2000

Effects of cumulative practice on mathematics problem -solving behavior

Kristin Hobbs Hazlett
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Hazlett, Kristin Hobbs, "Effects of cumulative practice on mathematics problem -solving behavior" (2000). *Graduate Theses, Dissertations, and Problem Reports*. 3178.
<https://researchrepository.wvu.edu/etd/3178>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Effects of Cumulative Practice on Mathematics Problem-Solving Behavior

Kristin H. Hazlett

**Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Doctor of Philosophy
in
Psychology**

**Philip N. Chase, Ph.D., Chair
Barry Edelstein, Ph.D.
Michael Perone, Ph.D.
David Schaal, Ph.D.
Joseph W. Wilder, Ph.D.**

Department of Psychology

**Morgantown, West Virginia
1999**

**Keywords: mathematics education, problem solving,
cumulative practice, behavioral education**

Copyright 1999 Kristin H. Hazlett

ABSTRACT

Effects of Cumulative Practice on Mathematics Problem-Solving Behavior

Kristin H. Hazlett

Mathematics education has long been in need of improved methods of instruction, particularly in the area of problem-solving skills. This study compared three methods of training rules about laws of exponents and order of operations. All three training methods used the same mastery criterion for training each rule and included the same number of practice trials during review sessions that preceded each test. The difference between conditions involved what types of problems were presented during the reviews. For each review session, the cumulative group ($n = 11$) practiced 50 problems covering all rules learned up to that review. The simple review group ($n = 11$) practiced 50 problems on one previous rule, and the extra practice group ($n = 11$) practiced 50 more problems of the same rule they had just mastered. Tests were administered after each review.

Though no initial differences existed between groups on any measure, the last test revealed that the cumulative group scored significantly higher than the other groups on items that involved novel applications of the individual rules. Moreover, the cumulative group outperformed the other two groups on untrained, complex problem-solving tasks that required novel combinations of the individual rules. In addition, the cumulative group performed the problem-solving tasks at a significantly faster rate than the other groups. There were no statistical differences among groups on a retention test, however, which was partially due to a reduction in sample size, as well as increases in variability of performance within groups.

Overall, the findings support the viewpoints of behavioral educators that mastery of component skills facilitates performance on higher-level skills and that novel behavior is fundamentally related to its component parts. The results also extend the research of behavioral educators by removing the confounded variables of simple review and extra practice found in previous studies and by showing the effects of cumulative practice on problem-solving behavior. Finally, the results suggest that an approach to training problem solving similar to the one presented in this study may yield higher levels of success than methods used by traditional mathematics educators.

This research was partially funded by dissertation grants from the Office of Academic Affairs and the Psychology Alumni Fund at West Virginia University.

ACKNOWLEDGEMENTS

This research was made possible by the hard work of many assistants including Carolina Aquilera, Karrie Wagoner, Brittany Wintry, Jenny Mankins, Karen Dotson, Shane Taylor, Gilbert Miller, and Bridget Hayes.

I would especially like to thank my adviser and committee chair, Dr. Philip N. Chase, for all his support and dedication to this project, as well as to the general pursuit of educational reform. Phil has been an exemplary model of what an educator and scholar should be. He has taught me much more than the fundamentals of behavior analysis and its applications to education; he has shown me how to use behavior analysis effectively to improve people's lives. Through his personable, energetic, insightful, and nurturing verbal and nonverbal behaviors, he has demonstrated how to strive for excellence as a scientist and to enjoy oneself along the way. I will always be grateful for his mentorship and friendship.

I also would like to thank my other committee members, Dr. Barry Edelstein, Dr. David Schaal, Dr. Michael Perone, and Dr. Joseph Wilder, for their time, energy, and ideas that have gone into this research. Joe Wilder has been a critical support of my mathematics research conducted over the past three years. In addition, Dr. Kent Parker and Dr. Stan Cohen have provided helpful information regarding statistical analyses.

I am most certainly indebted to my supportive family, whose love and prayers have been an encouragement from the beginning. Furthermore, just for the record, I would like my Dad to note that I have fulfilled his request in regards to my education. Years ago he commented that although I would likely be the first daughter to marry, I would do well to complete my Ph.D. first. All I can say, Dad, is—no sweat. Four days to spare.

I want to acknowledge and dedicate this work to the one who has been my motivation to finish in the shortest amount of time possible, and the one who would love me just the same even if I never finished at all. Bud, you are worth every sleepless night that I spent on my research, trying to shorten the time that we were apart. Thank you, and I will love you forever.

And, finally, my extreme behavioral persistence may be noted as I fully acknowledge and give all honor, glory, and praise to God, whom I look to as the Author of all laws of behavior and the only true Wisdom that allows me to understand them. I am very grateful for the gift of my education, the strength to endure it, and a fuller understanding of His nature as it is expressed through many beautiful and orderly principles of behavior.

TABLE OF CONTENTS

List of Tables.....	v
List of Figures.....	vi
Chapter 1. Introduction.....	1
Chapter 2. Method.....	13
Chapter 3. Results.....	18
Chapter 4. Discussion.....	26
Chapter 5. Conclusions.....	32
References.....	33
Appendix A: Subject Characteristics.....	37
Appendix B: Sample Practice Worksheets.....	40
Appendix C: Sample Test.....	42
Appendix D: Informed Consent Form.....	44
Appendix E: Instructions.....	46
Appendix F: Feedback Exceptions.....	48
Curriculum Vita.....	49

LIST OF TABLES

Table 2.1: Practice Sequence for each Condition.....	15
Table 3.1: Summary of Pretest Results.....	18

LIST OF FIGURES

Figure 3.1: Accuracy of Application.....	20
Figure 3.2: Rate of Application.....	21
Figure 3.3: Accuracy of Problem Solving.....	22
Figure 3.4: Proportional Problem-Solving Performance.....	23
Figure 3.5: Rate of Problem Solving.....	24
Figure 3.6: Accuracy on Retention Test.....	25
Figure 3.7: Rate on Retention Test.....	25

Chapter 1. Introduction

Beginning with the first international mathematics achievement test administered to American students more than thirty-five years ago, the problem of poor mathematics performance has emerged as an issue of national concern. In the 1960's, tests given by the International Project for the Evaluation of Educational Achievement (IEA) compared the rankings of students with those of their international peers. Findings from the study revealed that American 13-year-olds performed below the international average in arithmetic, algebra, and geometry. High school seniors ranked at the bottom of the list with overall mean scores about one standard deviation below the international average. A second IEA study conducted in the early 1980s reported that American seniors in the top 5% of their classes had only average international scores in algebra and calculus, which translated to scores around the 30th percentile in Japan and 50th percentile in England (Brown, 1996; Geary, 1996).

The National Assessment of Educational Progress (NAEP) also conducted two studies of American mathematics students in 1972-1973 and 1977-1978. The results of these studies revealed a serious deficit in problem-solving skills (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1980b). Moreover, 9 and 13 year olds appeared to decline in problem-solving and application skills over the five-year period between the studies (Carpenter, Kepner, Corbitt, Lindquist, & Reys 1980a.) Based on these findings, the National Council for Teachers of Mathematics (NCTM) recommended that problem solving be the main focus of mathematics instruction in the 1980s.

About this time, technological innovations like powerful calculators and personal computers started to become available for classroom use. Thus, math instruction had a new impetus for the shift away from basic skills towards problem-solving skills. Calculators could perform the basic skills, so higher-level skills could be the focus of instruction (Leitzel, 1989). Despite this shift in instructional focus, studies conducted in the late 80s still revealed deficits in problem-solving and "higher-level processing" skills, according to the Working Groups of the Commission on Standards for School Mathematics (as cited in Rech, Juhler, & Johnson, 1995). Moreover, basic skills were not learned. At the end of the 1980's, the National Assessment of Educational Progress reported that more than 2,800,000 thirteen-year-olds (about 85%) and 1,500,000 seventeen-year-olds (about 50%) could not reliably perform computations with decimals, fractions, and percents, or solve simple mathematical equations (Anrig & Lapointe, 1989). Thus, despite the problems in mathematics education at every skill level, teaching problem-solving skills was again advocated in the standards and philosophies adopted by the National Council of Teachers of Mathematics in 1989.

Now, exactly two decades after the initial shift in instructional emphasis, the United States still lags behind other countries in mathematics performance. According to the most recent international statistics, American eighth-grade students spent more time in class and received more homework than Japanese and German students, and yet could not compete with these peers on an achievement test administered in 41 countries. In fact, America's average scores were below the international average, while the top 10% of U.S. students performed at the level of the *average* student in Singapore, the world leader (Wingert, 1996).

According to the eighth and most recent national mathematics assessment administered by the National Assessment of Education Progress in 1996, American twelfth graders performed no better than they did on the first assessment, administered over twenty-five years ago (Vanneman, 1998). Students in the fourth and eighth grades did, in fact, score significantly higher than on the 1973 test (Vanneman, 1998). Despite these significant improvements, however, 38 – 39% of the students performed at a below basic level of mathematics proficiency for their grade level on the 1996 test (Reese, Miller, Mazzeo, & Dossey, 1997). These general statistics become quite alarming when considering a few specific examples from the 1996 results. That is, 70% of fourth graders could not do arithmetic with whole numbers and solve problems that required only one manipulation. Seventy-nine percent of eighth graders and 40% of seniors could not compute with decimals, fractions, and percents, recognize geometric figures, and solve simple equations. Moreover, beginning algebra and problems requiring more than one manipulation could not be performed by 93% of American seniors, an alarming statistic that is no better than the score reported in 1978 (Campbell, Voelkl, & Donahue, 1997).

It is apparent that the emphasis on problem-solving instruction, as advocated by the NCTM, was not a panacea for the crisis in mathematics education. Not only did the approach fail to produce successful higher-level skills, but it also left a critical core of basic skills unmastered by a large percentage of American students. It seems hasty, however, to abandon entirely the search for effective methods of training problem-solving skills. These skills are obviously important for success both in the classroom and in everyday life. Perhaps, then, a more successful approach could be developed that focused on training basic, or component, skills in a manner that led to the formation of higher-level skills. In order to develop such a program of instruction, it may be necessary first to reexamine what is meant by the term “problem solving.”

