Title
Role of rules in transfer of mathematical word problems

Permalink
https://escholarship.org/uc/item/71r17249

Author
Kanevsky, Inna Glaz

Publication Date
2006

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Role of Rules in Transfer of Mathematical Word Problems

A dissertation submitted in partial satisfaction of the requirement for the degree

Doctor of Philosophy

in

Psychology

by

Inna Glaz Kanevsky

Committee in Charge:

Professor Edmund Fantino, Chair
Professor Michael Cole
Professor Michael Fogler
Professor Gail Heyman
Professor John Wixted

2006
The Dissertation of Inna Glaz Kanevsky is approved, and it is acceptable in quality and form for publication on microfilm:

__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________

Chair

University of California, San Diego

2006
DEDICATION

For Alex, Mama Ira, and Mama Gala – for pushing, prodding, nagging, picking up the slack, and simply making this happen.

With special thanks to David and Elena Andressen and Soraya Newell, without whose invaluable help I would not be where I am.
# TABLE OF CONTENTS

Signature Page ................................................................. iii
Dedication ................................................................. iv
Table of Contents .......................................................... v
List of Figures ............................................................... vii
List of Tables .............................................................. viii
Acknowledgements ................................................. ix
Vita ........................................................................ x
Abstract ................................................................. xi

Chapter I.  Introduction ......................................................... 1

Chapter II.  Experiment 1 ....................................................... 13
  Method ................................................................... 14
  Results and Discussion ................................................. 16

Chapter III.  Experiment 2 ..................................................... 22
  Method ................................................................... 22
  Results and Discussion ................................................ 22

Chapter IV.  Experiment 3 ..................................................... 26
  Method ................................................................... 26
  Results and Discussion ................................................ 28

Chapter V.  Experiment 4 ..................................................... 32
  Method ................................................................... 33
  Results and Discussion ................................................ 38

Chapter VI.  Experiment 5 ..................................................... 55
LIST OF FIGURES

Figure 1: Average number of practice problems attempted before reaching the criterion of three consecutive problems solved correctly in Experiment 1.............. 17

Figure 2: Average percentage of transfer problems solved correctly in Experiment 1, out of six possible problems................................................................. 18

Figure 3: Average percentage of students who solved at least 4 out of 6 transfer problems correctly in Experiment 1.................................................................19

Figure 4: Average percentage of transfer problems solved correctly in each of the two three-problem consecutive segments in Experiment 1................................. 20

Figure 5: Average number of practice problems attempted before reaching the criterion of three consecutive problems solved correctly in Experiment 2......... 23

Figure 6: Average percentage of students who solved at least 4 out of 6 transfer problems correctly in Experiment 2.................................................................24

Figure 7: Average percentage of students who solved all of their attempted transfer problems correctly in Experiment 3.............................................................. 30

Figure 8: Average number of practice problems attempted before reaching the criterion of three consecutive problems solved correctly in Experiment 4......... 38

Figure 9: Average percentage of students who solved all of their attempted transfer problems correctly in Experiment 4..........................................................39

Figure 10: Average percentage of students who solved all of their attempted transfer problems correctly in Experiment 5......................................................... 58
LIST OF TABLES

Table 1: Distribution of answers to questionnaires in Experiment 4.........................41
Table 2: Intercorrelations for the Experiment 4 variables.........................................44
Table 3: Distribution of answers to questionnaires in Experiment 5.........................60
Table 4: Intercorrelations for the Experiment 5 variables.......................................61
ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Edmund Fantino, for his support, advice, and patience. Also, I thank Dr. Stephanie Stolarz-Fantino for her collaboration on parts of this project and her valuable advice.

This research was made possible through participation of teachers and students at the Preuss School, Washington Middle School, Del Mar Hills Academy, and Calle Mayor Middle school, and with assistance of the operant lab’s graduate and undergraduate students. Thank you!
VITA

1993 B. S. in Mathematical Instruction
State Teachers Institute, Nikolayev, Ukraine

1995-1997 Research Assistant, Department of Psychology, California State University, Los Angeles

1998 M. S. in Applied Behavior Analysis
California State University, Los Angeles

1998-2003 Teaching Assistant, Department of Psychology, University of California, San Diego

2002-2005 Adjunct Instructor, Department of Psychology, Grossmont College, San Diego

2005-present Assistant Professor, Psychology, Department of Behavioral Sciences, San Diego Mesa College, San Diego

2006 Ph. D. in Psychology, Experimental Behavior Analysis
University of California, San Diego

FIELDS OF STUDY

Major Field: Behavior Analysis

  Studies in Verbal Behavior
  Professor Barry Lowenkron

  Studies in Choice (Base-Rate Fallacy)
  Professor Edmund Fantino

  Studies in Transfer of Problem-Solving
  Professor Edmund Fantino
ABSTRACT OF THE DISSERTATION

Role of Rules in Transfer of Mathematical Word Problems

by

Inna Glaz Kanevsky

Doctor of Philosophy in Psychology

University of California, San Diego

Professor Edmund Fantino, Chair

Mathematical word problems are of interest to educators because of their importance in the curriculum, and to psychologists because of their value as a context for the study of transfer of problem solving. A classic issue in transfer of problem solving questions whether or not rule learning leads to inflexibility when the learned rule is no longer appropriate (e.g., Luchins, 1942). The present set of studies investigated the differences in success rates on a transfer test and rule use between schoolchildren who were either given direct instruction with a rule different from the transfer test rule, or had an opportunity to discover that different (base) rule through contingency-based practice. The studies were conducted directly in the classrooms and with the classroom teachers providing the instruction in order to maximize their ecological validity and offer an immediate educational application. Overall, it was shown that although the instruction group students always acquired the base rule faster than the contingency group students, the latter usually performed better on the transfer test. Thus, contingency-based learning had a transfer advantage over the rule instruction.
In Experiments 4 and 5 subjects’ understanding of the problems’ structure, rule use, and rule inference were assessed through questionnaires and, in Experiment 5, through the use of irrelevant information in the problems. Independently inferring solution-based rules during practice enhanced transfer test performance which required different rules. Also, subjects who inferred solution rules in the process of taking the transfer test were the subjects who did best on that test. Subjects who focused on the problems’ context instead of the solution rules did not perform well. Therefore, problem-solving involving rules is not necessarily inflexible and may instead enhance the transfer performance as long as the base rules are not learned by rote through direct instruction. Contingency-based practice which allows the participants to infer rules can be recommended as a way of teaching for transfer of solving mathematical word problems.
Chapter I

Introduction

Why word problems? I hope that I shall shock a few people in asserting that the most important single task of mathematical instruction in the secondary schools is to teach the setting up of equations to solve word problems. Yet there is a strong argument in favor of this opinion.

In solving a word problem by setting up equations, the student translates a real situation into mathematical terms: he has an opportunity to experience that mathematical concepts may be related to realities, but such relations must be carefully worked out. (Polya, 1962, p. 59)

George Polya, a renowned mathematician and the author of several books on problem solving, wrote this statement over forty years ago, but his reasoning still rings true with mathematics educators today. However, while word problems are incorporated in mathematics curriculum and students engage in problem solving on regular basis, they are having considerable difficulties, and often cringe at a mention of word problems (G. Cerelli, personal communication, March 2001; T. Parsons, personal communication, May 2002; Kanevsky, Stolarz-Fantino, & Fantino, 2001; Stacey & McGregor, 2000). Research in mathematics education that concerns itself with problem-solving often focuses on things like prerequisite knowledge or problem classifications, though some scientists investigate more broadly applicable factors, such as analogical transfer or instruction methods.

On the other hand, there has been much research on problem solving in cognitive and educational psychology, though only a small portion of it with typical word problems. A look at the literature reveals that psychologists tend to study more abstract mathematical or puzzle problems, such as Luchins’ (1942) water jar problems
or Dunker’s (1945) radiation problem. This incongruity might have contributed to the lack of connection between the work of psychologists and educators lamented by Stephen Reed (1999) who expressed a concern that his and other cognitive psychologists’ research had little influence on education, and vice versa. Though some publications (e.g., Bernardo, 2001; Berliner & Calfee, 1996; English, 1997) include contributions from both fields, further psychological investigations into problem solving that focus on mathematical word problems in the classroom instead of in the laboratory could be of use to educators as well.

Much of the up to date research on transfer of problem solving focused on analogical problem solving (e.g., Chen, 1999; English, 1998; Gick & Holyoak, 1980, 1983; Reed, Dempster, & Ettinger, 1985). In typical problem-solving experiments using analogical reasoning, participants are required to solve new, or target, problems by analogy after solving or learning about model, or base, problems (Goswami, 1992). The usual questions of interest are whether the similarities between the two problems will be noticed and whether they then will be used to help solve the target problem. Most investigations of analogical problem solving use base and target problems that are isomorphic or similar. According to Reed’s (1987) commonly accepted classification, problems that have different story contexts but can be solved in the same way are isomorphic, and problems that share a story context but require some modification of the solution for original, or base, problem to solve the target problem are similar. While isomorphic problems share structural features, such as the mathematical relations of their elements, similar problems share surface features, such as characters in the story (Mayer & Wittrock, 1996; Reed, 1987).
Most researchers found that positive transfer effects in analogical problem solving rarely occur spontaneously. For example, Gick and Holyoak (1980) presented their subjects with a story about a general who had to take a fortress with many roads radiating from it. Since the roads were all mined to prevent the passage of a large number of troops, the general dispersed his army into small units, which then converged onto the fortress at the same time. After studying this story, the subjects were asked to solve Dunker’s (1945) classic problem where a tumor has to be destroyed with radiation without destroying surrounding healthy tissue. Even though the two situations were structurally identical, a great majority of subjects failed to find the analogous solution unless given a hint, and many failed to come up with a dispersal solution even then. In their further research, Gick and Holyoak (1983) found that even adding a general solution statement at the end of the military story did not help subjects to notice the analogy better than when they were given the story without the solution statement.

Reed, Dempster, and Ettinger (1985) obtained similar findings using algebra word problems. They found that experience with problems similar to the target problems does not benefit students, while experience with problems equivalent to the base problems, with both structural and superficial similarities (Reed, 1987), does offer a positive transfer effect. The findings of Reed et al. are complemented by those of English (1998), who established that children did not benefit from the experience with isomorphic problems. In addition, English found that most children in that study did not even see a relationship between the isomorphic arithmetic word problems used in the study.
Clearly, experience with solutions to analogous (isomorphic or similar) problems does not necessarily lead to an improved ability to solve subsequent problems. In such situations, transfer is sometimes improved by using multiple analogies (Gick & Holyoak, 1983) or supplying hints or prompts to use the analogy (Catrambone & Holyoak, 1989; Gick & Holyoak, 1980, 1983). Thus far, the research presents a rather pessimistic view on training in problem solving, if even close analogies do not seem be of great help. Some (Novick, 1988; Silver, 1981; Weaver & Kintsch, 1992; Goswami, 1992) have suggested that this might be due to individuals’ focusing on the superficial similarities such as story context in the base and target problems. These superficial similarities would then obscure the structural similarities that would allow for use of the base problem’s solution in solving the target problem.

Understanding of structural instead of only superficial similarities between the base problem and the target problem can affect transfer on two levels: mapping and abstracting the solution. For mathematical word problems, mapping is a process of finding corresponding quantities in the base and target problems (Reed, 1999). If structural similarities are indeed more important than surface features, then an ability to overlook the story and find these corresponding quantities in the base and target problem should contribute to successful transfer. Indeed, Silver (1981) established that good problem solvers classify algebra word problems on the basis of mathematical structure, while poor problem solvers classify algebra word problems based on the story context. Reed (1987) also demonstrated that college students are significantly better at constructing equations for isomorphic problems than for similar problems, noticing and using the structural similarities more than the superficial similarities.
While successful mapping of structural features appears to be essential for analogical transfer, it is not sufficient. Novick and Holyoak (1991) argued that the ability of students to abstract solutions from base to target problems is most important in analogical transfer. Indeed, when they provided college students with already-mapped problems, the students were still largely unsuccessful at adapting the solution of the base problem to the target problem. Novick and Holyoak also found that students who inferred correct and complete solution rules, or schemas, from the base problem were the ones who demonstrated positive transfer effects.

The latter finding of Novick and Holyoak (1991) can be also taken as evidence that inferring rules from base problems should be beneficial for future problem solving. This conclusion may be correct for isomorphic problems, but the findings with similar and unrelated problems appear to be different. A classic experiment by Luchins (1942) introduced the notion of rigidity in rule-based problem solving, also known as the Einstellung effect. This effect occurred when subjects who practiced with a set of problems that were all solved by the same rule then continued to apply that rule even when it was no longer efficient for solving similar problems. Luchins’ subjects inferred a rule for solving a series of problems requiring them to get a set amount of water using three water jars of predetermined size, and continued to use that rule with novel problems that could better be solved by a shortcut instead.

While this could be interpreted as a positive transfer effect as far as analogical problem solving is concerned, it certainly is not beneficial. Such rigidity would be even more detrimental if the old rule no longer applied at all. In one of the conditions of Luchins and Luchins’ (1950) study, the subjects who had prior experience with a
rule persisted in using it on a problem that could not be solved by that rule. Likewise, Sweller, Mawer, and Howe (1982) found that subjects who were encouraged to use a rule while solving a series of increasingly complex problems experienced difficulty when faced with an unrelated simple problem, in marked contrast to subjects who practiced with the same series but without using rules. Though this finding agrees with Novick and Holyoak’s (1991) findings on difficulty of spontaneous rule induction, it puts in question the benefits of inducing subjects to infer rules in problem solving if the desired transfer is to problems that are neither equivalent nor isomorphic to the base problems.

However, not all rules may impede problem solving. Fantino, Jaworski, Case, and Stolarz-Fantino (2003) challenged Luchins’ (1942) notion of rigidity of rule-based problem solving. They used water jar problems resembling those used by Luchins, but unlike Luchins they made sure that there was a clear distinction between the training and transfer portions of the studies. In addition to examining the effects on transfer of self-generated rules, as in Luchins’ or Sweller et al.’s (1982) work, Fantino et al. examined the effects on transfer of rules provided by experimenters. In this case, the Einstellung effect was very short-lived for the students who induced the rule on their own, and non-existent for the students who were instructed with the rule. These results, however, could be attributed to procedural differences between Fantino et al. and Luchins’ experiments that suggest influence of problem solving context on transfer, similar to that observed by Catrambone and Holyoak (1989) and Spencer and Weisberg (1986) in analogical problem solving. That is, while instructed subjects learned the rule for the set of base problems, they did not expect the same rule to apply
when the set of target problems was presented. Instead, they might have learned to look for rules when faced with problems of that type, which could also explain their superior performance when compared to control subjects.

