top three classes were considered by the school to be of parallel ability. The top eight students from each of these three classes were used in the experiment.

Procedure. Twelve subjects each were assigned to conventional problem-dissimilar-similar and worked example-dissimilar-similar groups analogous to those of Experiment 4.

The procedure was identical to that of Experiment 4 except that subjects were matched with respect to the time spent on the acquisition phase. Subjects were paired according to ability based on class tests and then assigned to one of the two conditions. The subject from each pair assigned to the conventional problem was run first. These subjects were presented the eight acquisition problems used in Experiment 4. The paired subject from the worked example group was then allocated an acquisition time identical to that used by the conventional problem subject. A sufficient number of acquisition problems were constructed to ensure that worked example subjects did not exhaust the problem pool during the time allowed. In order to do this, the four problem categories used in Experiments 3 and 4 were duplicated several times with altered variables. The four worked example-conventional problem pairs from a set were all presented before beginning problems from a new set. It was thus possible for worked example subjects to complete several of these sets in the time taken by conventional problem subjects to complete one set. The similar and dissimilar test problems were identical to those of Experiment 4.

Results and Discussion

Based on our previous results, we might expect the worked example group to process more problems during acquisition than the conventional problem group. The worked example group was presented a mean of 24.92 problems to study or solve compared to the 8 problems solved by each of the conventional problem subjects during the same time period. Eleven of the 12 worked example subjects processed more than eight problems. The remaining subject ran out of time on the eighth problem.

Table 10 indicates mean times and errors per problem for each phase of the experiment. The differing number of problems presented to the two groups during the acquisition phase have not affected time to solve the dissimilar test problems that immediately followed the acquisition phase. According to a Wilcoxon matched pairs signed ranks test, $T\,(12) = 31$, which is not significant. (Subjects were matched according to time spent on the acquisition phase. The figure in parentheses following the T indicates the number of matched pairs whose members received differing scores. Pairs whose members received the same score are ignored by the Wilcoxon test.) On the similar problem set, which followed the dissimilar problems, significantly less time was required by the worked example group than by the conventional problem group, $T\,(12) = 8$. Significantly more time was spent on the dissimilar than on the similar problems, $T\,(24) = 0$.

As was the case in Experiment 4, this pattern of results was repeated using mathematical errors as the dependent variable. There was no significant difference between groups on the dissimilar problems, $T\,(10) = 20$, but there was a significant difference on the similar problems, $T\,(7) = 3$. With respect to expansion (see Table 11), significantly more conventional problem subjects used expansion on the dissimilar problems compared to the worked example group, $T\,(5) = 0$. On the similar problems, expansion was used rarely and the difference was not significant, $T\,(3) = 0$. More errors were made on the dissimilar than on the similar problems, $T\,(20) = 0$.

TABLE 10
Mean Seconds and Errors Per Problem in Initial and Repeat Problem Presentation During Acquisition, and on Similar and Dissimilar Test Problems in Experiment 5

	Acquisition		*Test*	
Group	*Initial Presentation*	*Repeat Presentation*	*Similar*	*Dissimilar*
Worked Example	57.9 (–)	55.8 (0.12)	52.4 (0.13)	163.7 (1.06)
Conventional Problem	232.2 (0.96)	66.5 (0.08)	101.1 (0.58)	183.3 (0.88)

Note: Mean errors appear in parentheses.

TABLE 11
Mean Frequencies with which Expansion Was Used in Experiment 6

	Acquisition		*Test*	
Group	*Initial Presentation*	*Repeat Presentation*	*Similar*	*Dissimilar*
Worked Example	–	.04	.08	0
Conventional Problem	.67	.33	.29	.33

Equating subjects on time spent on acquisition rather than number of problems presented, has increased the size of effects on similar problems compared to those found in Experiment 4. Despite the interpolation of dissimilar problems between the acquisition phase and the similar problems, less time was required and fewer errors were made by the worked example group than the conventional problem group on the similar problems. Using these measures, differences were not obtained on the dissimilar problems, replicating the results of Experiment 4. The only difference on these problems was with respect to the use of expansion. The conventional problem group used expansion more frequently than the worked example group. This also replicates the results of Experiment 4.

We may conclude that while the use of worked examples facilitated the acquisition of schemas, these were only beneficial on a restricted range of problems. Nevertheless, there also appears to have been a change in strategy due to schema strength, which had a somewhat more general effect. After studying worked examples that did not demonstrate expansion, its use was reduced relative to subjects presented conventional problems and this effect occurred on both similar and dissimilar problems.

GENERAL DISCUSSION

Our results suggest that schemas required for efficient problem solving are acquired relatively slowly during practice on conventional problems. This can be explained by assuming that conventional, goal-directed problem-solving strategies may not direct attention to those aspects of a problem conducive to schema acquisition. The use of worked example problems may redirect attention away from the problem goal and toward problem-state configurations and their associated moves. In providing evidence for this conclusion, our findings also suggest that problem-solving skill may require many schemas, each limited to a narrow set of problems.