Problem Solving

Misconceptions. Before offering a clear definition of problem solving, it is helpful to mention a few things that problem solving is not. First of all, problem solving is not a good description of every behavior trained in mathematics. This misconception may arise from the ambiguities of the English language. That is, almost everything in mathematics is called a “problem” to solve, whether solving simple addition problems or complex calculus problems. This characterization of problem solving is not specific enough to be useful, however. Second, problem solving is not functionally equivalent to word problems. Because word problems may offer a good example of a type of problem-solving skill, it is easy to form the misconception that training problem solving means training story problems. Indeed, studies that have addressed “problem solving” have often used word problems as the subject matter (e.g., Darch, Carnine, & Gersten, 1984; Moore & Carnine, 1989). This characterization, however, does not adequately pinpoint the critical features of problem solving.

Operational definition. Problem solving can be operationally defined as emitting a synthesis of responses that has never been reinforced as a unit (i.e., the solution) in a situation for which no previously reinforced response will be reinforced (i.e., the problem). Skinner (1966) described the extinction process of previously reinforced responses as leading to a series of responses that sequentially change the situation until the novel response, or the “solution” to the problem, is emitted. The consequence of each behavior emitted forms a discriminative stimulus for another behavior until the solution occurs.

A similar characterization of problem solving has been provided by behavioral educators. Becker, Engelmann, & Thomas (1975) described the process of academic problem solving as a “series of concepts and operations [that are] chained together in a problem solution” (p. 114). The appropriate responses to each concept and operation are under stimulus control such that they can be recombined into many different chains of behavior. In accordance with Skinner’s definition, then, these recombinations would result from the discriminative control over one response that is produced by the emission of a previous response. Thus, according to Becker et al., problem solving occurs when concepts and operations are used in novel combinations to produce a solution to a problem situation that has not been previously encountered.

It is important to note here that the novelty of the behaviors and the situation are based on an individual’s behavioral history; what is novel for one individual may have occurred repeatedly for another. Thus, it becomes necessary to specify that the event is novel *for that organism* when identifying a process as problem solving.

As applied to mathematics, problem-solving responses that lead to a novel solution may take the form of a sequence, or chain, of previously learned responses (e.g., formulating a step-by-step geometric proof based on previously learned theorems; see Epstein, 1985, 1987), or a combination of previously learned responses (e.g., using knowledge of the values of coins and knowledge of addition facts to compute a total amount of money without direct instruction on the task; see Johnson & Layng, 1992, 1994). Explicit instruction on the novel solution is never provided, and thus instruction from previous situations must be sufficient to produce a solution in the novel context. Therefore, training problem solving implies training skills in such a way that they can be integrated to form novel combinations that meet the demands of a novel situation. This involves establishing tight stimulus control over correct responses on component skills (i.e., training component skills to mastery) and teaching strategies (cued by discriminative stimuli) for how to combine the component skills to produce a problem solution (Becker et al., 1975; Darch et al., 1984; Moore & Carnine, 1989).

Defining problem solving in this way suggests that there may be at least two reasons that past efforts to train problem solving have failed to raise general achievement levels of American students. First of all, classroom methods that have emphasized “higher-level” skills may have failed to provide adequate training on the component skills (i.e., both basic number skills and component skills such as problem-solving strategies) that are fundamental prerequisites to solving problems in novel situations. Indeed, test scores have shown that basic skills have not been mastered (Campbell et al., 1997). Moreover, explicit instruction in component skills such as problem-solving strategies has been shown to very important in solving some kinds of problems (Darch et al., 1984; Xin & Jitendra, 1999) and would likely lead to improved scores if trained adequately.

A second major pitfall of problem-solving instruction, however, may be the failure to train for novel situations. A teacher may provide explicit instruction on a specific problem (e.g., how to solve one version of a distance-time word problem) such that one integrated, complex response is trained. Using this approach, however, does not guarantee that the component skills utilized to solve the “higher-level” problem will be available in students’ response repertoires as individual operants that can be recombined and applied to novel contexts such as achievement tests or situations in everyday life. Indeed, students have been reported to have the most

difficulty with test items presenting problems in novel, complex situations (Kouba, Brown, Carpenter, Lindquist, Silver, & Swafford, 1988). Therefore, it seems that instead of directly teaching specific types of novel problems, it should be the teacher's goal to provide instruction that prepares students to combine component skills in multiple ways to solve any number of novel types of problems that rely on those components skills.

Though it appears from the latest draft of the NCTM's 2000 standards (to be released in April 2000) that educators currently agree with this goal of training for novelty, it is unclear whether effective methods of training for novel problem-solving situations are understood. In the section of the NCTM's document that outlines the role of the teacher in supporting problem solving at grades 9–12, the opening sentence states that "the main responsibility of the teacher is to establish a climate conducive to the development of students' productive mathematical dispositions" (NCTM, Standards: Grades 9–12: Problem Solving). The section goes on to discuss rewarding perseverance and encouraging exploration. While these behaviors may be useful to some degree, they do not necessarily teach the students any content.

The document subsequently discusses the importance of teaching problem-solving strategies and providing opportunities for "comparison of methods, identification of connections, and discussion of planning, monitoring, and adjusting choices during work on the problem" (NCTM, Standards: Grades 9–12: Problem Solving). Though the document suggests some ways of carrying out these guidelines (e.g., asking students certain types of questions and using problems that have many possible solutions), it still remains incomplete and imprecise in describing how to train these skills. Consequently, even teachers who try to follow the guidelines are to some extent forced to rely on their own knowledge and experience, as opposed to research-based methods of effectively teaching problem solving. Therefore, it is probable that individual educators may fail to provide sufficient training on component skills, or may emphasize training on specific types of application problems, or both, which is likely to lead to poor instruction. The end result, then, may be low scores on both basic skills and problem-solving skills because the basic skills would not be trained, and the training on problem-solving skills would not transfer to novel situations. Indeed, these unfortunate outcomes describe well the current state of affairs of mathematics education, as evidenced by the past twenty years of instructional emphasis on "problem solving" coupled with the perpetually poor performance on all levels of international and national achievement tests.

It seems, then, that a more effective approach to training problem solving may be to train component skills so that they can be combined in many novel ways to produce complex solutions to novel task demands. In this way, performance on both basic and higher-level skills can be trained simultaneously. Moreover, problem-solving skills become transferable to novel situations. Overall, then, the more effective method of instruction should be judged as that which leads to mastery of component skills plus accuracy on novel tasks with no additional instruction. This ambitious outcome, however, is not a guaranteed result of teaching many individual problems. Thus, what will prove to transform the teaching of *problems* to the teaching of *problem solving* is the use of methods of instruction designed to produce novel behavior.

Novel behavior. Novel behavior that results from the integration, application, and combination of adequately trained skills has been called various names, such as generalization (Streifel, Bryan, & Aikins, 1974; Striefel & Wetherby, 1973; Streifel, Wetherby, & Karlan,

1976), recombinative generalization (Goldstein, 1983; Goldstein, Angelo, & Wetherby, 1987; Goldstein & Mousetis, 1989), contingency adduction (Andronis, Layng, & Goldiamond, 1997; Binder, 1996; Johnson & Layng, 1992, 1994), and productivity (Catania, 1980; Catania & Cerutti, 1986). Regardless of the terminology, the principle remains the same: instruction is provided on a given set of behaviors or skills so that novel tasks requiring a combination of these skills can be performed correctly. For example, Johnson and Layng (1994) have found that first teaching word problems involving whole numbers and then teaching fraction skills is sufficient to produce accurate performance on word problems with fractions, with no explicit instruction on the word problems containing fractions.

Studies using subject matter other than mathematics have also produced comparable findings. Research involving miniature linguistic systems, for example, has found that training a minimum number of nonsense-word combinations that are functionally equivalent to nouns and verbs is sufficient to generate many possible two-stimuli combinations without explicit training (e.g., Goldstein, 1983; Goldstein et al., 1987). Using this type of procedure, Streifel, Wetherby, and Karlan (1976) identified a training set of 12 nouns and 12 verbs that comprised 144 noun-verb instructions. They taught the first 12 noun-verb instructions by training one of the 12 verbs to criterion with each of the 12 nouns. They then trained the next 12 noun-verb instructions by training the second verb to criterion with each of the 12 nouns in a different order. This procedure was continued with additional verbs and was intermixed with probe trials that tested for correct performance on instructions that had not yet been trained. After a number of verbs had been trained with each of the 12 nouns, subjects were able to respond correctly to 11 novel noun-verb instructions involving the verb currently being trained after only 1 noun-verb instruction had been trained with that verb. Thus, the procedure successfully produced a type of “verbal problem-solving behavior.”

Using a more elaborate training procedure designed to produce sequences of three stimuli, Ellenwood, Chase, and Madden (1999) found that training 12 stimulus pairs, 2 three-stimuli sequences, and 6 additional stimulus relations is sufficient to produce 3 eight-stimuli classes and 512 sequences. This same principle also has been used to teach productive language skills to children with severe mental retardation (e.g., Goldstein & Mousetis, 1989).

One obvious advantage of adopting this approach to teaching problem solving is the built-in efficiency of the instruction. Instead of teachers attempting to cover every skill and its application, they can simply teach a core set of skills and strategies for their combination that will produce novel combinations and applications without any further training. Thus, the maximum learning can be gained from the minimum teaching. Alessi (1987) described this principle as training a generative set of skills to produce a universal set of skills. He referred to the ratio of the universal set to the generative set as the “generative power.” For example, teaching 655 morphographs (the smallest meaningful spelling unit) and some rules of combination can enable students to spell at least 10,000 words (Carnine & Becker, 1982). This produces a generative power of about 15:1. Robinson and Hesse (as cited in Carnine & Becker, 1982) tested this method of spelling instruction using a textbook called Morphographic Spelling (Dixon, 1980). One year after the training was completed, students were able to spell both trained words, as well as novel words never explicitly taught, without any decline in performance.

Examples of performance such as this make a strong case for basing problem-solving instruction on component skills that maximize learning and lead to accurate performance in novel situations. For many topics and subject matters, component skills have already been identified, making it easy to apply this principle. For example, teaching phonics and rules for blending will enable students to read novel words (Carnine, 1977). Teaching the meanings of each prefix and base word and how to convert between them will lead to accurate performance on novel metric conversions (Alessi, 1987). Teaching Latin prefixes, suffixes, and roots of medical vocabulary will lead to accurate identification of untrained terms (Ellenwood & Chase, 1997).