This notion of the beneficial effect of learning to look for rules corresponds with findings of Chen (1999), who demonstrated that children who devised abstract rules after practice with a series of problems performed better on transfer problems than children who devised specific rules. Children in Chen’s experiment had to practice solving problems equivalent or similar to Luchins’ (1942) water-jar problems, with the experimenter providing correct solutions if children failed. This practice technique is similar to methods used in guided discovery learning, which is often compared in the literature to direct instruction or studying worked-out examples.

Some researchers (Paas & Van Merriënboer, 1994; Reed & Actor Bolstad, 1991; Sweller & Cooper, 1985; Tuovinen & Sweller, 1999) make a convincing case for the value of using multiple worked-out examples in place of direct instruction on the problem solving principles or in place of problem solving practice. However, worked-out examples in most cases offer a transfer enhancement only if the base and the target problems are isomorphic. In such cases, multiple worked-out examples of increasing complexity allow the problem-solvers to focus on the structural, and not superficial similarities of the problems, and adapt solutions from the base to the target problems.

Such a benefit, though seemingly useful, offers only a short term advantage to the problem-solver, who can now solve a particular type of problem but must be trained on all others. Thus, a method that would allow for more generalized learning,
such as learning to look independently for common elements or common rules among several problems, would be more advantageous for instruction purposes. This method seems to be the guided discovery method, where students engage in independent or group problem solving with some guidance from their teacher, but without being explicitly told what the rules for problem solving are. McDaniel and Schlager (1990) demonstrated the advantage of this technique using river-crossing and water jar problems. They showed that college students who practice independent problem solving to criterion demonstrate positive analogical transfer unlike subjects in single base problem experiments (e.g., Gick and Holyok, 1980; Novick, 1988) or subjects who are given worked-out solutions (Reed, 1987). Likewise, subjects who discover a rule for solving a set of problems can discover another rule for a novel set of problems instead of demonstrating an Einstellung effect.

Muthukrishna and Borkowski (1995) collected additional evidence in support of discovery learning as a tool for problem solving transfer. They trained elementary school students on mathematical problem solving using direct instruction, guided discovery, or the combination of the two methods. They found that while the performance of the guided discovery group was comparable to the performance of the direct instruction group on near-transfer tasks, it was superior to the performance of the direct instruction group on far-transfer tasks, when the target problems were more different from the base problems.

The last study is different from most of the others. Here, instead of working with children in a laboratory setting one at a time, the researchers came into the classroom and used the children’s natural learning environment. Experiments of this
nature might not only be more appealing to educators, but also be more generalizable due to minimization of demand characteristics when compared to the laboratory studies. Recently, there has been a decline in randomized intervention-based research conducted in the classroom with children (Hsieh et al., 2005), despite a demand for such studies expressed in the *No Child Left Behind* federal legislation (2001).

However, Levin (1993) cautioned classroom researchers about compromises that have to be made in classroom settings, referring to them as “methodologically messy places” (p. 4). Any classroom experiment has to balance ecological and internal validity, where trying to set the conditions to be as close to real life classroom situations as possible can compromise control of the experiment. For example, children in different schools and even different classrooms within the same school have been exposed to different teaching practices, which can potentially affect the outcome of a study. Nevertheless, a classroom is a setting where problem solving is taught and where word problems are usually solved, and so it should be studied directly.

Therefore, the present series of investigations into the role of rules and instruction techniques in transfer of mathematical problem solving was conducted directly in middle school classrooms with word problems. Despite Reed’s (1999) assertion that transfer to unrelated problems is not very interesting, it seems that this type of transfer should be most telling when investigating the long-term usefulness of instruction techniques. Thus, in all five present experiments base problems are unrelated to the target problems.
The first two experiments investigated the effects of direct rule instruction or contingency rule learning on transfer of problem solving using sets of transfer problems, unrelated to the base problems, that either are solved by the same rule or are unrelated to each other. Experiment 1 used a set of transfer problems which were equivalent to each other, while Experiment 2 used a set of transfer problems which were unrelated to each other. Given findings of Chen (1999), Fantino et al. (2003), and McDaniel and Schlager (1990), in Experiment 1 it could be expected, despite Luchins’ (1942) finding of rigidity, that students who have learned a rule either through instruction or through learning it on their own by problem solving to criterion should not be impaired on transfer to problems with a novel rule in comparison to students who have not learned a rule. It was also hypothesized that the group that would learn the rule directly through contingencies by trying to reach a criterion would perform better on the test problems where a new rule would have to be learned. In Experiment 2, designed to serve as a control for Experiment 1, no new rule would be available for learning during the transfer test, since each problem would require a different solution rule. Given that Fantino et al. found that instructed rule training resulted in fewer changing rule transfer problems solved than either contingency-based rule training or training with changing rules, a similar effect might be expected with word problems.

Next, since many researchers in education advocate peer-group problem solving (Webb & Palinscar, 1996; see also Davenport & Howe, 1999), it appeared appropriate to investigate it in context of the rule-learning and transfer. Thus, in Experiment 3, in addition to independent problem solvers practicing with same-rule problems and independent problem solvers practicing with changing rule problems,
there were pairs of collaborating problem-solvers practicing with either same-rule or changing rule problems. A dyad was chosen over other group sizes due to evidence cited by Webb and Palinscar of more interaction and mutual attention in such a group as compared to a triad or a quad. Transfer effects and rule learning were assessed, with the goal of testing, in particular, whether the collaborative discovery learning strategy used by Muthukrishna and Borkowski (1998) had any advantages over individual discovery learning. As a pilot experiment, this study also presented participating students with some rule-related questions, aiming to investigate the potential of using these questions in further studies.

Further, Experiment 4 attempted to investigate whether students actually infer or use any rules during practice with word problems, either after direct instruction on a rule or after discovering a rule through contingency based problem solving. To assess this, students were asked to write out instructions for another child on how to solve a problem or a group of problems before and/or after solving the problems themselves. Procedures similar to this were used in several studies on transfer of problem solving (e.g., Chen, 1999; Novick & Holyoak, 1991; Rittle-Johnson, 2006).

Next, an experiment was conducted to assess which training method provides students with an opportunity to look past superficial details and infer rules. This was assessed at the mapping level, by providing practice problems that were solvable by the same rule but contain extraneous information. After practice, students were asked to describe how they solved the problems, as well as what the problems had in common. Then, they were asked to solve a set of problems unrelated to the base problems, but solvable by the same rule and containing irrelevant information. If
students learn during training to search out relevant information and compare it to the relevant information in previous problems, they should be able to infer rules by which problems are solved. This practice of filtering out irrelevant information, whether or not it would be interpreted as a general rule (Chen, 1999), was expected to benefit the students on the transfer test.
Chapter II

Experiment 1

This study sought to investigate the effects that direct rule instruction or contingency rule learning have on transfer of problem solving. Another aim of the study was to clarify conditions under which rule learning could result in the Einstellung effect discussed by Luchins (1942). This investigation was guided by Skinner’s (1969) analysis of the differences between rule-governed and contingency-shaped behavior. Skinner defined rule governed behavior as behavior that is controlled by "rules derived from the contingencies in the form of injunctions or descriptions which specify occasions, responses and consequences" (Skinner, 1969, p 160). He also proposed that people who follow rules or instructions would behave differently than people who are guided by direct exposure to environmental contingencies (Skinner, 1974). That distinction led to some operant laboratory experiments (e.g., Shimoff, Catania, & Matthews, 1981) showing that rule-governed behavior is relatively insensitive to changes in environmental contingencies. However, these studies were done using procedures that were far too remote from Luchins’ (1942) to draw conclusive connections. Thus, the present study intended to provide insight into the issue that would be of import to both learning psychologists and educators.

Chen (1999), Fantino, Jaworski, Case, and Stolarz-Fantino (2003), and McDaniel and Schlager (1990) collected prior evidence that students who learn a rule either through instruction or through learning it on their own by problem solving to criterion are not necessarily impaired on transfer to problems with a novel rule. However, while McDaniel and Schlager found evidence supporting positive transfer
effects of rule learning by discovery, where subjects learn the rule through experiencing contingencies, Fantino et al. observed larger positive transfer effects of rule learning by instruction. The present experiment was designed to investigate whether the positive transfer effects for the contingency-based learning rule group would be higher than the positive transfer effects for the instruction-based learning rule group.

Method

Subjects. Participants were 104 sixth grade students from five math classes at the Preuss School – a charter middle school located on the campus of the University of California, San Diego. Students participated in their classrooms during class time, with active involvement of their mathematics teacher. The data were collected during two different school years, with the same teacher participating both times. Due to the conditions of school participation in the study, no precise data on gender and ethnic composition of the sample are available. However, the school serves motivated children with high academic potential who come from low-income and predominantly minority households.

Materials and procedure. All students from the participating classes, which had 25 students each, were given parental consent forms by their teacher, 104 of which were returned signed. All participants also signed their own consent forms. Students who did not have parental consent or chose not to give their own consent were given alternative assignments by their teachers.

For each class, the experiment took place during a regularly scheduled, 90-minute math period. Prior to starting, the experimenters gave the participants and other
students in the class a 15-minute presentation about psychology and psychology research practices. Then, participating students in each class were randomly assigned to one of three groups: instruction (n = 35), contingency (n = 34), and control (n = 35), which were identified in the classroom by the colors red, green, and blue to keep subjects blind to their actual differences. Color-coded name tags that students pulled out at random were used to accomplish this assignment. Immediately afterwards, students from the contingency and control groups were taken outside of the classroom by undergraduate research assistants. Meanwhile, the instruction group received detailed instructions for solving a set of equivalent practice problems, all of which involved making round-trips of some distance a certain number of times per week, per month, etc. All problems could be solved by the rule Y = 2XZ, where X was the distance, Z was the number of round trips made, and Y was the total distance. The children’s mathematics teacher went over a worked example problem with this group of students, identifying known and unknown factors and discussing the solution. Students were also given a sheet with the problem worked out using a picture (Appendix I). The teacher instructed them to multiply distance by 2 (for the round-trip), then multiply the product by the number of times the trip was made, and the experimenter announced that they would now be given several problems that could be solved by this procedure.

Once the teacher answered all questions from the instruction group students, the other two groups were brought back to the classroom. The students were seated in different parts of the classroom according to their groups and given sets of 15 practice problems assembled into a booklet. The contingency group received the same set of
practice problems as the instruction group, but was not instructed how to solve them. The control group also received no instruction in solving the problems; in fact, their set of practice problems could not be solved by any general rule since their problems were unrelated. Students in all three groups were instructed verbally and in their booklets to solve correctly three practice problems in a row in order to move on to the next part of the study. Experimenters provided feedback after each problem, either telling students that they were correct with minimal praise (“Good,” “That’s right,” etc.) or telling them that they were wrong and giving them the correct answer. No other feedback was given.

After meeting the criterion on the practice problems, students solved two sets of three test problems, presented in counterbalanced order (Appendix II). These problems were structurally equivalent to each other but unrelated to the practice problems. The rule that could apply to these problems was $Y = 60X/Z$. No feedback was given during this phase.

Results and Discussion

A problem was considered to be solved correctly if a student found the right way to solve it, even if he or she made an arithmetical error in the process. This approach was taken because arithmetic skills of the subjects were not of interest for the study, and do not necessarily have a relationship with their problem-solving skills (Hegarty, Mayer, & Monk, 1995). Since the literature discusses successful problem solvers using various criteria for this designation, here the successful problem solvers were defined as those who solved at least 4 out of the 6 test problems (67%) correctly.
Based on prior findings such as those of Phye (2001) or McDaniel and Schlager (1990), it was expected that the instruction group would outperform the contingency and control groups on practice problems. Indeed, most students in the instruction group reached the criterion of three solved problems in a row immediately (M = 3.94, SD = 1.85). In the contingency (M = 6.38, SD = 3.88) and control groups (M = 7.96, SD = 4.00), however, some students had to work through all 15 practice problems to reach the criterion (Figure 1). The Levene’s test for homogeneity of variance indicated that the variance of the instruction group differed significantly from both the contingency group variance (F = 11.581, p < .01) and the control group variance (F = 27.951, p < .01). This necessitated a t-test assuming unequal variance, which indicated significant differences between the mean numbers of problems.

Figure 1: Average number of practice problems attempted before reaching the criterion of three consecutive problems solved correctly in Experiment 1.
attempted to criterion of the instructed and contingency groups ($t(46.93) = -3.321, p < .01$ two-tailed) and the instructed and control groups ($t(47.84) = -4.72, p < .01$ two-tailed). There were no significant differences with respect to the practice task between the contingency group and the control group.

The data on the transfer performance over all six transfer problems are presented in Figure 2, and support the hypothesis that contingency-based learning should produce higher positive transfer effects than instruction-based learning. The overall performance of the contingency group (M = 49%, SD = 44%) during the transfer stage was significantly better in terms of the percentage of test problems solved than the performance of either instruction (M = 29%, SD = 34%) or control (M = 30%, SD = 36%) groups, as indicated by a one-tailed t-test assuming unequal

![Figure 2: Average percentage of transfer problems solved correctly in Experiment 1, out of six possible problems.](image-url)
variances, with \( t(61.45) = 2.032, p < .05 \) and \( t(61.13) = 1.89, p < .05 \), respectively. In turn, the latter two groups were not significantly different from each other. Likewise, only 20% of students in the instruction group and 26% of students in the control groups could be classified as successful problem-solvers because they correctly solved at least four out of six transfer problems. The contingency group had a much higher proportion of successful problem-solvers, with fifteen out of thirty four students, or 44%, solving four or more problems correctly (Figure 3). A chi-square test indicated that these were significant differences, with \( \chi^2(2, N = 104) = 6.56, p < .05 \).