Any practical implications that may flow from these findings will depend on the extent to which our results can be generalized to the extended learning environments found in educational contexts. While this must be tested directly, it can be hypothesized that increased acquisition periods will generally increase the strength and number of schemas acquired and that this increase will be magnified by the use of worked examples. If this hypothesis is valid, the differences found here in favor of practice on worked examples over practice on conventional problem solving should be retained over extended acquisition periods. In consequence, we would like to argue that the usual emphasis on conventional problem solving in educational settings could be misplaced. Alternative techniques such as a heavy reliance on worked examples may be preferable. While worked examples are commonly employed until students are assumed to have obtained a basic familiarity with new material, the procedure is normally abandoned beyond this point to be replaced by conventional problems. It may be beneficial to persist with examples until complete familiarity with the material is attained. The acquisition of both appropriate schemas and an abstract representation of algebraic rules may be required for a high degree of proficiency. Expertise of this order tends to require protracted study. Heavier use of worked examples could reduce this period.

There are several possible criticisms of these conclusions that need to be considered. Firstly, it may be argued that if expertise really requires the acquisition of many highly specific schemas, expertise in large domains could never be attained. A worked example for every conceivable problem would require too many examples. We suggest that Experiment 1 provides evidence that experts do have cognitive representations of many problem types and these representations need to be acquired whether or not worked examples are used. The number of worked examples may not, in fact, exceed the number of conventional problems most students are expected to solve in the normal course of events, but require far less time to process for an equivalent or perhaps superior educational outcome.

We have provided evidence that learning may be retarded by problem-solving search and have suggested that worked examples may provide a good substitute by teaching students how to solve specific categories of problems. It might be argued that this eliminates "real" problem solving because problem solvers are learning how to solve particular problems rather than acquiring general problem-solving skills. On this point we agree with Greeno (1980b) who pointed out that students are frequently not credited with problem-solving skill if they have already learned how to solve particular types of problems. Yet, if problem-solving skill is domain-specific as is suggested by the literature on novice-expert distinctions (see Chi, Glaser, & Rees, 1982), then learning how to solve particular types of problems may be the only skill that can be acquired. In any case, it needs to be noted that novel problems, not covered by previously seen worked examples, must inevitably be encountered by students irrespective of teaching techniques. Problem-solving search strategies will still be needed to solve these problems.

Despite the aforementioned, the claim that students will lose something essential by a reduction in time spent on an activity seen frequently as being very valuable cannot be discounted until more evidence has been collected. Long-term studies using a variety of materials are needed before our suggestions can be considered totally convincing. Nevertheless, there appears to be no strong evidence that conventional problem-solving activity is generally beneficial other than in its side effect of schema development. There is no evidence, for example, that superior problem-solving strategies generalizable to a variety of problems are acquired through practice at problem solving. It should in any case be emphasized that our suggestions concerning the use of worked examples are intended to improve problem-solving skill, which must be tested in the conventional manner by the presentation of new problems to be solved by the student. Problem solving can provide both teachers and students with information concerning which aspects of an area have been fully learned and understood, and which aspects require further study. Its primary function may be in guiding rather than in accelerating learning.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Australian Research Grants Scheme to the first author.

The assistance of the New South Wales Department of Education and the following mathematics masters is gratefully acknowledged: W. Akhurst, Narrabeen High School; R. Clarke, Mosman High School; C. Hannon, Mackellar Girls' High School; C. Kerie, Freshwater High School; W. Masters, Forest High School; J. Smith, Manly High School; T. Wells, Balgowlah Boys' High School.

REFERENCES

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

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: Lawrence Erlbaum Associates.

De Groot, A. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving* (pp. 19–50). New York: Wiley.

Greeno, J. (1980a). Some examples of cognitive task analysis with instructional implications. In R. Snow, P. Federico, & W. Montague (Eds.), *Aptitude, learning and instruction, Vol. 2: Cognitive process analyses of learning and problem solving* (pp. 1-21). Hillsdale, NJ: Lawrence Erlbaum Associates.

Greeno, J. (1980b). Trends in the theory of knowledge for problem solving. In D. Tuma & F. Reif (Eds.), *Problem solving and education: Issues in teaching and research* (pp. 9–23). Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980a). Expert and novice performance in solving physics problems. *Science, 208,* 1335–1342.

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

Lewis, C. (1981). Skill in algebra. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 85–110). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Shulman, L. S., & Keislar, E. R. (Eds.). (1966). *Learning by discovery: A critical appraisal.* Chicago: Rand-McNally.

Simon, D., & Simon, H. (1978). Individual differences in solving physics problems. In R. Siegler, (Ed.), *Children's thinking: What develops?* (pp. 325–348). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Sweller, J., & Gee, W. (1978). Einstellung, the sequence effect and hypothesis theory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 513–526.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-end 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,* 639–661.