Though the identification of component skills is a powerful tool for designing successful problem-solving instruction, alone it is not sufficient to produce competent problem-solving skills. The choice of what to teach is only half the task of teaching. The other half involves the method of instruction. A potentially successful curriculum will fail unless presented to students in an appropriate manner.

Behavioral educators have identified several critical variables that relate to the presentation of instruction. For example, when teaching a new concept, it is important to present positive and negative examples of the concept, varying any irrelevant features (Engelmann & Carnine, 1982; Carnine et al., 1982; Tiemann & Markle, 1990). Moreover, immediate, individualized, and corrective feedback is critical to the success of all behavioral methods of instruction including Programmed Instruction (e.g., Markle, 1983; Skinner, 1954, 1968), Personalized System of Instruction (e.g., Keller, 1968; Kulik, Jaksa, and Kulik, 1978), Direct Instruction (e.g., Engelmann, Becker, Carnine, & Gersten, 1988; Engelmann & Carnine, 1982; Moore, 1986), and Precision Teaching (e.g., Beck & Clement, 1991; Lindsley, 1992). One of the most important instructional variables, however, is the type of practice used during training. There is general agreement that skills should be practiced to mastery, though some educators define this in terms of accuracy, and some in terms of fluency, or accuracy plus rate (see discussion in Binder, 1992). There is also sufficient support among educators for practicing skills in learning hierarchies or skill sequences (Carnine, Jones, & Dixon, 1994; Englemann & Carnine, 1982; Johnson & Layng, 1992, 1994; Moore, 1986; Resnick, Wang, & Kaplan, 1974; Tiemann & Markle, 1990). When skills are taught in sequences, a common addition to the training is some type of review procedure integrated into the learning of new skills. Several researchers have advocated a review procedure that relies on cumulative practice.

Cumulative Practice

Definition. Engelmann and Carnine (1982) have suggested that when training coordinate members of a learning set (as when training skill sequences), a method of cumulative programming should be used. Cumulative programming starts by independently training two skills to criterion and then reviewing them together (either both skills practiced in some alternating fashion within the same practice set, or both skills practiced as subskills of a more complex skill). After a criterion is met on the cumulative review, a third skill is trained to criterion. Next, the new skill is added to the two previously trained skills in a cumulative review of all three skills. This procedure is continued until all the skills in the sequence or hierarchy have been trained, with the mastered skills accumulating across the reviews (Becker et al., 1975; Carnine, 1997; Carnine et al., 1994).

Learning outcomes. Cumulative practice has been shown to produce positive learning outcomes such as improved efficiency in acquisition of skills and higher posttest scores. For example, studies of computer-assisted instruction (CAI) have compared programs that include cumulative review with ones that do not include cumulative review. Johnson, Gersten, and Carnine (1987) tested two CAI programs designed to train vocabulary words. One program taught 50 words in two sets of 25 each (called the Large Teaching Set). The other program taught 50 words in sets of up to 7 words each (called the Small Teaching Set). Subjects assigned to the Large Teaching Set were introduced to the first 25 words, their definitions, and sentences using the words. Subjects then practiced the words in multiple-choice exercises. After subjects reached an 84% correct mastery criterion on the exercises for two days, they practiced a sentence completion task for one day. Then, on the following day, the subjects were introduced to the second set of 25 words and began the same sequence of training activities. If the mastery criterion was met before the 20-minute daily training session was completed, subjects were allowed to play an arcade-type game that incorporated the vocabulary words. Training continued until mastery was achieved, or for 11 sessions, at which time subjects who had not mastered the words were not showing any improvement.

Subjects assigned to the Small Teaching Set were first tested on all 50 words, and those already known by a subject were eliminated from the respective subject's training set. Then unlearned words were introduced (no more than three at a time) using the same initial instruction provided to subjects in the Large Teaching Set. After the initial instruction, no more than seven words at a time were assigned to practice sets and trained using multiple-choice exercises (also similar to those used by subjects in the Large Teaching Set). When a word was identified correctly four consecutive times across two lessons, it was considered to be mastered. After 10 words were mastered, a cumulative review test was administered on all 10 words. If any of the words were responded to incorrectly, they were returned to the practice set for retraining. Training continued until mastery was achieved, or for 11 sessions, like the Large Teaching Set procedure.

A 50-item multiple-choice test was administered to all subjects as a pretest, posttest, and maintenance test. In addition, an oral quiz requiring students to formulate definitions was administered after session seven, and a reading comprehension task using frequently missed pretest words was completed after finishing the training (or after session 11 for subjects who did not meet the mastery criteria). Performance on all the measures was similar for both groups; however, more subjects in the Small Teaching Set were able to meet the mastery criteria (10 out of 12 vs. 8 out of 12). Furthermore, it took subjects in the Small Teaching Set significantly fewer sessions to reach mastery (7.6 vs. 9.1). Although this study did not control for many differences between the groups (e.g., number of words taught, mastery criteria, etc.), the study suggests a positive relation between cumulative practice and the rate of skill acquisition, as demonstrated by the efficiency of the cumulative training procedure.

A similar type of study by Gleason, Carnine, and Vala (1991) compared what was called a cumulative versus a rapid introduction method of teaching the identification of seven countries. The cumulative introduction group was required to identify two countries correctly on 6 consecutive trials before a third country was added. Then, after 13 consecutive identifications were made correctly with three countries, the fourth country was added, and so on, up to the seventh country. After the fourth, fifth, sixth, and seventh countries were taught, the students

were required to make 13 to 17 consecutive identifications correctly on a cumulative mix of all the countries trained.

In contrast to the cumulative introduction group, the rapid introduction group was only required to identify each country once before moving on to the next country. After each country had been taught in this manner, students were required to make 13 correct consecutive identifications using a mix of all seven countries. The results of the study showed that the performance of the two groups on a posttest, maintenance test, and application test was comparable. The cumulative group, however, needed significantly fewer responses to criterion during training (a mean of 103 versus 258) and significantly less time to reach final mastery (a mean of 8.7 versus 22.7 minutes).

Another learning outcome that may be facilitated by cumulative practice is posttest achievement. The best examples of high posttest gains associated with a large amount of cumulative practice come from studies of a math curriculum written by an educator named John Saxon. Called an “incremental approach,” Saxon’s texts present only a few new problems in each lesson with the rest being cumulative review problems (Saxon, 1982). In accordance with the cumulative review procedure, a new topic is introduced, practiced, and then added to a second new topic. Then both are practiced in a cumulative fashion until a third new topic is introduced and practiced cumulatively across the next several lessons, and the cycle continues.

Results of studies evaluating Saxon’s basic algebra text have shown very high levels of performance for students using the book. For example, one study involving approximately 1,360 ninth graders revealed that scores on algebra subtests (e.g., equations, word problems, etc.) were two to three times higher for students trained by Saxon’s text versus a traditional Algebra I text (Saxon, 1982). Another study by Klingele and Reed (1984) confirmed these results, showing that students using the Saxon text outperformed control students on both a departmental final exam and a standardized test published by the Mathematics Association of America. According to Saxon, additional benefits reported by schools that have adopted his texts include three times as many students enrolling in a fourth year of high school mathematics, a 50% decrease in the number of students who do not complete high school algebra, and a 20% increase in college board scores (Finn, 1988). Though the specific cause of these improvements cannot be determined, the incorporation of cumulative practice in Saxon’s texts lends support for the value of cumulative practice in producing high levels of posttest achievement.

In summary, then, cumulative practice has been associated with improvements in skill acquisition as well as high overall levels of posttest achievement. For at least some of these results, however, the source of the effect is unclear because the studies have failed to control a number of the confounding variables. Moreover, no study has analyzed the differential contributions of the components of cumulative practice. That is, several elements of practice make up what has been defined as cumulative practice. Any one of these elements may be sufficient to produce positive learning outcomes. It may be that the benefits of cumulative practice are produced merely from the extra practice trials that are inherently part of a cumulative practice procedure. Or, effects may be due to the process of review. Review (in what will be termed a “simple” form) is composed of practice distributed over time and an alternation of new items with previous items, both of which have been shown to be beneficial to learning. Or, it may indeed be that certain learning outcomes will only be produced as a result of a

cumulative review procedure in which all previously trained items are practiced simultaneously after each period of new training. Evidence for each of these cases will now be reviewed.

Components of cumulative practice

Extra practice. An experiment by Wilson, Majskerek, and Simmons (1996) suggested that the number of practice trials is the critical variable that affects skill mastery. In their study, an alternating-treatments design was used to test the acquisition of multiplication facts by four elementary students with learning disabilities. The software program Math Blaster, an example of computer-assisted instruction (CAI), was used as one of the alternating treatments. The other treatment was one-on-one teacher directed instruction (TDI). Both treatments included faded prompts, immediate, corrective feedback, remediation of errors, controlled practice designed to build fluency or automaticity, and timed cumulative practice drills covering all the facts learned in previous lessons as well as the facts learned in the current lesson. Mastery of a math fact was defined as a correct answer written in less than three seconds on two consecutive daily pencil-and-paper probes. The results showed that the two treatments produced a similar number of mastered facts at the beginning of the training. After a number of lessons, however (ranging from 8 to 21 across students), significantly more facts (as measured by visual inspection) were mastered using the TDI. The difference in success rate was reported to be 4% to 34% higher for the TDI instruction. Though many differences in the treatments (mode of presentation, type of rate-building practice activities, etc.) may have produced the difference in performance, one critical variable measured by the researchers was the opportunity to respond (i.e., number of practice trials). Under the TDI, students had two to four times as many opportunities to respond. Thus, as suggested by the researchers, the amount of practice may be the cause of the differential performance. Other differences between the treatments, however, prohibit specific conclusions.