The performance of the contingency group was superior to that of either the instruction or the control group on both test segments, though statistically significant.
only for Segment 2, with $t(66.22) = 1.81$, $p < .05$ and $t(64.75) = 2.31$, $p < .05$ (Figure 4). Students in instruction and control groups did equally poorly on both test segments.

Thus, students who learned the rule on their own (the contingency group) did best on the problems with a novel rule. Interestingly, none of the groups showed any significant change from first to second test segments, when they had to work without feedback. The superiority of both rule groups to the controls suggests that rule-based problem solving need not be inflexible. Indeed, it appears that at least when all target problems can be solved by the same rule, the best method of practice is to allow students to practice independently with multiple problems that also solve by the same rule, though different from the target rule. An important point that distinguishes this
experiment from those of Luchins (1942) or Fantino et al. (2003) is that in this experiment, students were required to practice until they reached a pre-set mastery criterion, without any time limits. This difference might account for results supporting contingency-governed rule learning in this experiment, similar to the findings of McDaniel and Schlager (1990), who also used a mastery criterion with their college-age subjects. This advantage of contingency-governed learning over the rule-governed learning such as that taking place for the instruction group is in line with Skinner’s (1966/1988) assertion that contingency-governed learning is more efficient than rule-governed learning.

The results also support the educational practice of guided discovery learning through feedback over the practice of direct instruction. Mayer (2004) proposed that there should be a “three-strikes rule against pure discovery learning,” citing research from three diverse areas among which were studies of problem-solving rules. Pure discovery means practice without feedback, similar to what occurred during test segments in the present experiment. Just as Mayer discussed, there was no learning apparent from that experience. However, guided discovery approaches often include instruction-based methods in addition to contingency-based methods. The present results indicate that there is a difference between the two, and that using rule instruction may contribute to difficulties in transfer when the transfer is to novel problem types.
Chapter III

Experiment 2

This study sought to expand on the results of Experiment 1, serving as a control experiment. Since Fantino et al. (2003) found that instructed rule training resulted in fewer changing rule transfer problems solved than changing rule training, then a similar effect might be expected with word problems. Thus, effects of direct rule instruction or contingency rule learning on transfer of problem solving were investigated using sets of transfer problems, unrelated to the base problems, that were also unrelated to one another.

Method

Subjects. Participants were 48 sixth grade students from two math classes at the Preuss School. Students participated in their classrooms during class time, with active involvement of their mathematics teacher.

Materials and procedure. The design of this experiment closely resembled that of Experiment 1, with the only difference in transfer materials. Instead of two sets of three problems solvable by the $Y = 60X/Z$ rule, two sets of three problems completely unrelated to each other or to any of the practice problems were used. Random assignment resulted in 13 in the instruction group, 16 in the contingency group, and 18 in the control group.

Results and Discussion

The results of this study did not support the hypothesis that rule training in either contingency or instructed group would be detrimental to transfer to changing-rule problems. First, as shown in Figure 5, contingency and control group students
took about the same average number of problems (M = 6.63, SD = 3.56 and M = 6.71, SD = 4.48), to reach the criterion of three problems in a row solved correctly. The instruction group performed as expected, with most students there achieving the criterion with the first three problems in the set and M = 4.38, SD = 2.26, significantly different from the instruction group (t(27) = -1.97, p < .05, one-tailed) and the control group (t(28) = -1.70, p < .05, one-tailed). The almost identical performances of the contingency group and the control group are consistent with the results of Experiment 1, especially given that identical materials and procedures were used.

The main finding of this experiment is that no disadvantage to any of the experimental groups over the control group was observed on the transfer test, as might

Figure 5: Average number of practice problems attempted before reaching the criterion of three consecutive problems solved correctly in Experiment 2.
have been expected from Fantino et al.’s (2003) results. All groups correctly solved almost the same number of transfer problems on average: $M = 3.62$, $SD = 1.71$ in the instruction group, $M = 3.13$, $SD = 1.89$ in the control group, and $M = 3.0$, $SD = 1.77$ in the contingency group. Though, as shown in Figure 6, a larger proportion of students in the instructed group (54%) than in both contingency (44%) and control (29%) groups, and a larger proportion of students in the contingency group than in the control group, solved at least four out of six transfer problems, a chi-square test detected no significant differences.

These results, given the changing rule nature of the transfer problems, cannot be easily reconciled with prior findings. However, it appears that mathematical ability
of the students, more than the treatment manipulation, affected whether or not they were successful on the test. If, regardless of group placements, the subjects are divided according to their practice performance into those who took only three problems to reach the criterion and those who had to solve 4 or more problems, these two groups differ in terms of their test performance. For the 22 subjects who reached the criterion in three problems, the mean number of test problems solved was 4.00, SD = 1.48, while for the remaining 24 subjects the mean number of test problems solved was 2.5, SD = 1.75, a significantly lower result (t(44) = -3.13, p < .01, two-tailed). Such difference did not exist for subjects in Experiment 1, indicating that for the test problems which were all unrelated to each other, direct instruction or contingency practice of any kind did not matter as much as the preexisting skills.

Overall, though, it appears that if the rules for equivalent problems are, in fact, learned during instruction or inferred during uninstructed contingency-based practice, they do not have any transfer effect on solving a set of problems that are unrelated to either each other or to the practice problems. However, the present experiment only assumes that the rules are inferred or learned by the students, since it did not directly measure rule learning. Experiment 4 and Experiment 5, below, attempted to assess rule inference and/or learning from instruction directly.
Chapter IV

Experiment 3

Since Experiment 1 demonstrated that discovering a rule for solving a particular type of a problem through practice with feedback has a positive transfer effect on solving problems unrelated to the training problems but equivalent to each other, the question arises of how to make this guided discovery easier and the effect stronger. The direction for the present experiment was suggested by the mathematics teachers who participated in Experiments 1 and 2, and who told us that their students often work on problem solving in pairs or small groups. This seems to be a popular and widely investigated way to teach problem solving (Muthukrishna and Borkowski, 1995). An issue of interest in this exploratory study was whether learning how to solve base problems in collaboration with a partner would have a stronger positive effect on transfer to unrelated target problems than doing it alone. Also, would students working together require fewer trials to criterion than students working alone?

Method

Subjects. Participants were 62 6th grade students from three science classes taught by the same teacher at a public elementary school in a middle to higher income suburban area. Students participated in their science classroom during regularly scheduled 50-minute periods, without active involvement of their teacher. In this school, 80% of students are white, 11% are Asian, and the rest are Hispanic or African-American. Due to the circumstances of school participation in the study, no precise data on gender and ethnic composition of the sample are available.
Materials and procedure. A week before the study was scheduled to take place, the experimenters visited all classes to discuss their participation with the students. In a 15 minute presentation, psychology as a science and research methods in psychology were briefly explained to the students. Then, after a brief description of the study, parental consent forms were distributed. This was a more focused presentation than in Experiments 1 and 2, since the experiment became a part of the science curriculum at this site. On the day of the experiment students who did not return parental consent were given alternative science assignments by their teacher. All participants also signed their own assent forms.

Because this was a science classroom instead of the math classrooms used in previous experiments, it was decided that this experiment would not have a direct instruction-based group. The positive transfer effect of a contingency-based method of teaching is well documented in other experiments in the series, so the present experiment focused on refining the understanding of its use in the classroom. Thus, students in each class were randomly assigned to one of the four groups: same rule (n = 10), same rule/collaboration (n = 16), changing rule (n = 14), and changing rule/collaboration (n = 22). As in previous experiments, the groups were identified in the classroom by different colors with the aid of color-coded name tags to keep subjects blind to their actual differences.

Simultaneously, the subjects in the two collaboration groups were also randomly assigned to pairs. Same rule groups were provided with practice problems that could be solved by the same rule, while changing rule groups were provided with practice problems that had nothing in common. The practice problems for same rule
groups were 10 practice problems taken from the set of 15 practice problems used for the instruction or contingency groups in Experiment 1, and the practice problems for changing rule groups were 10 practice problems taken from the set of 15 practice problems used for the control group in Experiment 1. All groups were asked to work on the practice problems until they either reached the criterion of 3 consecutive correct solutions or solved all of their practice problems. Individual students or student pairs received feedback after each problem they reported as completed. As in Experiments 1 and 2, the feedback consisted of checking the answers, saying “Good!” or “This is right” if the answer was correct, or writing down the correct answer and asking the students to try the next problem if the answer was incorrect. Students were also reminded how many more correct solutions they had to achieve after each problem they reported solved.

In the test phase, all students were individually given a set of 3 transfer problems that could be solved by the same, novel rule. These problems were a subset of the 6 problems used in Experiment 1. The experiment also included pilot versions of questionnaires to be later used in Experiments 3 and 4, which had no significant effect on the outcome and are not analyzed here.

Results and Discussion

Mean numbers of problems required during practice to reach the criterion were as follows: M = 3.89, SD = 1.32 for the same rule group, M = 3.90, SD = 1.50 for the same rule/collaboration group, M = 3.00, SD = 0 for the changing rule group, and M = 3.36, SD = .90 for the changing rule/collaboration group. Every student in the changing rule group reached the criterion on their first three problems, and significant
differences were found between this group and all of the remaining groups: the same rule group \((t(9) = -1.96, p < .05, \text{one-tailed})\), the same rule/collaboration group \((t(17) = 2.85, p < .05, \text{two-tailed})\), and the changing rule/collaboration group \((t(21) = 1.89, p < .05, \text{one-tailed})\). This result is contrary to the findings of both Experiment 1, where the equivalent of the current same rule group outperformed the equivalent of the current changing rule group, and Experiment 2, where the two groups were not significantly different.

There are two factors that could have contributed to this outcome. First, due to the low parental consent rates the group sample sizes were very small, and a possibility exists that a larger proportion of students proficient at math was assigned to the changing rule group. If this was the case, then the ceiling effect is most likely obscuring any potential differences. Second, the reduction of the practice problem set from 15 to 10 was accomplished mainly through removing all problems involving percent, which used to comprise a third of the original set, and one of which occupied number 3 slot in that set. This was done for the sake of time, since this type of problem, as potentially more challenging, could have prevented some students from completing the work soon enough to advance to the next phase.

On the transfer test, as seen in Figure 7, more students in both of the collaboration groups were successful on the transfer test, solving at least 2 out of 3 problems correctly. However, these differences were not significant. Likewise, there were no significant differences among the means of the four groups, though they showed a similar trend. A post hoc t-test detected that students who achieved the practice criterion with the first three problems did significantly better on the transfer
Thus, for these subjects the type of practice they engaged in had no significant effect on the transfer test outcome, while it appears that their pre-existing ability might have had an effect. According to the California Department of Education’s standardized testing conducted in the spring of that school year (the study was conducted in the fall), 68% of sixth graders at this school performed at or above proficient level, compared with the state overall 34%. It is very likely that the testing materials were too basic for some of these subjects, resulting in a ceiling effect.

Though this study did not find support for group-based practice (Muthukrishna and Borkowski, 1995), it does not warrant a rejection of the idea as not useful for
transfer (Mayer, 2004) because there was a trend favoring the collaboration groups during transfer. Given the small sample size and potential ceiling effects, it is not possible to draw from this study firm conclusions regarding either collaboration during contingency-based learning, or the effect rule based practice has on transfer. However, repeating the study at the same school to increase the sample size is not necessarily a good way to improve this study, unless other design features are modified. During a single short class period, the dynamics of dyad interactions coupled with the presence of individual learners working on the same task were difficult to control. It does not appear that the format used in the present series of studies will lend itself to the study of group discovery learning, since it is not compatible to the long-term projects commonly used in education research, such as in Muthukrishna and Borkowski (1995). Therefore, further investigation focused on assessing the role of rules learned through contingency-governed practice or through direct instruction in transfer of problem solving in more direct ways. The following two experiments included questionnaires, piloted in Experiment 3, which asked students to verbalize the rules they might have used to solve the word problems used in the study.
Chapter V

Experiment 4

Experiment 1 demonstrated that contingency-based practice with a class of equivalent problems is beneficial for transfer to a class of equivalent problems that are unrelated to the base problems. That finding bears on a dispute regarding the role rule learning plays in transfer (Fantino et al., 2003; Luchins’, 1942). Though it appears that contingency-based rule learning is beneficial for transfer, can we be sure that rules were in fact learned during contingency practice? Taking up this question, Experiment 4 investigated whether students actually infer or use any rules during practice with word problems, either after direct classroom instruction on a rule or after discovering a rule through contingency-based problem solving. Two procedures were utilized for assessing this issue. In one procedure, students were asked to describe similarities among the problems and to write out instructions for another child on how to solve a problem or a group of problems instead of solving the problems themselves. In another, students were asked to do the same after solving the problems.

These procedures posed a concern about the effects that answering such questions and consequent heightened awareness of the problem similarities and rules may have on the participants’ performance. Later, a Rittle-Johnson’s (2006) study showed that requiring subjects to self-explain how they solved a problem or why a particular solution is right or wrong enhanced their transfer test performance. Since such an effect has been considered a possibility, about a third of the students were asked questions about practice problems immediately after the practice, a third of the students were asked questions about the test problems before the test, and all of the
students were asked questions about the test problems after the test. If verbalizing the rules and problem similarities was a factor that can enhance transfer performance, then it would be expected that the groups who were questioned either about practice problems or about test problems before the transfer test would do better on the test than a group that was not questioned before the test.

Method

Subjects. Participants were 236 6th grade students from eleven math classes taught by four different teachers at a public middle school in an urban area with predominantly Hispanic (78%), low income (69%) students. Students participated in their classrooms during regularly scheduled 50-minute math periods, with active involvement of their mathematics teacher. Due to the circumstances of school participation in the study, no precise data on gender and ethnic composition of the sample are available.

Materials and procedure. A week prior the study was scheduled to take place, the experimenter visited all classes to discuss their participation with the students. In a 10-15 minute presentation during regular math periods, psychology as a science and research methods in psychology were briefly discussed, and parental consent forms were distributed after a brief description of the study. On the day of the experiment students who did not return parental consent or did not sign their own assent forms were given alternative assignments by their teachers.