A second study suggesting the critical role of additional practice was conducted by Beck, Perfetti, and McKeown (1982). The purpose of the study was to demonstrate the effects of vocabulary training on application measures including a generalized vocabulary test, a sentence verification task, and a reading comprehension task. The procedure involved training 104 vocabulary words in sets of 8-10 words each. Training on each set incorporated drills such as word association tasks, sentence generation tasks, speed drills in matching the words and definitions, and a multiple-choice test. Approximately two and one-half hours of instruction was spent on each set of words, spread across five days. An experimental group also received extra practice on target words after the training of word sets 3, 5, 7, 9, 11, and 12. The target words were practiced a total of 26 to 40 times throughout the study, while the other words were only practiced a total of 10 to 18 times during the initial 5-day training set. A control group did not receive extra practice on any of the words trained.

The results of the study showed that the experimental group outperformed the control group on a generalized vocabulary posttest that sampled untrained vocabulary words, and a reading comprehension test that used words the experimental group had practiced during the reviews. The experimental group was also faster and more accurate at answering questions using the trained vocabulary words. Furthermore, the experimental group performed significantly better on the words that had received extra practice versus the words that had not received extra practice on a vocabulary posttest involving trained and untrained words. These results were replicated by a follow-up study conducted by McKeown, Beck, Omanson, & Perfetti (1983). The follow-up study also showed that performance on application questions and reading

comprehension questions was superior for words that had received extra practice versus words that had not received extra practice. These results suggest that the extra practice led to improvements in performance on trained skills and application skills. This conclusion may not provide the best description of the source of the effect, though, because the addition of extra practice was implemented in the format of a simple review, or practice that is distributed across time and interspersed with the training of new items. Thus, this study, like the previous one, fails to show that an increase in the amount of practice alone is sufficient to produce benefits associated with cumulative review. Moreover, it suggests turning to an analysis of the effects of simple review procedures.

Simple review. A simple review procedure has two critical features: it incorporates practice across time (distributed practice), and it alternates the training of new items with the practice of old items. Both of these features have been shown to produce positive effects on learning. One of the longest-standing findings of experimental psychology is the benefit of distributed, or spaced practice (see review by Dempster, 1988). For example, an early study by Pyle (1913) showed that third-graders who practiced addition facts for 10 minutes once a day for 10 days outperformed those who drilled for 10 minutes twice a day for 5 days. The group practicing once a day achieved a rate of performance exceeding 200 additions per practice period, while the group practicing twice a day only reached a rate of around 165 additions per practice session. In addition, Bahrick and Phelps (1987) found that subjects trained on Spanish vocabulary with 30-day intervals between sessions were about two-and-a-half times as likely to recall the Spanish words on an 8-year retention test, as compared with subjects trained under massed practice conditions (0 days between sessions). Moreover, they were almost two times as likely to recall the words as subjects trained using a 1-day spacing between sessions. Similarly, Dempster found that subjects who read a passage of text with an interval of 48 hours between readings could recall significantly more information than subjects who read the passage at intervals of 30 seconds and 5 minutes (Experiment 1). Moreover, subjects who read a text twice with an intervening interval of 30 minutes could recall significantly more than subjects who read the passage with an intervening interval of 5 minutes (Experiment 2; both cited in Dempster, 1988). Based on these studies, it seems reasonable to conclude that practice distributed across time, in and of itself, is beneficial to learning outcomes, especially the retention of material.

In addition to distributed practice, simple review procedures incorporate the “interspersal” or alternation of new items with trained items (e.g., Rothkopf & Coke, 1966; Melton, 1970; Underwood, 1970). For example, Kryzanowski and Carnine (1980) studied letter-sound training on the target letters “e” and “i.” One group of subjects received spaced training on “e” (other items placed between presentations of “e”) and massed training on “i,” (no other items placed between presentations of “i”), and the other group received spaced training on “i” and massed training on “e.” All training used sheets containing three rows of five stimuli each. For the spaced training, the specified target letter (e.g., “e”) appeared interspersed with non-target letters (s, m, or c) or the similarly sounding target letter (e.g., “i”). For the massed training, the specified target letter (e.g., “i”) appeared in blocks of three, four, or five in a row. Subjects read the stimuli across the rows with the correct response modeled by the experimenter on the first appearance of each stimulus. The results showed that for both groups, performance on a posttest was superior for the target letter trained with spaced (or interspersed) training.

The interspersing of the trained stimuli also may have led to the effects found by Underwood (1970). He conducted five experiments in which the recall of sentences, common words, and nonsense syllables were tested after differing amounts and distributions of practice trials. Massed practice trials involved consecutive presentations of a stimulus, while distributed practice trials included the presentation of at least one other stimulus between any two trials of a given stimulus. All five studies showed that distributed practice produced better recall than massed practice, and the differences in performance increased across frequency of presentations of the stimuli.

An interspersed procedure that can be more readily identified as a traditional review procedure was tested by Reynolds and Glaser (1964; Experiment 2) in a study that used programmed instruction to teach biology units (e.g., cells, protozoa, tissues, etc.). A unit on mitosis was chosen as the target unit to receive spaced reviews. Some groups received all training on mitosis during one series of frames (massed condition) and some groups received the same amount of training across three series of frames (an initial set of frames plus two spaced reviews provided after other units were taught). Retention tests were administered after 10 days and 3 weeks following the final session of instruction.

The results showed that students receiving the spaced reviews on mitosis scored significantly higher on all measures of performance on mitosis for both retention tests, as compared with students receiving the same amount of practice in one massed presentation. Furthermore, the similar pretest scores and retention test scores of both groups on units not receiving review practice indicated that the groups were equivalent in other regards. Thus, the simple review practice led to improved retention of training.

Statement of the Problem

Based on the above studies, it appears that extra practice and simple review procedures can produce most of the same learning outcomes as cumulative practice, such as improved skill acquisition and retention of training. These learning outcomes, however, are related to performances of skills directly trained. Thus, the question remains whether cumulative practice may be necessary for positive learning outcomes on skills not directly trained. Could cumulative practice facilitate problem solving? If so, would extra practice or simple reviews be sufficient to produce the same results?

The only study discussed thus far that suggested review or extra practice may be sufficient to produce a significant improvement on an untrained skill was the vocabulary study replication by McKeown, Beck, Omanson, & Perfetti (1983). It showed significantly higher performance on application skills of words that had been reviewed (or received extra practice across time) as compared with those that had not been reviewed. This finding for application skills, however, does not necessarily address skills defined as problem solving. Though application of training is similar to problem solving in that it tests behavior in novel situations, it can be distinguished from problem solving by examining what is trained and what is reinforced. When a trained skill or concept is tested in a novel situation and a previously reinforced response occurs, it is not an instance of problem solving because the *same* response is still reinforced. If a child says “chair” in the presence of a novel example of a chair, then the trained response “chair” is still reinforced. Problem solving, on the other hand, involves training more than one response independently such that a synthesis of the responses forms a different response than any

independently trained response. That is, no previously trained response will be reinforced. For the above example to be an instance of problem solving, then, the student would have to be trained separately on the concept of chair and the concept of wooden and then, given the problem of identifying a novel item having the critical features of both concepts, she might combine the responses to say “wooden chair” (the only reinforced response). That would be an example of “verbal problem solving.”

Based on this distinction between application and problem solving, no studies presented assessed the effects of cumulative practice on the novel behavior of primary interest (problem solving), or whether extra practice or review practice would be sufficient to produce the same results. Thus, the question is posed again, what are the effects of cumulative practice on problem-solving behavior? Furthermore, are extra practice and simple review practice procedures sufficient to generate the same results as a cumulative practice procedure? Answers to these questions may reveal what method of teaching component math skills would lead to the successful development of mathematics problem-solving skills.

Current Study

The present study attempted to answer these questions by training five component math skills using extra practice, simple review practice, or cumulative practice procedures. The effect of the training procedures on problem-solving skills was assessed as the main dependent variable. In addition, applications of the trained skills were tested to measure any differential effects of the practice procedures on novel behavior in which trained responses were reinforced in novel situations. The differences in the three training procedures occurred during intermittent “review periods.” For the cumulative practice group, the review periods involved a simultaneous review of all skills trained to that point in the study. For the simple review group, the reviews consisted of practicing one previously trained skill. For the extra practice group, the reviews provided additional practice on the component skill just mastered. During each review period, all subjects completed the same number of practice trials and had the same opportunity for reinforcement.

The five component skills trained included four laws of exponents and one rule about simplifying mathematical expressions (an “order of operations” rule). The problem-solving skills required subjects to combine the basic rules of manipulating exponents and the order of operations rule to solve complex exponent problems involving 3, 4, or all 5 component rules trained (with no direct training on the complex tasks). The application skills required students to use each of the individual rules separately to solve novel examples of each rule.

Chapter 2. Method

Subjects and Setting

Thirty-three students from West Virginia University (WVU) participated in the study. Eleven students were male and 22 were female. None of the students had taken any college level math classes at WVU except Math 23, which is the least rigorous math class available for credit. In addition, the subjects had not received passing credit for either pre-calculus or calculus in high school. For a detailed description of subject characteristics including specific math classes taken in high school and WVU major of study, refer to Appendix A.

To further facilitate the selection of subjects, potential participants completed two days of pretesting on the exponents skills related to the study. The overall percent correct scores on Test 1 and Test 2 were averaged for each subject, and subjects were matched based on their mean scores. Then each matched subject was randomly assigned to one of the three conditions.

The subjects received monetary reimbursement and extra credit in psychology courses for participating in the pretesting and the study. Sessions were held in a room that contained three large desks and one small student desk. During the main body of the study, up to four students were able to participate in experimental sessions concurrently. Typically, however, only one or two subjects reported to the experimental room at a given time.

Materials

Practice worksheets. Five similar versions of all the worksheets were constructed by changing the numbers and letters. See Appendix B for examples of worksheets. On the first version of each worksheet that trained a component rule, one of the five rules was introduced, along with an explanation and several examples of the rule. The five component rules trained were the following: how to multiply variables and coefficients with exponents, how to divide variables and coefficients with exponents, how to raise variables and coefficients with exponents to a power, how to find the roots of variables and coefficients with exponents, and how to simplify multiple-step arithmetic problems using the rule for order of operations.