Students in each class were randomly assigned to one of six groups: instruction with post-training questions and post-test questions (instruction/post-training, n = 45), instruction with pre-test questions and post-test questions (instruction/pre-test, n = 36),
instruction with post-test questions only (instruction/post-test, n = 37), contingency with post-training questions and post-test questions (contingency/post-training, n = 37), contingency with pre-test questions and post-test questions (contingency/pre-test, n = 40), and contingency with post-test questions only (contingency/post-test, n = 40). Somewhat uneven group sizes are due to very low participation numbers in some of the classes, which resulted in a decision to have fewer group divisions in those classes. Overall, 118 students received instruction, and 117 had to learn from contingencies, so the main group divisions were close in size. The groups were identified in the classroom by different colors to keep subjects blind to their actual differences. Just as in the above experiments, color-coded name tags that students pulled out at random were used to accomplish this assignment.

Immediately afterwards, students from all three contingency groups were taken outside of the classroom by the research assistants. Meanwhile, the instruction groups received detailed instructions for solving a set of equivalent practice problems. The problems were the same as in Experiments 1 and 2 and could be solved by the rule \( Y = 2XZ \). The children’s mathematics teachers went over a worked example problem with this group of students, identifying known and unknown factors and discussing the solution. Students were also given a sheet with the problem worked out using a picture. All teachers involved their students in discussing the problem, and reviewed general problem-solving practices such as careful reading. The teachers eventually instructed the participants to multiply distance by 2 (for the round-trip), then multiply the product by the number of times the trip was made, and the experimenter
announced that they would now be given several problems that could be solved by this procedure.

Once the teachers and the experimenter answered all questions from the instruction groups’ students, the contingency groups were brought back to the classroom. The students were seated in different parts of the classroom according to their groups and given sets of 10 practice problems assembled into a booklet. A subset of the 15 problems originally used in Experiments 1 and 2 was used here due to shorter periods at this school as compared to the Preuss School class periods. The contingency group received the same set of practice problems as the instruction group, but was not instructed how to solve them. Students in both groups were instructed verbally and in their booklets to solve correctly three practice problems in a row in order to move on to the next part of the study. Experimenters provided feedback after each problem, either telling students that they were correct with minimal praise (“Good,” “That’s right,” etc.) or telling them that they were wrong and giving them the correct answer. In addition to this standard feedback, students in the instruction groups were directed to their worked-out example page when they first presented an incorrect answer. If they persisted in not solving the problem correctly, they were given the correct answer and told to move on to the next problem. Once students either met the criterion of 3 problems in a row solved correctly or worked through all 10 practice problems, their practice booklets were collected and they moved on to the next phase of the study.

Students in the instruction/post-practice and contingency/post-practice groups were given a separate page with questions related to the practice problems. One of the
questions asked: “You just worked with some word problems, where numbers were included in a story. Do you think these problems had anything else in common? If they did, what was it?” A second question asked: “If you needed to explain to a friend how to solve these problems, could you come up with a rule for your friend to use with them? If you can, what is the rule that can help solve all these problems?” These questions are similar to those asked in other studies (e.g., Moreau & Coquin-Viennot, 2003) under like circumstances.

Students in the instruction/pre-test and contingency/pre-test conditions were given a separate page with questions related to the test problems, which listed the three test problems and instructed students to read but not solve them. The first question asked: “Do you think these problems have anything else in common? If they do, what was it?” The second question asked: “If you needed to explain to a friend how to solve these problems, could you come up with a rule for your friend to use with them? If you can, what is the rule that can help solve all these problems?” These questions are virtually identical to those asked in a post-practice questionnaire.

After completing their respective questionnaires, students in both of the post-practice and both of the pre-test groups moved on to the test problems, as did the students in both of the post-test groups immediately following practice. These problems were equivalent to each other but unrelated to the practice problems.

For this experiment, using 6 test problems was impractical due to the time constraints. Therefore, only 3 test problems were used, especially since no learning took place in Experiment 1 over the two 3-problem test segments. Another change as compared to Experiments 1 and 2 was that the new problems now were solved by a
different formula, \((X+Y)/Z\). This change in the procedure was instituted because formerly used problems involved time conversions, which appeared to be a difficult issue for the six-graders, as evidenced by low success rates in Experiment 1. Coupled with the time constraints, this could prevent too many students from completing the test, so the problems were replaced by those the teachers judged to be simpler. No feedback was given during testing.

Finally, following the test phase, all groups were given a post-test questionnaire which was similar to the pre-test questionnaire and contained the same questions.

Questionnaire coding. The answers to the two questions on the three questionnaires were coded using several variables, because multiple answers were possible. “Solution” variable was coded using “0” if a subject did not mention any mathematical operations necessary for solving the problem, such as “add,” “times 2,” or “multiply,” and “1” if he or she mentioned mathematical operations. “Context” variable was coded using “0” if a subject did not mention story features such as “roundtrip,” “there and back,” “taking things somewhere,” “how many will fit,” and so on, and “1” if he or she mentioned these story features. “General” variable was coded using “0” if a subject did not mention careful reading, looking for numbers, and other general aspects of solving word problems, and “1” if he or she mentioned these general aspects of solving word problems. “Nothing” variable was coded using “1” if a subject wrote in “no,” “I can’t,” or “nothing” as the answer.
Results and Discussion

Practice. As Figure 8 shows, students in the instructed groups outperformed students in the contingency groups during practice, taking on average 3.07 problems to reach the criterion as opposed to contingency groups’ average 5.42 problems. This significant \( t(94.10) = -8.86, p < .01 \) finding is consistent with the results of Experiment 1 and prior research results such as Reed, Dempster, and Ettinger (1985) which show that worked out examples and other forms of direct instruction offer an advantage in solving equivalent problems.

Transfer test. Several students across all conditions did not complete the practice booklet in time to proceed to the next phase. Also, in the four groups that had
either a post-practice questionnaire or a pre-test questionnaire, some students did not finish it in time to work on the test problems. Thus, for the test phase the number of subjects varied in the following way: 36 for the instruction/pre-test group, 31 for the instruction/post-practice group, 36 for the instruction/post-test group; 27 for the contingency/post-practice group, 19 for the contingency/pre-test group, and 34 for the contingency/post-test group. The implications of this change will be discussed below.

In the test phase, there were no significant differences between any of the three instruction groups or any of the three contingency groups, suggesting that asking the students to directly self-infer the rules did not have an effect on transfer. On the other hand, each contingency group outperformed a corresponding instruction group.

**Figure 9:** Average percentage of students who solved all of their attempted transfer problems correctly in Experiment 4.
Therefore, the data from all three contingency groups and from all three instruction groups were combined for analysis. As seen in Figure 9, students in the instruction group averaged 75% correct, significantly less than students in the contingency groups who averaged 89% correct, with t(185) = 2.72, p < .001. Also, the percentage of students who solved 100% of the test problems they attempted correctly was 70% for instruction group and 89% for contingency group, a significant difference with \( \chi^2(1, N = 177) = 8.42, p = .004 \). Thus, the results of Experiment 3 so far replicated the results of Experiment 1. However, the main goal of this study was to investigate rule inference by students and its effects, which are described below.

**Questionnaires.** Table 1 presents the numbers and the percentages obtained for solution-related, context-related, general, and negative answers on each question of each questionnaire. It is notable that Question 2, which specifically asked for a solution rule, drew out very few context-related answers in all cases.

**Post-practice questionnaire.** On the post-practice Question 1 there was a trend for the contingency group students to give more solution-related answers than were given by the instruction group students (Table 1). However, it was not significant, with \( \chi^2(1, N = 69) = 3.493, p = .06 \). Nevertheless, the opposite might have been expected since the instruction group students reached the practice criterion significantly faster than the contingency group students, thus appearing to be better at solving this kind of problem. Only about a quarter of the instruction group students and almost half of the contingency group students were able to identify structural similarities between the practice problems at this point, seemingly contrary to the findings of Silver (1981). If the faster achievement of practice criterion is the
Table 1: Distribution of answers to post-practice, pre-test, and post-test questionnaires in Experiment 4. The answers are grouped by the teaching condition (instruction or control) and question number. Solution, context, general, and nothing categories are coded according to the description in text.

<table>
<thead>
<tr>
<th></th>
<th>Post-Practice</th>
<th>Pre-Test</th>
<th>Post-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Instruction</td>
<td>Contingency</td>
<td>Instruction</td>
</tr>
<tr>
<td>N</td>
<td>42</td>
<td>27</td>
<td>35</td>
</tr>
<tr>
<td>Q 1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solution</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>11</td>
<td>15</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>%</td>
<td>26.2%</td>
<td>35.7%</td>
<td>34.3%</td>
<td>31.4%</td>
<td>28.6%</td>
<td>35.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>14</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>%</td>
<td>33.3%</td>
<td>2.4%</td>
<td>34.3%</td>
<td>0.0%</td>
<td>17.9%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>6</td>
<td>15</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>%</td>
<td>14.3%</td>
<td>35.7%</td>
<td>8.6%</td>
<td>20.0%</td>
<td>9.0%</td>
<td>20.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nothing</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>17</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>26.2%</td>
<td>26.2%</td>
<td>22.9%</td>
<td>48.6%</td>
<td>10.7%</td>
<td>32.1%</td>
</tr>
</tbody>
</table>

important feature, though, then it should have an effect of its own on identification of problem similarities. Indeed, students in the contingency group who reached the criterion by 3 problems were significantly less likely to give context-related answers than other kinds of answers, while students in the instructed group were almost equally likely to do both, \( \chi^2 (1, 25) = 3.949, p < .05 \). Thus, those who were in fact successful problem solvers tended to identify problem similarities in a way that was consistent with Silver’s findings, but only in the contingency group. This suggests that the students in the instruction group might have solved the problems by mechanically following their teacher’s instructions and the practice worksheet, without analyzing the problems themselves. If so, they behaved as would be expected based on Luchins’ (1942) Einstellung effect.
Students in the instruction groups were less likely than students in the contingency groups to be unable to answer Question 2, $\chi^2(1, N = 69) = 4.684, p < .05$. In the contingency group, 51.9% of subjects did not come up with an answer to this question (Table 1), showing that they did not necessarily infer the relevant rule. This was true even for the contingency group subjects who reached the criterion in 3 problems, 46% of whom gave a negative answer to Question 2. A reversal of the Question 1 picture is interesting here, because it shows that when the instruction group students were specifically prompted to come up with a rule they could do it. Equal proportions (35.7%) of the instruction group students recalled either a solution-based rule or general problem-solving rule taught to them by the teacher.

There were significant correlations between the variable indicating whether students reached the practice criterion after only 3 problems solved and the variable indicating pooled solution-related answers on both post-practice questions (one-tailed Spearman’s $r = .249, p < .05$). Thus, it appears that, overall, students who were able to verbalize either structural similarities or solution rules for the problems were somewhat more successful during practice. This also fits with the existing literature (Silver, 1981; Reed, 1987; Moreau & Coquin-Viennot, 2003) which shows that more successful problem solvers focus on structural features of the problems while less successful problem solvers focus on superficial, context features.

Recent transfer literature makes a distinction between procedural learning, procedural transfer, and conceptual knowledge (e. g., Rittle-Johnson, 2006). While procedural learning involves application of learned rules to familiar problems, procedural transfer requires using them to solve novel problems. Conceptual knowledge is usually
assessed with tasks for which already learned procedures do not apply, but the same
general concepts must be used to generate an answer. Using mathematical equivalence
problems, Rittle-Johnson found that direct instruction is most effective in achieving
procedural learning, which is a kind of learning that subjects in the present series of
studies experience in the practice phase. Thus, this finding is in agreement with the
results of the present experiment. However, the more important conclusion of Rittle-
Johnson’s study is that prompting students to self-explain by discussing how and why
a correct or an incorrect answer could be obtained is helpful in enhancing procedural
transfer. Though this technique is more elaborate than the questionnaire used in the
present experiment, Rittle-Johnson’s results still suggest that the subjects who used
solution-related answers on the post-practice questionnaire should have been more
successful on the transfer test than the students who did not use solution-related
answers. However, neither the kind of answers students gave to the post-practice
questions, nor the fact of taking it as opposed to taking a pre-test questionnaire or no
questionnaire before testing, affected students’ performance on the transfer test. Thus,
contrary to Rittle-Johnson’s findings, it does not appear that asking children to
verbalize the similarities between practice word problems or the rules they used to
solve them results in higher success on novel problems.

Pre-test questionnaire. For students in the pre-test group, solution-related
answers on both questions correlated on pre-test and post-test, with $r = .603$ for
Question 1 and $r = .620$ for Question 2, $p < .01$. Similar correlations were observed for
most other answers (Table 2).

On the pre-test Question 1, there were no significant differences for any of the
Table 2: Intercorrelations for the Experiment 4 variables. The questionnaire data are grouped as a yes/no (1/0) for each type of answer, as well as by questionnaire (post-practice, pre-test, or post-test) and question number.

Note: * Correlation is significant at the 0.05 level.