The independent variable was the type of review worksheet presented to the three groups of subjects on the review day before each test. The cumulative practice group received a cumulative worksheet that presented practice problems covering every rule that had been trained up to that review period. The simple review group received a review worksheet covering one previously mastered rule (e.g., Rule 3 was reviewed subsequent to mastering Rule 4). The extra practice group received another version of the component rule worksheet for the rule they had just mastered (e.g., Rule 3 after Rule 3).

Tests. The primary dependent variable was performance on a test designed by the experimenters. The test was divided into two general sections: application items and problem-solving items. Seven similar versions of the test were constructed by changing the numbers and letters of each problem and the order of presentation of the application problems (see Appendix C for a sample test).

The application items tested the subjects' ability to apply the individual rules to solve novel problems that had more than one variable. For example, during the instruction a student

may have answered the problem $2x^3 \cdot 6x^3 = ?$ while working on Rule 1 practice problems. Then, on a test, the student might have been presented with an application problem such as $3a^6b^2 \cdot 5a^8b^2 = ?$. The critical differences between the practice problems and the application problems, then, were the use of different numbers and letters, and the use of two or three variables instead of only one variable.

All five rules were tested individually in the section containing application items. There were 2 problems for each of the four exponent rules and 4 problems for the order of operations rule, for a total of 12 application items. Each answer for the application items was divided into parts for scoring purposes. The items on the exponent rules were divided into three or four parts (e.g., an answer of $15a^4b^6$ had three parts: 15, a^4 , and b^6). The items on the order of operations only had one part (a number answer). Each part was scored as correct or incorrect.

The second section of the test consisted of the problem-solving items. This section assessed whether novel complex problems that involved combinations of the trained rules could be solved correctly. The items involved combinations of two, three, or all four exponent rules, as well as the order of operations rule. For example, a problem may have been similar to the following: $[(3g^5 \cdot 8g^9) / 4g^8]^2 = ?$. To solve this problem, the student needed to combine four rules: the multiplying variables and coefficients with exponents rule, the dividing variables and coefficients with exponents rule, the raising variables and coefficients with exponents to a power rule, and the order of operations rule. The subjects were never taught how to combine any of the rules; furthermore, they only were shown problems that required combining the rules on the tests.

There were 12 problem-solving items. Four items tested the combination of two exponent rules and the order of operations rule; 4 tested the combination of three exponent rules and the order of operations rule; and 4 tested the combination of all four exponent rules and the order of operations rule. Answers to the problem-solving items were divided into two parts for scoring purposes (e.g., $10b^2$ was separated into 10 and b^2). Each part was scored as correct or incorrect.

Procedure

Subject assignment. During the recruitment process, students were told the subject characteristics that were needed for the study. Students who met the characteristics and were interested in participating in the experiment made an appointment for a pretest session. During the first pretest session, subjects completed one version of the experimenter-designed test (see Appendix C.) All subjects who correctly solved application problems on more than two of the exponent rules did not qualify to continue. Furthermore, subjects who answered more than 4 (out of 24) parts correctly on the problem-solving section of the test did not qualify to continue.

Subjects who did not qualify on the first day's pretest were paid for their correct answers, given an extra-slip validating their participation, and dismissed. All subjects who performed at or below the qualifying criteria were asked to take a second version of the test on a subsequent day. The same qualification criteria also applied to the second pretest. Subjects whose performance on the second day's pretest exceeded the criteria were paid and dismissed.

The overall test scores from both pretests were averaged. All subjects whose scores fell in the qualifying range (10-39% correct) were asked to participate in the study. Subjects who

agreed to participate took the Basic Algebra Exam published by the Mathematics Association of America and signed an Informed Consent Form (see Appendix D). Subjects were then matched on their mean scores on the experimenter-designed pretests and randomly assigned to one of the three groups. A total of 11 subjects were randomly assigned to each group.

Overview of practice sequences. All three experimental groups (cumulative practice, simple review, and extra practice) were exposed to the same training procedure on the five component rules. That is, all subjects were required to complete 25 problems of each rule with 100% accuracy for three sessions before progressing to the next step in the training. Subjects were given unlimited attempts at meeting this mastery criterion during each of the three sessions. When this mastery criterion was met on three separate days, subjects were given a review session (cumulative, simple review, or extra practice). A test was then administered during the subsequent experimental session; and on the following day, the next component rule was introduced.

Table 2.1 provides the general layout of the practice sequences for each group. Subjects were allowed multiple attempts at achieving the daily mastery goal of 25 correct problems for a component rule and typically achieved the final mastery of each rule in three sessions. Therefore, the general cycle of events was three days of practice on an individual rule, one day of review (experimental manipulation), and one day of testing. The only variation in this general cycle occurred at the beginning of the study because two rules were trained before a review could occur. Thus, the cycle was repeated four times (for rules 1 and 2, rule 3, rule 4, and rule 5). During the review sessions, the cumulative practice group received one of the cumulative worksheets (e.g., cumulative 3 after mastering Rule 3). The simple review practice group received worksheets on the rule prior to the most recently trained rule (e.g., Rule 2 after mastering Rule 3), and the extra practice group received additional worksheets on the rule that

Table 2.1 Practice Sequence for Each Condition

Session	Group		
	Cumulative Practice	Simple Review	Extra Practice
1 – 2	Pretests 1, 2	Pretests 1, 2	Pretests 1, 2
3 – 5	Rule 1	Rule 1	Rule 1
6 – 8	Rule 2	Rule 2	Rule 2
9	Cumulative 2	Rule 1	Rule 2
10	Test 3	Test 3	Test 3
11 – 13	Rule 3	Rule 3	Rule 3
14	Cumulative 3	Rule 2	Rule 3
15	Test 4	Test 4	Test 4
16 – 18	Rule 4	Rule 4	Rule 4
19	Cumulative 4	Rule 3	Rule 4
20	Test 5	Test 5	Test 5
21 – 23	Rule 5	Rule 5	Rule 5
24	Cumulative 5	Rule 4	Rule 5
25	Test 6	Test 6	Test 6
26 (Retention)	Test 7	Test 7	Test 7

was just mastered (e.g., Rule 3 after mastering Rule 3). All subjects completed 50 problems on their respective worksheets, which ensured that the number of practice trials was held constant across groups.

Procedure for daily sessions. Individual sessions were conducted daily (Monday – Friday) for approximately 15-30 minutes per subject. In addition, makeup sessions were occasionally conducted on the weekends for subjects who missed days during the week. During the final pretest session, general instructions about the study were provided for all subjects (see Appendix E). During the subsequent session, training began on Rule 1. When Rule 1 and every subsequent rule was introduced, subjects received Version 1 of the practice worksheets containing an explanation of the rule and appropriate examples. If questions were asked, the relevant part of the instructions were pointed out for the subject, or the subject was told that no further explanations could be given at that time (whatever response was appropriate for the respective question).

Subjects then completed 25 problems requiring the use of the rule. The problems were graded by the experimenter or an assistant, and feedback was given to the subjects. Feedback included whether each response was correct or incorrect and an explanation of how to achieve the correct answer for all problems answered incorrectly. For example, if the subject completed a problem as follows: $3^3b^2 \cdot 3^6b^8 = 9^9b^{10}$, the grader gave the following feedback: “For this problem, you want to use the rule that says if the coefficients are the same, then keep the number the same and add the exponents. So, you want to keep the coefficient 3 as a 3 for the answer.” The grader also made the appropriate corrective markings on the paper (i.e., a “3” above or below the coefficient “9” in the incorrect answer). Cycles of practice and feedback occurred in this manner until 25 problems were completed correctly. If a subject needed to leave before meeting the mastery criterion for that day, then the session was continued on the subsequent day.

On a few occasions, exceptions to the method of presenting general corrective feedback had to be made. Each exception has been documented in Appendix F. All the exceptions involved such things as a prerequisite rule that a subject had to be told explicitly (e.g., $3^2 = 3 \cdot 3$) or a clarification of how two rules were not contradictory. No exceptions provided information about how to complete the problem-solving items.

On review days, subjects were presented with their respective review worksheets. They completed 50 problems and then received feedback in the same manner as for the regular practice days. No additional problems were completed, however, regardless of performance.

On the day following a review, subjects took a test. After the sixth test was completed (after two pretests and four tests during training), the main body of the study was finished. There were no specific time limits imposed on any of the tests, but the experimenter recorded the time taken to complete the application items and the time taken to complete the problem-solving items.

After a retention interval, a seventh test was administered to all subjects who were available. There were two different retention intervals used in the study: 4-6 weeks (approximately 1-1½ months) and 9-12 weeks (approximately 2-2½ months). These retention intervals were developed around the general availability of subjects after the date they finished

the main body of the study. (They were not based on specific theoretical grounds other than the assumption that the retention interval should be at least one month for the purpose of seeing differential effects on performance). Seven subjects from each group were available for retention testing. Four of the subjects from each group were tested after the shorter retention interval and three from each group after the longer interval.

Reinforcement procedures. Subjects earned money for correct answers on the tests (including the pretests and retention test) and were not penalized for incorrect answers (refer to the instructions provided for the subjects in Appendix E, which include test earnings). During the session following a test, subjects were presented with a record of their earnings on the test. This was the only feedback provided concerning their performance on the tests.

Subjects also earned \$1.50 on every practice day that they met the mastery goal of 25 problems correct. On review days, they earned \$2.00 if they completed all 50 problems correctly on the first (and only) attempt. Subjects were paid their earnings halfway through the study and at the completion of the study. They also were awarded extra credit in psychology courses at the conclusion of the study.

Interobserver agreement. Interobserver agreement scores were calculated for both the tests and practice worksheets. Approximately 20% of the tests administered during the main body of the study were regraded, and an agreement score was calculated by dividing the agreements by the total responses (agreements plus disagreements) and multiplying by 100%. The average agreement score for all the tests was 99%. Approximately 20% of the retention tests were also regraded. The agreement score for the retention tests was 100%.

For each subject, approximately 20% (4 out of 19) of the practice and review days were randomly selected. On each of these days, every set of problems completed by the subject as an attempt to meet the daily goal was regraded. If the second observer disagreed with any marking of the first observer (e.g., one part of a problem should have been marked as incorrect but was not), the whole attempt of 25 or 50 problems was counted as a disagreement. Using this conservative measure, the total agreements for all the subjects was divided by the total number of attempts (number of agreements plus disagreements), and the quotient was multiplied by 100%. This formula resulted in an interobserver score of 89% for the practice and review session problem sets.

Chapter 3. Results

Pretests

There were five pretest measures of mathematics performance administered to all participants. Table 3.1 presents a summary of descriptive and inferential statistics for all five measures. The first four performance measures were assessed using Test 1 and Test 2 of the experimenter-designed tests. These tests covered skills related to the current study (i.e., manipulations with exponents and order of operations) and were administered on two consecutive daily sessions. The four measures evaluated were percent correct on application items, rate correct on application items, percent correct on problem-solving items, and rate correct on problem-solving items. Although the cumulative group's mean scores appeared to be slightly higher on these measures, the numerical differences were small and the variability large, such that no differences were statistically significant for Test 1 or Test 2. The fifth pretest measure, the Basic Algebra Exam published by the Mathematics Association of America, was a test of general algebra skills. Once again, the slightly higher score of the cumulative group was not statistically significant. In summary, then, there were no statistical differences among groups on any of the pretest measures.

Table 3.1 Summary of Pretest Results

Measure	<u>Cumulative Practice</u>		<u>Simple Review</u>		<u>Extra Practice</u>		F(2,30)
	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	<u>M</u>	<u>SE</u>	
Application % correct							
Test 1	38.09	4.92	36.72	5.00	36.81	4.07	.027
Test 2	44.00	3.34	43.18	3.54	43.00	4.86	.018
Application rate ^a							
Test 1	1.72	0.26	1.58	0.22	1.61	0.27	.079
Test 2	3.12	0.59	2.20	0.37	2.46	0.45	.974
Problem solving % correct							
Test 1	4.09	1.27	3.73	1.64	4.00	0.76	.022
Test 2	3.36	1.59	2.27	1.55	1.45	0.81	.493
Problem solving rate ^a							
Test 1	0.13	0.04	0.09	0.04	0.12	0.03	.341
Test 2	0.08	0.05	0.05	0.04	0.04	0.02	.191
General algebra skills % correct							
Basic Algebra Exam	18.9	3.54	15.3	3.40	15.6	2.38	.404

Note. None of the F-values are significant at the $p < .05$ level. M = mean; SE = standard error.

^a Measured in number of correct responses per minute.

Training

A three-way repeated measures ANOVA with a between subjects factor of Group and within subjects factors of Rule and Day was conducted to analyze the number of attempts taken to meet the daily mastery criterion during the practice sessions. The factor of Group consisted of the cumulative group, the simple review group, and the extra practice group. The factor of Rule reflected performance on each of the five component rules. The factor of Day represented performance on each of the three training days during which subjects were required to master each rule before moving ahead in the training.

There were no significant interactions or main effects of group. Significant main effects were found for rule, $F(4, 120) = 13.06$, $p < .001$, and day, however, $F(2, 60) = 7.47$, $p < .01$. Analyses of main comparisons revealed that the mean number of attempts taken to meet the mastery criterion on Rule 1 ($M = 2.10$, $SE = .13$) was not different from that of Rule 2, $F(1, 30) = 4.11$, $p > .05$; however, it was greater than the mean number of attempts taken on Rules 3, 4, and 5, $F(1, 30) = 48.12$, $p < .001$, $F(1, 30) = 43.66$, $p < .001$, and $F(1, 30) = 6.30$, $p < .05$, respectively. Moreover, the subjects took significantly more attempts to master each rule on practice Day 1 than on Days 2 and 3, $F(1, 30) = 4.98$, $p < .05$ and $F(1, 30) = 16.86$, $p < .001$, respectively. There was no significant difference, however, in the mean number of attempts taken on Days 2 and 3, $F(1, 30) = 2.30$, $p > .05$.

Testing

The main dependent variable of the study was performance on the experimenter-designed tests presented in Appendix C. The tests were broken into two parts: application and problem-solving items. Accuracy (percent correct) and rate (number of correct answers/minute) were measured for both parts. Figure 3.1 shows the mean percent correct and standard errors of application items for all groups across the six tests. A two-way repeated measures ANOVA with a between subjects factor of Group and a within subjects factor of Test revealed no interaction of Group x Test, $F(10, 150) = 1.47$, $p = .156$; a main effect of test, $F(5, 150) = 119.42$, $p < .001$; and no main effect of group, $F(2, 30) = 3.08$, $p = .06$. The data of primary interest, however, were scores on the final tests. On Test 5, the cumulative group scored an average of 95.00% correct ($SE = 1.54\%$), whereas the simple review group's mean score was 84.55% correct ($SE = 4.07\%$), and the extra practice's mean score was 85.27% correct ($SE = 3.07\%$). On Test 6, the cumulative group's average score was 97.00% correct ($SE = 0.90\%$); the simple review group's mean score was 85.36% correct ($SE = 3.84\%$); and the extra practice group's mean score was 85.45% correct ($SE = 3.59\%$). These differences of approximately 10% (one letter grade) on Test 5, and 12% (more than one letter grade) on Test 6 were statistically significant based on one-way ANOVAs, $F(2, 30) = 3.61$, $p < .05$, and $F(2, 30) = 4.72$, $p < .05$, respectively. In addition, comparisons revealed that the cumulative group outperformed the simple review group, $t(12.80) = 2.40$, $p < .05$, and $t(11.11) = 2.95$, $p < .05$, respectively for Tests 5 and 6, and the extra practice group, $t(14.73) = 2.84$, $p < .05$, and $t(11.26) = 3.12$, $p < .05$, respectively for Tests 5 and 6. (The degrees of freedom for the t-tests reflect the appropriate values for groups with unequal variances. This was necessary because of significant differences as measured by Levene's test for homogeneity of variance on Tests 5 and 6, Levine statistic[2, 30] = 11.48, $p < .001$ and Levine statistic[2, 30] = 5.65, $p < .01$, respectively.)

Figure 3.1. Mean percent correct on application items for all three groups across pretests and tests. Tests were administered to all groups at the same time, though data points are spaced apart for legibility. Error bars represent plus and minus one standard error.

Figure 3.2 displays the mean rate of correct responses and standard errors of application items for all groups across the six tests. The results of a two-way repeated measures ANOVA with a between subjects factor of Group and a within subjects factor of Test showed no significant interaction of Group x Test, $F(10, 150) = 1.54$, $p = .13$; a main effect of test, $F(5, 150) = 71.21$, $p < .001$; and no main effect of group, $F(2, 30) = 1.53$, $p = .23$. Once again, however, the trend was in the expected direction, with the cumulative group responding at a mean rate of 11.51 corrects per minute ($SE = 1.72$) by Test 6. The simple review group's mean response rate, however, was only 8.20 correct responses per minute ($SE = 0.85$), and the extra practice group's mean response rate was only 8.89 correct responses per minute ($SE = 1.15$) on Test 6. The variability within the groups, though, was too large to result in any statistical differences between groups.

Figure 3.2. Mean number of correct application items per minute for all three groups across pretests and tests. Tests were administered to all groups at the same time, though data points are spaced apart for legibility. Error bars represent plus and minus one standard error.

A two-way repeated measures ANOVA with a between subjects factor of Group and a within subjects factor of Test was also conducted to analyze differences in accuracy on the problem-solving items. Figure 3.3 shows the mean percent correct and standard errors on problem-solving items across tests for all groups. Because the Group x Test interaction was significant, $F(10, 150) = 3.09$, $p < .01$, simple effects of group at each test were examined. There were no significant differences in performance on Test 1, $F(2, 30) = 0.22$, $p = .98$, Test 2, $F(2, 30) = 0.49$, $p = .62$, Test 3, $F(2, 30) = 2.42$, $p = .11$, Test 4, $F(2, 30) = 2.66$, $p = .09$, or Test 5, $F(2, 30) = 2.80$, $p = .08$. There was a significant difference on Test 6, however, $F(2, 30) = 5.78$, $p < .01$. The cumulative group ($M = 82.64\%$, $SE = 5.18\%$) outperformed the simple review group ($M = 56.27\%$, $SE = 7.61\%$), $t(17.62) = 2.86$, $p < .05$, and the extra practice group ($M = 45.27\%$, $SE = 10.33\%$), $t(14.73) = 3.23$, $p < .01$. There was no significant difference between the means of the simple review group and the extra practice group, however, $t(18.39) = 0.86$, $p = .40$. (The degrees of freedom for the t-tests reflect unequal variances, Levine statistic $[2, 30] = 4.47$, $p < .05$.)

Figure 3.3. Mean percent correct on problem-solving items for all three groups across pretests and tests. Tests were administered to all groups at the same time, though data points are spaced apart for legibility. Error bars represent plus and minus one standard error.

Further analysis was conducted on the problem-solving performance at Test 6 to see if the differences in accuracy among groups were solely due to differences in acquisition of the component rules (as measured by application performance). To conduct this analysis, a proportional problem-solving score was calculated for each subject by dividing the percent correct on the problem-solving items on Test 6 by the percent correct on the application items on Test 6. Using this relative analysis, a subject in the cumulative group who scored 100% on the problem-solving section and 100% on the application section would receive the same score (1.0) as a subject in the extra practice group who scored 65% on both sections.

Figure 3.4 shows the mean proportional problem-solving scores and standard errors for all groups on Test 6, as well for all previous tests. A one-way ANOVA conducted on the proportional problem-solving scores at Test 6 revealed a significant main effect of group, $F(2, 30) = 3.95$, $p < .05$. The cumulative group ($M = .85$, $SE = .05$) scored significantly higher than the simple review group ($M = .64$, $SE = .07$) and the extra practice group ($M = .53$, $SE = .11$), $t(18.43) = 2.34$, $p < .05$, and $t(14.37) = 2.62$, $p < .05$, respectively. Once again, however, there was no significant difference between the means of the simple review and extra practice groups, $t(17.14) = 0.86$, $p = .40$. (There were unequal variances, Levine statistic $[2, 30] = 7.36$, $p < .01$).

Therefore, the same relation among the means was found on this relative measure of problem-solving performance.

Figure 3.4. Mean proportional problem-solving scores for all three groups across pretests and tests. Tests were administered to all groups at the same time, though data points are spaced apart for legibility. Error bars represent plus and minus one standard error.