** Correlation is significant at the 0.01 level
<table>
<thead>
<tr>
<th></th>
<th>Teach.</th>
<th>Cr. at 3</th>
<th>Success</th>
<th>Post-Practice</th>
<th>Solution</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Post-Practice</th>
<th>Context</th>
<th>Post-Practice</th>
<th>Post-Practice</th>
<th>Post-Practice</th>
<th>General</th>
<th>Post-Practice</th>
<th>Post-Practice</th>
<th>Post-Practice</th>
<th>Nothing</th>
<th>Posttest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cr. at 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Success</td>
<td>-.021</td>
<td>.052</td>
<td>.106</td>
<td>-.120</td>
<td>-.079</td>
<td>.194</td>
<td>.068</td>
<td>-.171</td>
<td>-.063</td>
<td>-.026</td>
<td>.100</td>
<td>-.020</td>
<td>.061</td>
<td>-.003</td>
<td>.261*</td>
<td>-.159</td>
<td>.166</td>
<td>.014</td>
<td>.124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.082</td>
<td>.010</td>
<td>.042</td>
<td>-.082</td>
<td>.137</td>
<td>.169</td>
<td>-.248</td>
<td>.242</td>
<td>.071</td>
<td>-.036</td>
<td>.029</td>
<td>.134</td>
<td>.097</td>
<td>-.037</td>
<td>.164</td>
<td>-.119</td>
<td>-.018</td>
<td>-.375**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.204</td>
<td>.274</td>
<td>.030</td>
<td>-.467**</td>
<td>-.089</td>
<td>-.108</td>
<td>-.245*</td>
<td>-.194</td>
<td>-.184</td>
<td>-.076</td>
<td>-.434**</td>
<td>.019</td>
<td>-.153</td>
<td>.034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.211</td>
<td>.620**</td>
<td>.085</td>
<td>-.077</td>
<td>.333*</td>
<td>-.215</td>
<td>-.452**</td>
<td></td>
<td>.067</td>
<td>-.147</td>
<td>-.161*</td>
<td>-.482**</td>
<td>.121</td>
<td>-.527**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.241</td>
<td>.603**</td>
<td>.111</td>
<td>-.591**</td>
<td>-.087</td>
<td>-.387*</td>
<td>-.123</td>
<td>-.200</td>
<td>-.141</td>
<td>-.324</td>
<td>-.125</td>
<td>-.314*</td>
<td>-.087</td>
<td>.074</td>
<td>.017</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.090</td>
<td>.607**</td>
<td>.000</td>
<td>-.090</td>
<td>-.004</td>
<td>-.130</td>
<td>-.208</td>
<td>-.395**</td>
<td>-.192</td>
<td>-.308</td>
<td>-.148</td>
<td>-.593**</td>
<td>.087</td>
<td>-.386*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.246**</td>
<td>.291</td>
<td>-.242</td>
<td>-.171</td>
<td>-.642**</td>
<td>-.113</td>
<td>-.000</td>
<td>.112</td>
<td>-.006</td>
<td>-.014</td>
<td>-.390**</td>
<td>-.118</td>
<td>.000</td>
<td>.105</td>
<td>-.452**</td>
<td>-.021</td>
<td>-.409**</td>
<td>-.146</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.116</td>
<td>.148</td>
<td>-.181</td>
<td>-.004</td>
<td>-.089</td>
<td>-.066</td>
<td>.363*</td>
<td>-.019</td>
<td>-.174</td>
<td>-.197*</td>
<td>-.482**</td>
<td>-.125</td>
<td>-.278</td>
<td>-.330*</td>
<td>-.399*</td>
<td>-.216*</td>
<td>-.663**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.190</td>
<td>.115</td>
<td>-.215</td>
<td>-.045</td>
<td>-.072</td>
<td>-.091</td>
<td>-.041</td>
<td>-.086</td>
<td>-.176</td>
<td>-.105</td>
<td>-.176*</td>
<td>-.124</td>
<td>-.141</td>
<td>-.165</td>
<td>.043</td>
<td>.153</td>
<td>-.078</td>
<td>-.200</td>
<td>-.053</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.170</td>
<td>.372*</td>
<td>-.145</td>
<td>-.200</td>
<td>-.046</td>
<td>-.102</td>
<td>.196</td>
<td>-.141</td>
<td>.440**</td>
<td>-.533**</td>
<td></td>
<td>.093</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td>-.135</td>
<td>.112</td>
<td>.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.112</td>
<td>.105</td>
<td>-.094</td>
<td>-.135</td>
<td>.112</td>
<td>.083</td>
<td>-.561**</td>
<td>-.159</td>
<td>-.469**</td>
<td>-.029</td>
<td>-.198</td>
<td>.068</td>
<td>.016</td>
<td>.315**</td>
<td>-.016</td>
<td>.162</td>
<td>.152</td>
<td>-.268</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.033</td>
<td>.109</td>
<td>.394</td>
<td>.094</td>
<td>.053</td>
<td>.319</td>
<td>.356*</td>
<td></td>
<td>.092</td>
<td>.590**</td>
<td></td>
<td>.254**</td>
<td>.033</td>
<td>.527**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
answers among the students in the instruction and the contingency group. Out of the subjects who came up with an answer to that question, 92.5% later got all of the test problems they attempted to solve right. This was significantly more than 60% of the subjects who answered “no” or “I can’t” ($\chi^2 (1, N = 50) = 7.018, p < .01$).

Interestingly, none of the 18 contingency/pre-test group subjects remaining at the point of taking the transfer test answered in the negative.

On Question 2, context-related answers were nearly impossible, and only a single subject chose to answer that way. Significantly more of the instruction group (47%) than of the contingency group (22%) subjects who tried to answer this question could not come up with an answer ($\chi^2 (1, N = 76) = 5.145, p < .05$).

All of the instruction group subjects who took the pre-test questionnaire reached practice criterion in 3 problems, and only 31.4% of these subjects used a solution-related answer on Question 2. In the contingency group, 90.9% of the subjects who reached the practice criterion on the first 3 problems later came up with an answer to Question 2, as opposed to 50% of the subjects who needed more than 3 practice problems to reach criterion. This difference is significant, with $\chi^2 (1, N = 27) = 4.909, p < .05$. Also, 54.5% of the contingency group subjects who reached the practice criterion by 3 problems and 18.8% of the contingency group subjects who reached the practice criterion after 3 problems used a solution-related answer on Question 2. This difference had only marginal significance, with $\chi^2 (1, N = 27) = 3.759, p = .053$. It appears that the contingency group students who figured out the practice problems quickly were more likely to infer rules for solving novel problems.

Would exposure to the problems and attempting to infer rules and similarities
affect problem-solving on the transfer test? A questionnaire condition interacted with the teaching condition (instruction or contingency), so that the teaching group placement had a significant effect for the pre-test group ($\chi^2(1, N = 50) = 4.160, p < .05$). On the test, 94.7% in the contingency group and 71% in the instruction group got all of the problems correct. Only one subject out of 19 in the contingency/pre-test group failed to solve all test problems attempted. This is interesting in the context of Rittle-Johnson’s (2006) findings regarding the effect of self-explanation, especially given the post-practice questionnaire effect. There were no significant effects on test success for solution-related, context-related, or general answers for either of the pre-test questions, but there was a significant effect on test success of simply coming up with an answer for Question 1. Thus, it appears that exposure to and answering the questions about the test problems before solving them was beneficial for the transfer test success. If this can be considered a form of self-explanation (Atkinson, Renkl, & Merrill, 2003), then self-explanation for word problems is more effective when it is targeted to the actual problems to be solved, and when there was no prior instruction with a different rule. Though Rittle-Johnson found no advantage for guided discovery learning combined with self-explanation as compared to direct instruction combined with self-explanation, it was in evidence in the present experiment.

**Post-test questionnaire.** Neither the teaching condition nor previous experience with post-practice or pre-test questionnaires affected whether a subject would give a negative answer on either of the post-test questions. Group placement also had no effect on whether or not subjects used solution-related answers, though a higher proportion of students in the contingency/post-test group than in the instruction/post-
test group did not use general answers, with $\chi^2(1, N = 48) = 4.076, p < .05$.

A difference in post-test answers between students who were and were not successful on the test was affected by their group placement. On Question 1, the students in the instruction/post-test group, who had not taken a questionnaire before this point, were more likely to use a solution-related answer if they solved all of their problems correctly ($\chi^2(1, N = 30) = 4.451, p < .05$). On the whole, students from the post-test only groups who did well on the test were much less likely to use a context-related answer to Question 1 ($\chi^2(1, N = 48) = 7.763, p < .01$) than students who did poorly on the test. This effect was mostly explained by differences in the instruction/post-test group, with $\chi^2(1, N = 30) = 8.72, p < .01$, because all 4 contingency/post-test students who did poorly on the test did not have time to take the post-test questionnaire. Analyzing general answers on Question 1 revealed that they were more likely to be given by students who were not successful on the test, $r = - .205, p < .05$. Further analysis revealed that this difference was significant for the instruction groups, with $\chi^2(1, N = 78) = 4.459, p < .05$. Also, for Question 2 in the instruction groups only, success on the test corresponded to a higher proportion of students (59.1%) using solution-related answers, and lack of success corresponded to a much lower proportion of the students (25%) using solution-related answers, with $\chi^2(1, N = 78) = 4.748, p < .05$.

No one in the contingency groups who did not do well on the test had time to complete the post-test questionnaire. Thus, it is unknown whether the effect would exist there as well. Also, in both post-practice and pre-test groups, but not in the post-test only groups, students who were not successful on the test were more likely to give
general post-test Question 1 answers, with $\chi^2(1, N = 43) = 4.118$, $p < .05$ and $\chi^2(1, N = 36) = 4.906$, $p < .05$, respectively.

Overall, in agreement with prior findings (Silver, 1981; Reed, 1987; Moreau & Coquin-Viennot, 2003), it appears that students who were successful on the test were more likely than unsuccessful students to be able to verbalize structural similarities and solution rules for the test problems. This ability was not impaired by being taught a different rule, just as expected given Fantino et al.’s (2003) results.

Did it matter on the post-test questionnaire whether the subjects did well during practice? Solution-related answers were given on Question 1 by 66.4% of the students who reached the criterion on the first 3 problems and only 38.1% of students who had to solve more than 3 problems to reach the criterion, with $\chi^2(1, 128) = 5.934$, $p < .05$. This effect was mostly expressed in the contingency groups, with $\chi^2(1, 50) = 7.025$, $p < .01$. Specifically, among the students who took a pre-test, 57.6% of those who reached the criterion on the first three problems and none of the students who had to solve more than 3 problems to reach the criterion used solution-related answers to Question 1, with $\chi^2(1, 37) = 4.734$, $p < .05$. The contingency/pre-test students who reached the criterion in 3 problems were also much less likely than those who did not reach it in three problems to use context-based answers, $\chi^2 (1, 13) = 5.306$, $p < .05$.

Thus, success during practice, though more common in the instruction group, was more likely to lead to post-test verbalizing of structural features of the test problems in the contingency group. This was most evident in students who previously answered the same questions on the pre-test questionnaire. Previously it was established that the contingency group’s subjects who were more successful during
practice were also more likely to verbalize solution rules for practice problems, showing an ability to infer rules without being taught. On the other hand, learning a rule for practice problems from a teacher and thus quickly reaching the practice criterion might have a negative effect on inferring rules for transfer problems.

**Discussion.** This study offered strong evidence in support of Experiment 1 findings. Namely, while direct instruction using an example problem and a rule offers immediate advantage with problems equivalent to the example, it is contingency-based practice that leads to most positive transfer to problems unrelated to the practice problems but equivalent to each other.

A potential problem for the study is that the drop-out rate in the contingency/pre-test group was proportionately much higher than in the other 5 groups, with 52.5% of subjects not proceeding to the test phase. More than half of these missing subjects, or 30% of the original 40 in the group, did not complete their practice booklets in time to answer their pre-test questionnaire. Overall, only .4% of the total number of instructed group subjects versus 24% of the contingency group subjects took too long in practice to complete the next phase. This agrees with the findings on benefit of instruction for solving equivalent problems discussed above, but poses problems for analysis of data collected with these reduced numbers of subjects. According to informal observations in the classrooms, students who did not move on past practice problems were among the weakest mathematically. Assuming that the group assignment was truly random, it is not unlikely that the instruction groups had more of the poor problem-solvers than the contingency groups remaining after the practice. This raises the possibility that if the contingency groups outperformed the
instruction groups on the transfer test as in Experiment 1, the results could be attributed to superior composition of the contingency groups.

Due to the field nature of this study, this concern cannot be addressed through replication because students who have difficulty with math will most likely always drop out of the contingency groups more frequently than out of the instruction groups. Another possibility would be timing the practice, but that does not seem practical because children would have to wait for their answers to be checked after each problem, and this time would vary unpredictably. However, one implication of this situation for actual classroom instruction is that even though contingency-based practice has advantages for immediate transfer as seen in Experiment 1, it may work best in classrooms with more mathematically advanced students, unless it is a part of a long-term teaching strategy.

Though there was no opportunity to directly assess the alternative explanation offered above, there is a way to do so through a post-hoc comparison. If the students who were working too slowly to proceed to testing in the contingency group were to make it to testing nevertheless, an alternative explanation predicts that the overall success rates in this group would be lower, thus diminishing the effect. Likewise, this implies that students who worked more slowly than others during practice should have done worse than others on the transfer test. To test this assumption, a chi-square analysis compared test success of the students who reached the criterion of 3 problems in a row correctly on their first three problems to students who reached this criterion after making some errors. There were no significant differences in the success rates between these students. Considering as a success at least 2 out of 3 test problems
solved correctly, 80% of the students who reached the criterion on the first 3 problems and 82% of the students who reached the criterion after 4 or more problems were successful on the test, $\chi^2 (1, N = 177) = .08, p = .778$. This implies that there was no relationship between success during practice, taken apart from the teaching method, and success during the transfer test. Thus, it is most likely that the contingency-based practice in fact leads to higher transfer test success than the direct instruction does.

A goal of this study was to identify whether the subjects were using rules, or at least were able to verbalize them, and what conditions affected this. There was definite evidence of subjects verbalizing rules for both practice and test problems, as well as evidence that this verbalizing is affected by group placement and problem-solving success. Contingency groups’ students’ success was influencing and was influenced by verbalizing solution-related rules and similarities for practice problems, in agreement with the evidence collected by Fantino et al. (2003). On the other hand, instruction groups’ students tended to show more of the rigidity in their rule use evident in Luchins’ (1942) original work. This, however, was not the case for the instruction group students who were successful on the test, showing that the taught rules are not always an impediment.

Successful problem-solvers in this study tended to be more likely to verbalize rules and similarities for the test problems, and these rules and similarities were more likely to be mathematical solution-based and less likely to be general or context-based. Thus, these subjects were probably successful because they were dealing with the problems’ structure and not the problems’ story lines (context). Additional evidence in support of the idea that expressing solution rules indicates that the students learn about
structural features of the word problems instead of concentrating on the superficial, context features of the word problems comes from the fact that in every questionnaire there were strong negative correlations between solution-related and context-related answers (Table 2). Likewise, mentioning solution-related similarities negatively correlated with mentioning general rules for problem-solving. All of these findings agree with prior claims made by Silver (1981) or Reed (1987).

Also, the positive transfer effect of the pre-test questionnaire supports a recent Rittle-Johnson’s (2006) claim that self-explanation is advantageous for transfer. However, it appears that unlike in Rittle-Johnson’s study, the self-explanation effect only held for subjects who did not have rule instruction, and did not manifest itself for transfer problems when self-explanation was used with practice problems.

An additional consideration for the present experiment is the bilingual nature of the majority of the students. Recently, studies investigating second language learners working on word problems (Barwell, 2005; Bernardo & Calleja, 2005) found that understanding of the problems is impaired when these students work in their second language. It is likely that students in the instructed group were less affected by this during practice because they were given an example and told that all practice problems had the same structure. Also, since Question 2 on each of the three questionnaires was designed in a way that specifically required a rule, students who understood the question should not have used the problems’ context. However, Table 1 shows that some students did mention the context, indicating a potential reading comprehension issue. Since no data on language ability were collected at the time, this has to remain a speculation and a possibility for further inquiry.
Overall, this study supported findings of Experiment 1 and allowed for some additional insight into the role of rules in transfer of word problem solving. It showed that not only it is not harmful for transfer to infer the rules through contingency-based practice, but also it can be advantageous to have such experience. Potentially, it offers an interesting take on Skinner’s (1969, 1974) view of rule-governed and contingency-based learning. That is, contingency-based learning has been shown here to be more beneficial for transfer than rule-governed learning achieved through direct instruction. However, rules were not completely absent in the contingency-based learning process, because they were inferred by the subjects themselves. Since the subjects who inferred appropriate rules were more successful on the transfer test, then it could be said that the contingency-based learning is best when it is designed for inferring the rules.

Still, this study left some questions open. Conceptual knowledge, when applied to word problems, relates to the understanding that the problems have underlying mathematical structure (Rittle-Johnson, 2006). As discussed earlier, there was no significant effect of the instruction or contingency teaching method on mentions of problems’ mathematical structure or solutions in any of the questionnaires. However, a way to further assess this issue is though introducing irrelevant information into the problems, so that understanding of the problem structure (conceptual knowledge) would be necessary to be successful. Experiment 4 was designed to assess children’s inference and use of rules with or without instruction at the mapping level, by providing practice problems that solve by the same rule but contain extraneous information.
Chapter VI

Experiment 5

This experiment further investigated which training method provided students with an opportunity to look past superficial details and infer rules. Using irrelevant information imbedded into the problem’s story line was shown to be a useful method for assessing problem solvers’ understanding of the underlying problem structure (Cook, J. L., & Rieser, J. J., 2005; Low & Over, 1989, 1993). Thus, though the instruction and contingency conditions in the present experiment resembled those in Experiments 1 and 4, all practice and test problems contained additional story and numeric content that was irrelevant to their structure. If students learn during training to search out relevant information and compare it to the relevant information in previous problems, they should be able to infer rules by which problems are solved. This practice of filtering out irrelevant context should also benefit the students on the transfer test (Chen, 1999). Therefore, it was hypothesized that students who learned to solve a set of equivalent problems by contingencies would be able to express better understanding of their structure, as well as would be more successful solving a set of transfer problems and expressing understanding of the structure of those problems.

Method

Subjects. Participants were 135 6th grade students from five math classes taught by two different teachers at a public middle school in Torrance, CA where students are 49% White, 34% Asian, 11% Hispanic, 3% African-American, and 3% other ethnicities. At this school, 10% of students participate in the free lunch program. Due to the circumstances of school participation in the study, no precise data on ethnic
composition of the sample are available.

Students participated in their classrooms during regularly scheduled 54-minute math periods, with active involvement of their mathematics teachers. Twenty-eight students of Teacher 1 and 89 students of Teacher 2 participated in the study.

Materials and procedure. A week before the study was scheduled to take place, the experimenter visited all classes to discuss their participation with the students. As in Experiment 3, in a 10-15 minute presentation, psychology as a science and research methods in psychology were discussed, and parental consent forms were distributed after a brief description of the study. On the day of the experiment, students who did not return parental consent were given alternative assignments by their teachers. All participants also signed their own assent forms.

Students in each class were randomly assigned to one of two groups: instruction (n = 57) and contingency (n = 60). As in previous experiments, the groups were identified in the classroom by different colors with the aid of color-coded name tags to keep subjects blind to their actual differences.

Once the group assignment was completed, students from the contingency groups were taken outside of the classroom by the undergraduate research assistants. Meanwhile, the instruction group was instructed by their teacher in a way similar to prior experiments. All teachers involved their students in discussing the problem, and reviewed general problem-solving practices such as careful reading and watching out for irrelevant information. The teachers eventually instructed the participants to multiply distance by 2 (for the round-trip), then multiply the product by the number of
times the trip was made, and the experimenter announced that they would then be given several problems that could be solved by this procedure.

Once the teachers and the experimenter answered all questions from the instruction group students, the contingency group students were brought back to the classroom. The students were seated in different parts of the classroom according to their groups and given sets of the same 5 practice problems assembled into a booklet. The problems were a subset of the 10 problems used in Experiment 4, chosen in consultation with the teachers to ensure that the practice would be completed in time. Five problems were deemed sufficient because most students who reached criterion of 3 correct in a row in past experiments did so within 5 problems. Also, mathematical skills of students at this school were expected to be higher than at the school which participated in Experiment 4 (2005/2006 mathematics CST score for the Experiment 5 school: 62%, 2004/2005 mathematics CST score for the Experiment 4 school: 26%, and 2004/2005 California average CST mathematics score: 40%, all according to the California Department of Education).

Students in both groups were instructed verbally and in their booklets to solve correctly three practice problems in a row in order to move on to the next part of the study. Feedback protocol from Experiment 4 was followed here. Once students either met the criterion of 3 problems in a row solved correctly or worked through all 5 practice problems, their practice booklets were collected and they moved on to the next phase of the study.

After practice, all students were given a separate page with questions related to the practice problems. The questions were identical to those used in the post-practice
questionnaire in the Experiment 4. After completing their respective questionnaires, students in both groups moved on to the three test problems. These problems were equivalent with each other but unrelated to the practice problems, and contained irrelevant numeric and story information. The students had to complete all three problems without feedback.

Finally, all subjects who finished the test phase were given a post-test questionnaire similar to that used in Experiment 4.

Figure 10: Average percentage of students who solved all of their attempted transfer problems correctly in Experiment 5.
Results and Discussion

Practice and transfer test. As was expected based on the results of the prior experiments in the series, the instruction group significantly outperformed the contingency group on practice problems (M = 4.07, SD = .97 and M = 4.52, SD = .79, respectively), with t(108.24) = -2.63, p = .01. Unlike in the results from the prior experiments, there was no significant difference between the two groups on average proportions of test problems solved correctly, with M = .90, SD = .21 for the instruction group and M = .95, SD = .17 for the contingency group. This appears to be a ceiling effect, based on the high level of math proficiency of the subjects. According to the California Department of Education, 60% of the 6th grade students at this school in this year are either advanced or proficient at mathematics, as compared to California’s overall 30%. However, the teaching manipulation appeared to have some effect, demonstrated in a small correlation (r = .197, p < .05) between the teaching conditions and the test success defined as solving all test problems a student attempted correctly (Table 4). As seen in Figure 10, a significantly higher proportion of the contingency group students (91%) than the control group students (77%) were able to achieve test success, \( \chi^2(1, N = 113) = 4.4, p < .05 \).

Questionnaires. The answers to the post-practice and post-test questionnaires were coded into variables of “solution,” “context,” “general,” and “nothing” using the Experiment 3 coding system, with an additional variable of “information” that was coded as “1” when the students mentioned the irrelevant information in their answers, and as “0” when they did not. Distribution of answers is presented in Table 3, while relevant correlations are presented in Table 4.
Table 3: Distribution of answers to post-practice and post-test questionnaires in Experiment 5. The answers are grouped by the teaching condition (instruction or control) and question number. Solution, context, general, and nothing categories are coded according to the description in text.

<table>
<thead>
<tr>
<th>Question</th>
<th>Post-Practice</th>
<th>Post-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Instruction</td>
<td>Contingency</td>
</tr>
<tr>
<td>N</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>Q</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Solution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>%</td>
<td>17.0%</td>
<td>45.3%</td>
</tr>
<tr>
<td>Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>33</td>
<td>8</td>
</tr>
<tr>
<td>%</td>
<td>62.3%</td>
<td>15.1%</td>
</tr>
<tr>
<td>Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>%</td>
<td>18.9%</td>
<td>18.9%</td>
</tr>
<tr>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>%</td>
<td>13.2%</td>
<td>28.3%</td>
</tr>
<tr>
<td>Nothing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>%</td>
<td>3.8%</td>
<td>13.2%</td>
</tr>
</tbody>
</table>

Post-practice questionnaire. Whether a student was in the instruction group or the contingency group correlated only with context-related answers on Question 1 (Table 4). That is, students in the contingency group were slightly less likely than students in the instruction group (Table 3) to cite context similarities among the practice problems, $r = -.165$, $p < .05$. A stronger effect for this kind of Question 1 answers existed for reaching the practice criterion in 3 problems or taking more problems to reach it. Out of the students who needed more than 3 practice problems to reach the criterion, 65.2% used a context-related answer, while out of the students who reached the practice criterion in 3 problems only 45.5% used a context-related answer, with $\chi^2(1, 112) = 4.257$, $p < .05$. Thus, once again it appears that the weaker problem-solvers are more likely to focus on context similarities (Silver, 1981).
Table 4: Intercorrelations for the Experiment 5 variables. The questionnaire data are grouped as a yes/no (1/0) for each type of answer, as well as by questionnaire (post-practice or post-test) and question number.
Note: * Correlation is significant at the 0.05 level. ** Correlation is significant at the 0.01 level
<table>
<thead>
<tr>
<th>Solution</th>
<th>Practice</th>
<th>Test</th>
<th>Context</th>
<th>Practice</th>
<th>Test</th>
<th>Information</th>
<th>Practice</th>
<th>Test</th>
<th>General</th>
<th>Practice</th>
<th>Test</th>
<th>Nothing</th>
<th>Practice</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>criterion</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>success</td>
<td>.196*</td>
<td>.197*</td>
<td>.122</td>
<td>-.099</td>
<td>.207*</td>
<td>-.013</td>
<td>-.165</td>
<td>-.074</td>
<td>-.199*</td>
<td>.097</td>
<td>.116</td>
<td>.152</td>
<td>-.011</td>
<td>.064</td>
</tr>
<tr>
<td>criterion at 3</td>
<td>-.065</td>
<td>.099</td>
<td>-.167</td>
<td>-.137</td>
<td>.001</td>
<td>-.195*</td>
<td>.096</td>
<td>.205*</td>
<td>.082</td>
<td>-.081</td>
<td>.079</td>
<td>-.103</td>
<td>-.027</td>
<td>.133</td>
</tr>
<tr>
<td>success</td>
<td>-.066</td>
<td>.104</td>
<td>.260**</td>
<td>.121</td>
<td>-.048</td>
<td>-.062</td>
<td>-.206*</td>
<td>.045</td>
<td>.012</td>
<td>-.027</td>
<td>-.014</td>
<td>-.058</td>
<td>.080</td>
<td>.025</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solution</th>
<th>Practice</th>
<th>Test</th>
<th>Context</th>
<th>Practice</th>
<th>Test</th>
<th>Information</th>
<th>Practice</th>
<th>Test</th>
<th>General</th>
<th>Practice</th>
<th>Test</th>
<th>Nothing</th>
<th>Practice</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>.173</td>
<td>.233*</td>
<td>.045</td>
<td>-.232*</td>
<td>-.008</td>
<td>-.052</td>
<td>.179</td>
<td>-.152</td>
<td>-.121</td>
<td>-.028</td>
<td>-.023</td>
<td>-.127</td>
<td>-.139</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.014</td>
<td>.329**</td>
<td>.142</td>
<td>-.034</td>
<td>.114</td>
<td>-.117</td>
<td>-.121</td>
<td>-.277**</td>
<td>-.068</td>
<td>-.128</td>
<td>-.126</td>
<td>-.408**</td>
<td>-.047</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.303**</td>
<td>-.032</td>
<td>.018</td>
<td>-.531**</td>
<td>-.099</td>
<td>-.084</td>
<td>-.062</td>
<td>-.140</td>
<td>.018</td>
<td>-.075</td>
<td>-.020</td>
<td>-.271**</td>
<td>-.187</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.117</td>
<td>-.010</td>
<td>-.021</td>
<td>-.128</td>
<td>-.154</td>
<td>-.108</td>
<td>-.224*</td>
<td>-.165</td>
<td>-.087</td>
<td>-.266**</td>
<td>-.190</td>
<td>-.520**</td>
<td>.051</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context</th>
<th>Practice</th>
<th>Test</th>
<th>Information</th>
<th>Practice</th>
<th>Test</th>
<th>General</th>
<th>Practice</th>
<th>Test</th>
<th>Nothing</th>
<th>Practice</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>.027</td>
<td>.183</td>
<td>.090</td>
<td>-.354**</td>
<td>-.267**</td>
<td>-.119</td>
<td>-.171</td>
<td>-.333**</td>
<td>-.011</td>
<td>-.105</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.022</td>
<td>-.038</td>
<td>.103</td>
<td>-.100</td>
<td>.172</td>
<td>.075</td>
<td>-.137</td>
<td>-.145</td>
<td>-.095</td>
<td>.065</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.152</td>
<td>-.226*</td>
<td>-.016</td>
<td>-.244*</td>
<td>.011</td>
<td>-.059</td>
<td>.000</td>
<td>-.171</td>
<td>-.054</td>
<td>-.014</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.058</td>
<td>-.058</td>
<td>-.052</td>
<td>-.033</td>
<td>-.067</td>
<td>-.026</td>
<td>-.053</td>
<td>-.020</td>
<td>-.035</td>
<td>-.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information</th>
<th>Practice</th>
<th>Test</th>
<th>General</th>
<th>Practice</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>.334**</td>
<td>.398**</td>
<td>.169</td>
<td>-.204*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.162</td>
<td>.384**</td>
<td>-.023</td>
<td>-.014</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.215*</td>
<td>-.046</td>
<td>.029</td>
<td>-.159</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.020</td>
<td>-.011</td>
<td>-.142</td>
<td>-.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General</th>
<th>Practice</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>.228*</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.102</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.309**</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-.112</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nothing</th>
<th>Practice</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>.030</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Context-related answers for Question 2 were unlikely because of the question make-up, though some students used context references in answers that also mentioned solutions, general rules, or irrelevant information. The question appeared to be more difficult than Question 1, with more students answering it negatively (Table 3). Since about 12% of both instruction and contingency group’s students gave a negative answer, it seemed that this was not directly influenced by a teaching group. Unexpectedly, 95.7% of students who did not reach the criterion in 3 problems and only 83.1% of those who did were able to answer this question, with $\chi^2(1, N = 112) = 4.743, p < .05$. When the teaching group was considered as a factor, it appeared that this effect occurred primarily in the instruction group, with $\chi^2(1, N = 53) = 4.808, p < .05$. Among these students, 96.4% of those who needed more than 3 problems to reach the practice criterion and 76% of those who reached the practice criterion in 3 problems were able to answer Question 2. Given the small number of subjects who gave negative answers (14 total), high mathematical proficiency at the school, and the presence in this set of some Japanese and Korean students who are immersed into an English classroom (Bernardo & Calleja, 2005), it is possible that this result is accidental.