The final dependent measure, the rate of correct responses on the problem-solving items, also was analyzed using a two-way repeated measures ANOVA with a between subjects factor of Group and a within subjects factor of Test. Figure 3.5 shows the mean number of correct responses per minute and standard errors across tests for all groups. A significant interaction of Group x Test was found, $F(10, 150) = 3.26$, $p < .01$. There were no differences among groups on Test 1, $F(2, 30) = 0.34$, $p = .71$, Test 2, $F(2, 30) = 0.19$, $p = .83$, or Test 3, $F(2, 30) = 2.03$, $p = .15$. There was a difference in performance on Test 4, however, $F(2, 30) = 4.20$, $p < .05$. The cumulative group ($M = 2.11$, $SE = .42$) performed significantly faster than the simple review group ($M = .84$, $SE = .19$), $t(14.09) = 2.78$, $p < .05$. (The group variances were unequal, Levine statistic[2, 30] = 5.36, $p < .05$.) On Test 5, the cumulative group ($M = 2.95$, $SE = .40$) again outperformed the simple review group ($M = 1.52$, $SE = .26$), $t(30) = 2.66$, $p < .05$. On Test 6, the cumulative group ($M = 2.76$, $SE = .30$) responded faster than both the simple review group ($M = 1.63$, $SE = .17$) and the extra practice group ($M = 1.54$, $SE = .38$), $t(30) = 2.70$, $p < .05$, and $t(30) = 2.93$, $p < .01$, respectively. There was no significant difference between the mean rates of the simple review group and the extra practice group, however, $t(30) = 0.23$, $p = .82$.

Figure 3.5. Mean number of correct problem-solving items per minute for all three groups across pretests and tests. Tests were administered to all groups at the same time, though data points are spaced apart for legibility. Error bars represent plus and minus one standard error.

Retention Test

Retention data for the 21 subjects who completed Test 7 are displayed in Figures 3.6 and 3.7. The data were analyzed using one-way ANOVAs with a between subjects variable of group. The same dependent variables (accuracy and rate on application and problem solving) were analyzed for the retention test as for the pretests and tests. Although the general patterns of results for both the accuracy and rate data were similar to those of Test 6, the analyses revealed no significant differences among the groups on any of the measures. This is probably due in part to the reduction in sample size and the increases in variability within groups.

Figure 3.6. Mean percent correct on application and problem-solving items for all three groups on the retention test. Error bars represent plus and minus one standard error.

Figure 3.7. Mean number of correct application and problem-solving items per minute for all groups on the retention test. Error bars represent plus and minus one standard error.

Chapter 4. Discussion

Effects of Cumulative Practice

The current study examined the effects of cumulative practice on problem-solving behavior, as compared with the effects of simple review practice and extra practice. Three groups of students with poor mathematics skills learned five algebra rules under similar training conditions. The only difference in the training procedures was the type of review session presented before each test (cumulative, simple review, or extra practice). The pretests showed that all three groups of subjects started out at the same level of performance on the target algebra skills. Furthermore, there were no group differences on the number of attempts needed to master each rule as subjects progressed through the training. On the review days, all groups answered their respective review items with a mean accuracy of at least 97% correct. This level of performance was quite high, especially for the cumulative group who was required to answer review items on all rules learned up to that point for each review session.

The cumulative group also performed well on the tests of application and problem-solving skills. Results from the final test administered to all groups indicated that the cumulative group outperformed the other two groups on accuracy of application and problem-solving skills, as well as rate correct on problem-solving skills. The cumulative group's average performance on the final test of problem-solving skills was 82.6% correct. This means that students who answered an average of around 1 problem-solving item correctly (out of 24) prior to training were able to answer an average of approximately 20 problem-solving items correctly after training, even though the training did not include *any* instruction on problem solving.

Although this finding is significant per se, it is likely that the effect of cumulative practice may be even greater than these data show. That is, the average level of performance of students in the cumulative group was lowered by two scores that deserve closer inspection. The lowest two scores on Test 6 were 42% and 63% correct. There is a relatively large gap between these scores and the other scores (79%, 83%, 83%, 83%, 88%, 92%, 96%, 100%, and 100%). The score of 42% was obtained by a student who probably belonged to a different subject population than the other 32 subjects in the experiment. She was the only subject who had taken no math classes in high school or at WVU, completing only an Algebra I course at a learning center. The reason the subject was included in the study was because she met the subject qualification criteria of passing no more than trigonometry in high school and no more than Math 23 at WVU. Furthermore, her mean score on the two experimenter-designed pretests was greater than 10% correct (the minimum qualifying score). Based on her post-training test results as compared to all other subjects, however, it is clear that the criteria should have included a minimum requirement for math classes taken and not just a maximum requirement.

The second lowest score was likely a result of several uncontrollable factors that can affect data collected from human subjects. That is, the subject began missing sessions frequently during the end of the study, which was likely due to the normal increase in the amount of work and number of tests at the end of a college semester. This particular subject, however, fell behind in the daily sessions to the point that she had to complete the final several sessions during exam week. Moreover, she was the only subject who had to complete two sessions in less than a 12-hour time span because it was one of the last days of exam week, and she was leaving for the

summer as soon as she completed the final test. All these factors likely contributed to her low performance on Test 6 (the final test). In summary, then, it may be useful to exclude these two low scores for the sake of examining the potential effects of cumulative review. If these outliers are discarded, the mean performance of the cumulative group jumps to almost 90% correct on Test 6.

On Test 7, retention data was collected from only one of the two subjects who had scored poorly. When excluding this score from the average score for the retention test, the mean problem-solving performance increases to 85.5% correct. Considering the many weeks of no practice, this level of performance is quite high. When statistical analyses were conducted on all the data collected for the retention test, including the low scorer's datum, neither the problem-solving results nor the other retention results proved to be statistically significant. This outcome is likely due to a number of factors, including increases in variability of the scores, slight increases in the other groups' performance coupled with slight decreases in the cumulative group's performance, and a decrease in power derived from the sample size.

Despite the lack of significant differences on the retention test, though, the overall effects of cumulative practice are quite impressive, especially when disregarding the two outlying scores on Test 6. Before cumulative practice is declared to be the best of the three methods for training problem-solving skills, however, potential outliers that affected the mean scores of the other two groups also should be addressed.

Although no unique factors were observed that may have affected the low-end performers, the high-end performers tend to raise more interesting issues. That is, two subjects from each of the other groups performed at a very high level (88% and 100% for two subjects in the simple review group, and 96% and 96% for two subjects in the extra practice group). Due to unknown variables in these students' learning histories, they were able to achieve the same performance as the top cumulative subjects simply through mastering the skills and receiving either extra practice or simple review practice. Their cohorts, however, were not. In fact, without including these top performers, the range of scores for the simple review group was 17%, 21%, 46%, 46%, 54%, 67%, 67%, and 67%, which yields an average of 47.9% correct. The range of scores for the extra practice group was 4%, 13%, 17%, 17%, 17%, 42%, 54%, 67%, and 75%, which yields an average of 34.0% correct.

Whether the "high outliers" of the extra practice and simple review groups and the "low outliers" of the cumulative group are included in the discussion or not, however, the overall performance of the cumulative group was much less variable than the performance of the other two groups. Excluding the outliers, the scores for the cumulative group (as presented earlier) covered a range of 21% (79%-100%). Scores of the simple review group and extra practice group (excluding their outliers) covered ranges of 50% (17%-67%) and 71% (4%-75%), respectively.

It seems, then, that a simple set of principles can be derived from the individual problem-solving data. That is, the best method of training problem solving, as tested in this study, involves cumulative practice. Cumulative practice will override most other historical and current contextual learning variables for this target population. With weaker, less effective methods,

however, the outcomes are much more variable due to individual learning histories. For a few top performers, the less than ideal training methods will not help nor hinder the outcome. For the rest of the students, though, the results will fall along a wide range of performance outcomes (mostly much lower than those of cumulative subjects) based on each subject's training history. This is similar to what seems to occur in the school systems. That is, a few students do well no matter what type of instruction is used. Other students can do well if some other variables come into play, but typically perform very poorly without external help. Almost all students do well if the best instructional method is used.

Overall, the results from the study suggest that incorporating cumulative practice into training procedures will lead to high levels of performance on novel, untrained skills. More specifically, what are typically thought of as "higher-level" mathematical skills such as applying individually trained rules in a novel situation and synthesizing rules into novel combinations can be facilitated through a cumulative practice training procedure. Neither providing extra practice on each component rule nor incorporating individual reviews of previously trained rules proved adequate to produce comparable results, particularly on problem-solving skills.

Furthermore, based on the subject population used in the study, the results suggest that even students with low levels of math skills can successfully perform "higher-level" skills through adequate training on component skills. Component skills can be mastered while simultaneously programming for novelty through the use of cumulative review procedures. The end result is that students perform well on the component skills *and* are prepared to extend the individual skills to novel situations (application) and to synthesize the skills into novel solutions derived from combinations of the skills (problem solving). These are the results that mathematics educators have sought for years.

Reasons for the Effectiveness of Cumulative Practice

Multiple simple reviews. Because the results of the current study provide strong support for incorporating cumulative practice into teaching procedures, the next logical question that educational researchers might ask is whether cumulative practice can be analyzed further to determine what basic principles are at work. There seem to be at least two possible explanations of the current findings that could be explored with follow-up research. First, the cumulative review procedure might have been superior to the other two procedures simply because every trained skill was reviewed at every review period. That is, the benefits of simple review practice were multiplied because every possible skill received interspersed practice spaced over time. This explanation would simply be due to the benefits of a simple review procedure applied to every rule in a cumulative fashion.

To test this explanation, a study similar to the present one could be run again with two groups. One group would be the current cumulative group, who would receive one worksheet covering a mix of all problems learned at each review period; and one group would be a group that received multiple simple review worksheets covering every rule learned (each worksheet covering only one rule) at each review period. The only difference between the groups, then, would be that the cumulative group practiced the problems in a mixed format, and the other group reviewed each rule individually on multiple simple review worksheets. If the number of practice trials at each review period were controlled and both groups performed the same, then

the key to the success would appear to be in the simple review of each rule learned so far at every review period. If the cumulative group outperformed the other group, however, then the effects would be due to factors arising from the format of the review sheets. This brings out the second proposed explanation—that cumulative practice facilitates discrimination training.