In Experiment 4 the instruction group students were able to recall either general or specific rules taught to them by the teacher. In the present study, among students who reached the practice criterion in 3 problems, those from the instruction group were also more likely than those from the contingency group to mention discarding irrelevant information as a rule on this question, $\chi^2(1, N = 66) = 5.533, p < .05$. 


An interesting issue, of course, is whether verbalizing problem similarities and rules after the practice had any impact on transfer test success. None of the answer variables for post-practice questions correlated directly with test success. However, certain patterns emerged when students’ placement in the instruction or the contingency group was factored in. For students who used a context-related answer on Question 1, group placement did not matter, since about 80% of them in each group were successful on the test. However, among students who did not use a context related answer, 96.8% of students from the contingency group and only 70% of students from the instruction group were successful on the test, with $\chi^2(1, N = 51) = 7.359, p < .01$. Given a significant negative correlation between solution-based and context-related answers on Question 1 (Table 4, $r = -.232, p < .01$), the effect was reversed for solution-related answers. Among those who used a solution-related answer on Question 1, 93.8% of students from the contingency group and only 55.6% of those in the instruction group were successful on the test, with $\chi^2(1, N = 25) = 5.252, p < .05$. Thus, it appears that the ability to see structural similarities and not focus on superficial similarities of the practice problems contributed to the main effect of superiority of the instruction group on the transfer test. Even though instruction and contingency groups did not significantly differ in their mention of solution-related similarities, for the instruction group the answers were most likely a result of what the teachers told them, while the contingency group had to find the answers through their experience. Thus, what they learned during practice helped the contingency group students and hindered the instruction group students during the transfer test.
**Post-test questionnaire.** Only 2 students, both in the contingency group, gave a negative answer to Question 1. As seen in Table 3, significantly more of the contingency group students than the instruction group students listed structural, solution-based problem similarities, with $\chi^2(1, N = 106) = 4.554$, $p < .05$. The reverse was true for context-based similarities, with $\chi^2(1, N = 106) = 4.19$, $p < .05$. This finding offers additional evidence for Experiment 4 results that suggested that the contingency group offers an advantage in problem solving on transfer test through enhancing the subjects’ ability to infer solution rules. It is particularly important here, because students had to map problems in a way that would ignore irrelevant information and extract underlying problem structure.

As was expected based on Experiment 4 results and prior findings (Silver, 1981; Reed, 1987), students who were successful on the transfer test were more than twice as likely as those who were not successful (56.8% and 22.2%, respectively) to refer to solution-based similarities of the problems, $\chi^2(1, N = 106) = 7.64$, $p < .01$. In contrast, they were only half as likely as those who were not successful on the test (25% and 50%, respectively) to refer to context-based similarities of the problems, $\chi^2(1, N = 106) = 4.514$, $p < .05$.

When the placement in the instruction or the contingency group was factored in, it was clear that students in the instruction group who identified solution similarities were dramatically more likely to be successful on the test (90.5%) than those who did not identify solution similarities (64.5%), $\chi^2(1, N = 52) = 4.5$, $p < .05$. There were no significant differences in the contingency group with respect to test success among students who gave or did not give solution-related answers, though it
conformed to the same pattern (94% versus 85%). On the other hand, students in the contingency group who identified context similarities were less likely to be successful on the test (72.7%) than those who did not identify context similarities (95.3%), $\chi^2(1, N = 54) = 5.335, p < .05$. There were no significant differences in the instruction group with respect to test success among students who gave or did not give context-related answers, though it also conformed to the same pattern (68.4% versus 78.8%). The pattern partially corresponds to that for the post-test questionnaire of Experiment 4, though there contingency group students who did not do well on the test had no time to complete the questionnaire. It is interesting that for the contingency group, where only about 9% of students were unsuccessful on the test, it really did not matter as much as for the instruction group whether students could verbalize structural problem similarities. It was more of a hindrance for these students to focus on problem context.

Question 1 was constructed as more open-ended, and so more likely to detect the differences in answers. The questions were deliberately phrased in such a way that the answers to Question 2 on both questionnaires were unlikely to include context references, because that question specifically asked for a rule. Only one subject mentioned problems’ context on the post-test questionnaire. As Table 3 shows, students in the contingency and the instruction groups were nearly equally likely to offer a mathematical, solution-based rule as an answer to this question. There were also no significant differences between groups for any other answers. However, students in the instruction group who were not successful on the test were shown to be more likely to answer with general problem-solving rules than those who were successful on the test (46% versus 18%), $\chi^2(1, N = 52) = 4.137, p < .05$. In the
contingency group, about 20% of both successful and unsuccessful on the test students answered with general problem-solving rules. Thus, it appears that the instruction group students who were having difficulty solving the transfer problems could have been unable to infer specific rules to solve them. This is likely a consequence of learning a rule for the base problems through instruction, which for at least some of the instruction group students caused the Einstellung effect (Luchins, 1942).

**Discussion.** As expected, in the present experiment the instruction group students reached the practice criterion faster than the contingency group students, despite the presence of irrelevant information in the problems. On the transfer test, the contingency group students were more successful that the instruction group students. The successful problem solvers had to be able to identify underlying structure of the problems in order to avoid using the irrelevant information (Low and Over, 1993). Thus, the students who had to learn how to solve the practice problems from a feedback contingency alone were better at learning the new, transfer rule than the students who were given a rule for solving the practice problems.

Though students from the instruction group were more likely than students from the contingency group to state a solution rule for practice problems, this was most likely simply because they remembered the teacher’s instructions. This conclusion is supported by the fact that they were less likely than the contingency group students to identify solution-based, structural similarities of the practice problems. Noting structural similarities of the problems is established in the literature (e. g., Silver, 1981; Hegarty, M., Mayer, R. E., & Monk, C, A., 1995) as a behavior of the more proficient problem-solvers.
After the transfer test, students from the contingency group demonstrated that they were more likely to verbalize solution-based similarities and less likely to verbalise context-based similarities of the transfer problems, than students from the instruction group. Additionally, referring to solution-based similarities and not referring to context-based similarities was correlated with the test success. This is consistent with the prediction, based on Fantino et al. (2003) and the results of Experiments 1 and 4, that rule discovery in practice should lead to a positive transfer effect. It also appeared to increase the likelihood of rule discovery during testing, when no feedback was given.

The hypothesis regarding detecting irrelevant information as a rule working to subject’s advantage was not supported, because the subjects mentioned it as a rule very infrequently. There were no significant correlations of information-related questionnaire answers with either reaching the practice criterion on the first three problems or solving all three test problems correctly, nor were there any effects of the teaching condition or teacher on information-related answers. It is possible, however, that this outcome is an artifact of the way questions were asked. For example, very few students used more than one category in their answer, contributing to a lot of between-categories negative correlations (Table 4). It is not unlikely that irrelevant information was not mentioned by successful problem solvers because they believed that they answered the question completely enough without it. Since there was no direct assessment of this possibility, it remains open for further exploration.

However, it is evident that despite the presence of irrelevant information in the problems, the subjects, especially those in the contingency group, were able to map
the problems’ underlying structure. Since there was a clear advantage to the contingency group in understanding of the transfer problems’ structure, it can be seen as evidence of conceptual knowledge transfer (Rittle-Johnson, 2006). Rittle-Johnson found that neither self-explanation nor the instructional conditions (direct instruction or discovery) improved conceptual transfer with equivalence problems. The present study supplies evidence that contingency based practice may be beneficial in promoting conceptual transfer.

Another interesting issue with regard to irrelevant information is that of gender differences. Low and Over (1993) found that male high school students performed better than female high school students when word problems contained irrelevant information, demonstrating that the boys had a better grasp of the problem structure. When the present sample was analyzed with respect to gender (49 girls, 64 boys), no differences in test performance or mentioning irrelevant information on questionnaires emerged. There were also no significant differences when the effects of teaching manipulation were analyzed controlling for gender. If the ceiling effect is the culprit for the lack of significant difference on transfer test based on the teaching condition, it can also account for the lack of gender difference on this test. However, the results based on answers to questionnaires are not so susceptible, and thus can be said to contradict Low and Over’s claim, at least at the middle school level with a high proportion of mathematically advanced or proficient students in the sample.

Overall, the findings of the present experiment support contingency-based problem solving practice as a method that offers more positive transfer effects to novel problems than direct instruction does, including conceptual transfer. It also appears to
enhance the students’ inference of solution rules as opposed to focusing on surface features of problems, which is known to be a correlate of successful problem solving (Silver, 1981; Reed, 1987). In turn, inferring rules in the context of solving word problems does not impair transfer as suggested by Luchins (1942), and may be a contributor to successful transfer instead.

This study has several limitations which originate in its field nature. First, a possible ceiling effect may have prevented the expression of differences among the groups. Second, because this was a field study relying on participation of the classroom teachers, the outcome may have been affected by their skill and class composition. This can be addressed by bringing students to the laboratory from different schools, but such a method of study would not offer the applied advantages sought after by this series of studies. Finally, the timing constraints of a class period, as in Experiment 3, imposed limits on the number of practice and test problems and the number of questions used, which restricted the amount of adapt available for analysis. However, this factor is just as much a factor in real classroom situations, where the teachers have to accomplish their objectives, such as teaching problem solving, in limited amount of time. Therefore, despite its limitations, this experiment represents an important bridge between academic understanding of transfer issues and bringing more efficient teaching techniques into mathematics classrooms.
Chapter VII

General Discussion

This study found evidence for positive transfer effects associated with contingency-based practice using word problems solvable by a rule different from the target problems’ rule. In a classroom setting, practicing with equivalent base problems while receiving immediate feedback allowed students to learn directly from the consequences of their work on the problems. Though this teaching method led to slower acquisition of base problems than direct instruction supplemented by a worked-out example, it resulted in higher success rates on a transfer test. This finding is novel with respect to the existing literature, which usually investigates analogous transfer where, unlike in the present study, base and target problems share structural and/or contextual similarities. Though this well studied type of transfer is difficult to achieve (Gick & Holyoak, 1980, 1983; Novick & Holyoak, 1991; Reed, 1987), it is best accomplished through instruction or through use of worked out examples (Novick, 1988; Rittle-Johnson, 2006; Silver, 1981). One practical implication of teaching for analogous transfer is that each problem type has to be taught separately, making it difficult for most students to independently attempt solving a problem of a novel type. It can be speculated that this is at least one of the reasons for the common “word problem phobia” lamented by students and teachers alike. Therefore, showing that a contingency based approach to practice, also known as the guided discovery approach (Mayer, 2004), can lead to higher success with a novel problem type than direct instruction, offers a useful technique to educators. In this study, unlike in typical analogous transfer experiments, base and target problems do not share any context or
structural features, yet contingency-based practice with the base problems enhanced both success in solving and understanding of the underlying structure of the target problems.

The main effects of faster criterion achievement during practice for the instruction group and higher rates of test success for the contingency group found in Experiment 1 were supported by the results of Experiments 4 and 5. Notably, this outcome occurred only when the test problems were all solvable by the same rule as each other. Experiment 2 was designed as a control study for Experiment 1, and it showed that neither the instruction group nor the contingency group had any transfer test advantages when all transfer problems were structurally and contextually different from each other.

The results hold important theoretical implications for the debate on the role of rules in problem solving. The Luchinses (Luchins, 1942; Luchins & Luchins, 1950, 1970) defined the Einstellung effect, also known as rigidity or set effect, which interfered with efficient solving of transfer problems. They found that the surface features similarities lead to applying an old rule to the novel problems, usually in situations where for the subjects there was no discernable change in problems from the practice to the test phase. This finding was challenged by Fantino et al (2003). In Luchins’ studies, both instruction and practice with equivalent problems led to subjects’ persistence with an old rule when a new rule had to be applied for successful solution, a finding that did not appear for contingency-based practice in Fantino et al. Likewise, the present studies show that contingency-based practice that allows
students to infer rules does not impair and may enhance their success on novel transfer problems, as long as the novel problems are solvable by the same novel rule.

Experiments 4 and 5 found some evidence that students do infer rules during practice and during the transfer test. Ability to overtly verbalize the solution rules for transfer problems was associated with successful problem solving on the transfer task. In addition, it appeared that when students attended to problem context more than to problem structure, they were more likely to make errors on the transfer test. This is consistent with both Fantino et al.’s (2003) notion that independently inferred rules do not interfere with transfer and Silver’s (1981) finding that successful problem solvers focus on problem structure instead of problem context. This was especially true when practice was solely contingency-based. However, in both Experiment 4 and Experiment 5, students in both groups who demonstrated their knowledge of practice problem solution rules on a questionnaire were not at any disadvantage during the transfer test. Thus, the widely cited Einstellung effect (e.g., Chen, 1999; Grysikou & Weisberg, 2005; Sweller, Mawer, & Howe, 1982) does not appear to occur in classroom situations where word problems are involved and changes in the type of problems to be solved are more distinct than in Luchins’ procedure.

Skinner (1974) asserted that rule-governed learning is not as sensitive to changes in the situation as contingency-based learning, essentially agreeing with the Luchinses (Luchins, 1942; Luchins & Luchins, 1950, 1970). That would mean that on a transfer test contingency group subjects should perform better than instruction group subjects should perform, just as observed in Experiments 1, 4, and 5. Skinner (1969) defined rules as stimuli that specify contingencies, and usually meant that the behavior
controlled by the rules and not directly by the contingencies is less flexible. However, what happened to the contingency group students during practice, as they learned through feedback to solve the set of equivalent problems? If they learned how to solve these problems, it can be said that they learned a rule and their behavior from then on became rule-governed. Overall, though contingency group subjects enjoyed a higher success rate on the test, it was not because they did not know what the rules were. In fact, it was evident that the subjects who inferred appropriate rules were more successful on the transfer test. In Experiments 1 and 2 subjects in control groups practiced, with feedback, with changing-rule problems, and did not do as well as the contingency group subjects on the transfer tests. Not only there is no harm in learning rules (Fantino et al., 2003), but also the contingency-based learning appears to be most efficient for transfer when it is designed for inferring the rules.

This is an important implication for classroom instruction. It suggests that if teachers offer their students practice with sets of equivalent word problems instead of direct instruction for each set of word problems, then the students will become more proficient at identifying solution rules of novel sets of word problems. Since word problems are important for curriculum and life outside of classrooms, instructional techniques that focus on teaching for transfer are much desired by educators (e.g., Fuchs, Fuchs, Hamlett, & Appleton, 2002; Weber-Russel & LeBlanc, 2004; Xin, Jitendra, & Deatline-Buchman, 2005). The contingency procedure used in this study resulted in significant positive transfer effects over the direct instruction procedure. There was also clear evidence of an advantage in conceptual understanding of the problems. In particular, in Experiment 5 students in the contingency group were more
likely to be successful with problems that included irrelevant information than students in the instruction group. Thus, Rittle-Johnson’s (2006) dismissal of different effects of these two instructional techniques (contingency-based learning, or invention, and direct instruction) on transfer appears to be premature. The reason for the discrepancy may potentially lie in the difference between the study materials. Mayer (1982) demonstrated that solving mathematical equation problems such as those in Rittle-Johnson study, and solving mathematical word problems, though not involving a story line such as those used in the present experiments, are qualitatively different processes. That is, different strategies may need to be used by subjects who work on equations as opposed to subjects who work on story problems, thus making different instructional techniques more successful.

This study did find limited support for benefits of self-explanation (Atkinson, Renkl, & Merrill, 2003; Chi, Bassoc, Lewis, Reimann, & Glasser, 1989; Rittle-Johnson, 2006), where subjects are asked to infer rules or explain solutions much as they had to do in Experiments 4 and 5 questionnaires. There was a positive transfer effect of taking the pre-test questionnaire in Experiment 4. However, it appears that unlike in Rittle-Johnson’s recent study, the self-explanation effect held only for subjects who did not have rule instruction. Moreover, there was no positive transfer effect for engaging in self-explanation on a post-practice questionnaire. Thus, the advantage conferred by self-explanation was secondary to the advantage conferred by the contingency-based practice. Potentially, this has to do with the age of the subjects, language issues, and the phrasing of the questions which could have all contributed to
the lack of influence of the pre-test and post-practice questionnaire on transfer test success.

This study had a number of limitations associated with the fact that it was conducted in the classroom setting. Levin (1993) cautioned classroom researchers about compromises that have to be made in classroom settings, referring to them as “methodologically messy places” (p. 4). Indeed, a quest for ecological validity of this work caused this study to be conducted not only within real-world classroom settings, but within time constraints of class periods as scheduled at participating schools. Also, classroom teachers directly participated in the study, thus introducing an additional variable that might have influenced the data. It turned out that the teachers affected the results, though only during the practice phase when they were giving instruction to some of the students.

Another limitation related to the real-life classroom setting was the preexisting mathematical and problem solving proficiency of the participants. Researchers often separate their subjects by ability (Hegarty, Mayer, & Monk, 1995) or chose to study only advanced students, creating the materials accordingly (Cooper & Sweller, 1987). However, this study did not deliberately sort the subjects this way because it was not concerned with determining how advanced students differ from less advanced students in problem solving, which is a widely investigated area (e.g., Hegarty, Mayer, & Monk, 1995; Silver, 1981). The participant choice depended on schools agreeing to participate in the study. Thus, two schools with mostly proficient or advanced students were used for Experiments 4 and 5, leading to ceiling effect obscuring the findings. If more challenging materials were used, it would be more likely that the data would
resemble those from other experiments more, but then there was a high likelihood that a lot of the participants would not complete the tasks on time.

This last issue, the time constraints, affected Experiment 4 most of all. The main effect of the study, the positive transfer effect of contingency-based practice, still appears to be a valid finding. However, it was weakened in Experiment 4 by a high number of participants not finishing all of the study tasks within the time allowed by a class period. Given the generally low level of mathematical ability at the participating school, this obstacle could not be overcome other than by using additional time, which was not possible in the study. However, this situation also raises a question of benefits of contingency based, guided discovery learning for low ability students within a classroom setting. These students as a population are not usually targeted by psychological research, perhaps out of sensitivity to the effect of labeling the students. However, students with learning disabilities affecting learning of mathematics were studied in context of mathematical problem solving (Fuchs et al., 2002; Xin et al., 2005). These studies, with findings mostly conforming to the results of the research with regular populations, were done over multiple sessions, as was a discovery learning study done by Muthukrishna and Borkowski (1995) with typical elementary school students. The attempt in Experiment 3 to incorporate group practice used by Muthukrishna and Borkowski was inconclusive mainly due to the time constraints which could not be overcome by replication.

Nevertheless, given the need to provide educators with scientific research suitable for classroom implementation (Sternberg & Lyon, 2002), and the decline of educational interventional research with children in the last decade (Hsieh et al.,
2005), this study aims to make an important contribution both to education and to psychology of learning. Sternberg and Lyon point out that “[L]aboratory research can be useful. But schools are extremely complex, multifaceted environments, and the transfer of phenomena and interventions from laboratory to school…cannot be assured” (¶ 8). Directly in the classrooms, allowing students to practice solving a set of equivalent problems with feedback repeatedly provided them with an advantage on a transfer test as compared to giving them instructions on solving the practice problems. Thus, conducting this study of transfer in the classrooms made its results themselves much more transferable, though at the expense of controlling extraneous variables. Still, even though B. F. Skinner might have disapproved of this approach itself, the study sheds additional light on the discussion of rule-governed versus contingency-based learning that he originated, along with other relevant phenomena. It supported Fantino et al.’s (2003) contention that problem-solving involving rules does not have to be inflexible. In fact, independently inferring solution-based rules during practice enhanced transfer test performance which required different rules. Also, subjects who inferred solution rules in the process of taking the transfer test were the subjects who did best on that test. Thus, contingency-based practice which allows the participants to infer rules can be recommended as a way of teaching for transfer of mathematical word problems.
Appendix A
Practice Worksheet Used for all Instruction Groups

**Practice Problem**

Five times a week Marin is going to a guitar workshop that is 3 miles away from her home. She likes to walk there and back. How many miles of walking does that involve every week?

2 * 3 = 6 miles

6 * 5 = 30 miles

Answer: 30 miles
Appendix B

Same-Rule Practice Problems

1. Every day Emily visits her grandmother, who lives 3 miles away. Emily takes her bike to go there and come back. How many miles a week does Emily have to bike?

2. Elderly Mrs. Smith wants to move in a new apartment. The entrance to that apartment is atop a flight of 16 steps. Mrs. Smith wants to know how many steps up and down would she have to climb a week if she will go for her daily walk.

3. When Alex went to college, he moved from Los Angeles to San Diego, about 120 miles away from his parents’ home. He drives to Los Angeles to visit them 3 times a month. How many miles of driving every month does Alex do to visit his parents and come back to San Diego?

4. Jonathan swam from shallow to the deep end of a pool and back 15 times. The pool is 20 feet long. How many feet did Jonathan swim?

5. Every Friday Jennifer walks to the library to exchange her books and then walks back home along the same road. The library is about 1.5 miles away from Jenny’s home. Jenny did this every Friday during the last year (there are 52 weeks in a year). How many miles of walking did going to the library involve for Jenny in 1999?

6. Louis visits his best friend Arthur’s house on average 10 times a month. Louis’ home is 3 miles away from Arthur’s, and Louis usually bikes to get there and back. How many miles, on average, does Louis bike every month because he visits his friend?

7. The Salazar family visits their grandparents about 6 times a year. The grandparents live 230 miles away from the Salazars’ home. How many miles of traveling do the Salazar’s do every year to visit the grandparents and come back home?

8. A boat of Catalina Company makes 4 trips a day from Dana Point to Avalon and back. The distance between San Pedro and Avalon is about 25.5 miles. How many miles a day does this boat travel by water?

9. Tom and Jerry go kayaking on the bay 2 times every month. They usually begin at a kayak rental place, paddle to the same spot in Seaport Village, and return by the same route. It is 5 miles by boat from the rental shop to the Seaport Village. How many miles do Tom and Jerry kayak every month?

10. Sandy tries to jog for exercise every day, but she can actually do it about 20 times a month. She usually jogs from her home to the park and back. The park is 3.5 miles away from Sandy’s home. How many miles does Sandy jog every month?

11. Michael used to drive to the gym 4 times a week. He realized that if he were to walk to and from the gym instead, he would get additional exercise. The gym is 2.3 miles away from Michael’s home. How many miles of additional walking would this involve every week?
12. Celine lives 2 miles away from her school. If she walks to and from school, how many miles of walking does she have to do for this purpose during one school week (5 days)?

13. Mrs. Morgan is a bus driver on a city line. Her route is 28 miles one way and is all along the same major street. Usually Mrs. Morgan makes 6 round trips per day. How many miles does she drive in her bus every day?

14. Stephanie’s soccer practice is 8.5 miles away from her home. Her mother drives her there and back 3 times a week. How many miles of driving does Stephanie’s mother do every week because of this?

15. A ski lift line is about 150 feet long. Each chair makes 80 roundtrips during the lift’s operating hours. How many feet would you travel if you were to sit in one of the chairs all day?
Appendix C

Changing-Rule Practice Problems

1. Joe and his sister played a computer game. Joe earned 157 points, but his sister earned twice as many. How many points did Joe’s sister earn?

2. Jim called on the phone his older brother who is away at college. They talked for 23 minutes, and the telephone bill charge for this call was $1.84. How much does a minute of phone time cost?

3. The “Go Far!” bus company is considering a 15% increase in the bus fare. If a bus ticket is now $25.00, how much would it cost after the increase?

4. Erica does 45 sit-ups in minute. If she is able to keep up this rate, how many sit-ups would she be able to do in 5 minutes?

5. There will be three astronauts abroad a space shuttle. They plan to spend 40 days in orbit. Every day these three people drink 1.5 gallons of water altogether. How much water should they bring, if they want to bring twice what they actually need in case of emergency?

6. A dress that used to cost $20 is now on 25% off sale. How much does it cost now?

7. The Hernandez family’s car uses up about a gallon of gasoline for every 32 miles of driving. If they are planning to go on a 288 miles long trip, how much gasoline will they use?

8. Tony used to wash his dad’s car since he was seven. It usually took him 90 minutes to do the job. Now that he is twelve, he does it 50% faster. How long does it take Tony to wash the car now?

9. Janet is 6 years younger than her sister Karen, who is 29 years younger than their mother. Janet just finished making a present for her mother’s 40th birthday. How old is Janet?

10. There are 32 students in Mrs. Nguen’s class. Today 25% of them are home sick with the flu. How many students are absent?

11. Andrea was helping to make wedding favors for her cousin’s wedding. So far she made 75 of them, while her cousin made 135. There will be 250 guests at this wedding. How many more favors do Andrea and her cousin need to make so that every guest gets one?

12. Brothers Jake and Eli were saving to buy a telescope together. Jake contributed $45.30. Eli saved $23.25 less than Jake did. How much money did the boys save together?

13. Judy reads 50 pages an hour. How much time will she need to finish a 350 pages long book?
14. Jessica was 60 inches tall when school got out in June, but she grew 3 inches over the summer. By how many percent of her old height did her height increase?

15. Natalie likes to write letters to her friends. She mails about 15 letters every month. A stamp costs 33 cents. How much money does Natalie spend on stamps every month, on average?
Appendix D

Test Problems Used in Experiments 1 and 3

Set 1

1. Ramona’s French teacher wants her students to read 45 pages of a French book every week. If Ramona spends 9 hours a week reading her French book, how many minutes does she spend per page, on average?

2. Gregory is writing thank-you notes for his birthday presents. He has 12 notes to write, and wants to be done with them in 1.5 hours. How many minutes, on average, does he have to write each note?

3. Olga was participating in a cookie-eating contest at a fair. She ate 36 large cookies in 15 minutes (and became very, very full). How many seconds did it take her to eat each cookie, on average?

Set 2

1. A gymnastics group decided to raise money for their costumes by setting up a car wash. They have decided to work for 6 hours, and estimated that they will be able to wash about 25 cars during this time. How many minutes do they plan to spend washing each car, on average?

2. Monica and Stanley are making sandwiches to serve at their party. They want to make 150 sandwiches in remaining 3 hours before the guests arrive. How many minutes, on average, can they spend per sandwich?

3. Tanya needs to memorize 48 Spanish words, but she only has 2 hours to do it. How many minutes per word does she have, on average?
Appendix E

Test Problems Used in Experiments 4 and 5

1. The students at Lincoln School are going on a picnic at Crystal Lake. There will be 339 students and 13 teachers going to the picnic. There will also be 27 parents going to the picnic, who will all ride in their own cars. Busses to carry all students and teachers can each take 44 people. How many busses will be needed?

2. Maria and her little sister Janet are making cupcakes for a bake sale, and packing them into boxes. Maria made 134 cupcakes, and Janet made 82. Each box can hold 24 cupcakes, which will be sold for $1.50 each. How many boxes do they need to pack all of the cupcakes?

3. Farmer Pickles needs to take his hens and roosters to his new henhouse. He has 230 hens and 15 roosters. He also has 165 rabbits that will need to be moved later. His truck can carry 49 birds on each trip. How many trips should Farmer Pickles make to move all of his birds?
References