Discrimination training. The other possible explanation of the effects of cumulative practice on problem solving is derived from the description of problem solving presented by Skinner (1966) and Becker et al. (1975). As explained previously, they have described the process of problem solving as occurring when one previously trained response creates an S^D for another response, which, in turn, changes the environment such that another S^D is formed. Ultimately, then, the S^D for the final response is presented and the solution is complete. Based on this description, the amount of stimulus control of specific S^D s over individually trained responses surfaces as a critical factor influencing the emergence of problem solving. According to the second possible explanation, then, discrimination training may be needed to increase the stimulus control over each trained response.

Discrimination training may result from cumulative practice through the differential reinforcement of correct responses for each type of problem in the presence of all other types of problems, as occurs on every cumulative review worksheet. This differential reinforcement would increase the stimulus control over responses to each problem type. Tighter stimulus control, in turn, would facilitate problem solving because each response in the series of problem-solving responses would be more likely to occur and produce the correct S^D for the next response, which, ultimately, would form the solution to the problem.

Previous applied research on cumulative practice and directly trained behavior lends support to this second explanation of the results. Fink and Brice-Gray (1979) showed that a method of cumulative programming involving more simultaneous discriminations produced superior performance as compared with a method of cumulative programming involving fewer simultaneous discriminations. They trained subjects with developmental disabilities to identify five two-syllable words. The training for both groups of subjects began with an introduction of two flashcards that contained one word each. The experimenter stated the first word while pointing to the appropriate card and asked the subject to touch the word. If the subject did so, a reinforcer (determined via a teacher conference) was presented immediately. If the subject failed to touch the correct word, the experimenter again pointed to the correct response and repeated the request to touch the correct word.

For the subjects in a group called the successive pairs group, word 2 was named after word 1 had been identified correctly on eight consecutive trials. Then eight consecutive correct responses were required on word 2. Next, eight consecutive discriminated responses were required on words 1 and 2. Word 3 was then introduced alone, and eight consecutive correct responses on it were required before discriminated responses were required for each pair of words trained (word 1 vs. word 2, word 1 vs. word 3, and word 2 vs. word 3). Words 4 and 5 also were introduced in this manner, with each word identified correctly on eight consecutive trials by itself before being discriminated from each previously trained word in a pairwise fashion.

The second group, the cumulative introduction group, was given more simultaneous discriminations on each word because each additional word was trained in the presence of all previously trained words after being trained in isolation. Thus, word 3 had to be discriminated from words 1 and 2 simultaneously; word 4 had to be discriminated from words 1, 2, and 3 simultaneously; and word 5 had to be discriminated from words 1, 2, 3, and 4 simultaneously. The same mastery criterion was used in this procedure as was used in the successive pairs procedure. The results of the study indicated that subjects trained under the successive pairs procedure needed significantly more trials to achieve all mastery criteria. Moreover, they performed significantly poorer on the posttest than did subjects trained using more simultaneous discriminations.

Although Fink and Brice-Gray (1979) only tested directly trained skills, additional support for the discrimination training explanation can be found from applied research on novel behavior, including productive and receptive generative language. Studies have shown that discrimination training resulting from a type of cumulative review procedure is necessary for the emergence of novel behavior. Frisch and Schumaker (1974) trained receptive language using a multiple-baseline-across-behaviors design. The three behavioral baselines used involved responding correctly to three categories of prepositional phrase requests: "Put the _____ next to the _____."; "Put the _____ under the _____."; and "Put the _____ on top of the _____." One specific request for one category of the prepositional phrases (e.g., "Put the dog next to the boat") was trained until a performance criterion was met. Then, a probe session occurred that included 16 instances of previously trained requests interspersed with 5 probe requests from each of the three categories of prepositional phrases. Responding correctly to trained requests, but not probed requests, was reinforced during the probe sessions. If all probe items from the prepositional phrase currently being trained were not answered correctly, then the subsequent training session involved training another request from the same prepositional phrase category until a performance criterion was met on that request. Then another probe session occurred. Training continued in this manner on one type of prepositional phrase until all probe requests were responded to correctly. Then either training on a new type of prepositional phrase or discrimination training (a cumulative review procedure of all prepositional phrases trained thus far) was presented, and probe sessions were conducted again across all types of prepositional phrases.

The general pattern of results across the three subjects used in two experiments revealed that the introduction of cumulative practice (called discrimination training by the experimenters) increased the performance of novel instruction following for all behaviors trained to that point in the research. Moreover, the best training procedure resulted from training each type of prepositional phrase request across a number of various requests and then in conjunction with all previously trained types of requests (cumulative review) before training a new type of request. This was similar to the procedure used in the current experiment.

Clark and Sherman (1975) found comparable results using a multiple-baseline-across-behaviors design that involved the training of productive language. When training on a new concept was introduced, novel responding to all previously trained concepts decreased (generally to 0% correct). Subsequently, when concurrent training on all previously trained skills (called discrimination training) was introduced, high levels of novel performance emerged across all

concepts trained thus far. These studies suggest that cumulative practice is a form of discrimination training that facilitates novel behavior such as concept formation. In the same manner, problem-solving behavior is likely to be facilitated through the incorporation of cumulative practice.

Chapter 5. Conclusions

The current study showed that cumulative practice leads to high levels of performance of novel mathematics behaviors, including application and problem-solving skills. The novel behaviors resulted from direct training on component algebra skills alone. Furthermore, the same level of problem-solving performance was not achieved by subjects exposed to either simple review or extra practice procedures. These findings support the viewpoints of behavioral educators: first, that mastery of component skills facilitates performance on higher-level skills; and second, that complex, novel behavior is fundamentally related to its component parts and can be explained using basic behavioral processes. The results of the current study also extend the research of behavioral educators by removing the confounded variables of simple review and extra practice as found in previous studies, and by showing the effects of cumulative practice on problem solving. In addition, the argument has been proposed that increased stimulus control resulting from discrimination training provided by cumulative practice might be one of the underlying basic processes involved in problem-solving behavior.

From an applied perspective, the findings suggest that it may be beneficial for the National Council of Teachers of Mathematics to promote the type of instructional approach used in the current study as a strategy for training problem solving in the new millennium. That is, the curriculum aimed at teaching “higher-level” skills should be broken down into its component parts. Then, instruction should focus on mastery of the component skills with sufficient cumulative practice incorporated into the training. This approach may produce significant increases in achievement scores, not only on the component skills directly trained (such as manipulations with fractions or exponents), but also on novel tasks that require a synthesis of students’ classroom training.

This type of improvement in mathematics instruction will be essential if the United States plans to achieve its goal of being a world leader in mathematics. Though four decades of consistently poor math performance suggest that the goal may be unreachable any time in the near future, educators who pursue data-driven reforms should persist in trying to implement change. It is never too late to start teaching effectively.

References

- Alessi, G. (1987). Generative strategies and teaching for generalization. The Analysis of Verbal Behavior, 5, 15-27.
- Andronis, P. T., Layng, T. V. J., & Goldiamond, I. (1997). Contingency adduction of “symbolic aggression” by pigeons. The Analysis of Verbal Behavior, 14, 5-17.
- Anrig, G. R., & LaPointe, A. E. (1989). What we know about what students don’t know. Educational Leadership, 47, 4-9.
- Bahrack, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 344-349.
- Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74, 506-521.
- Beck, R., & Clement, R. (1991). The Great Falls Precision Teaching Project: An historical examination. Journal of Precision Teaching, 8, 8-12.
- Becker, W. C., Engelmann, S., & Thomas, D. R. (1975). Teaching 2: Cognitive learning and instruction. Chicago: Science Research Associates.
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. The Behavior Analyst, 19, 163-197.
- Brown, M. (1996). FIMS and SIMS: The first two IEA international mathematics surveys. Assessment in Education: Principles, Policy, & Practice, 3, 193-213.
- Campbell, J. R., Voelkl, K. E., & Donahue, P. L. (1997). NAEP 1996 trends in academic progress. Achievement of U.S. students in science, 1969 to 1996; mathematics, 1973 to 1996; reading, 1971 to 1996; writing, 1984 to 1996. Report in brief. (Report No. TM027731). Washington: National Library of Education. (ERIC Document Reproduction Service No. ED 411 327)
- Carnine, D. (1977). Phonics versus look-say: Transfer to new words. Reading Teacher, 30, 636-640.
- Carnine, D. (1997). Instructional design in mathematics for students with learning disabilities. Journal of Learning Disabilities, 30, 130-141.
- Carnine, D., & Becker, W. C. (1982). Theory of instruction: Generalisation issues. Educational Psychology, 2, 249-262.
- Carnine, D., Jones, E., & Dixon, R. (1994). Mathematics: Educational tools for diverse learners. School Psychology Review, 23, 406-427.
- Carpenter, T. P., Kepner, H., Corbitt, M. K., Lindquest, M. M., & Reys, R. E. (1980a). Results and implications of the second NAEP mathematics assessment: Elementary school. Arithmetic Teacher, 27, 10-12, 44-47.
- Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Lindquest, M. M., & Reys, R. (1980b). Results of the second NAEP mathematics assessment. Secondary school. The Mathematics Teacher, 73, 329-338.
- Catania, A. C. (1980). Autoclitic processes and the structure of behavior. Behaviorism, 51, 175-186.
- Catania, A. C., & Cerutti, D. T. (1986). Some nonverbal properties of verbal behavior. In T. Thompson & M. D. Zeiler (Eds.), Analysis and integration of behavioral units (pp. 185-211). Hillsdale, NJ: Earlbaum.

- Clark, H. B., & Sherman, J. A. (1975). Teaching generative use of sentence answers to three forms of questions. Journal of Applied Behavior Analysis, 8, 321-330.
- Darch, C., Carnine, D., & Gersten, R. (1984). Explicit instruction in mathematics problem solving. Journal of Educational Research, 77, 351-359.
- Dempster, F. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. American Psychologist, 43, 627-634.
- Dixon, R. (1980). Morphographic Spelling. Chicago: Science Research Associates.
- Ellenwood, D. W., & Chase, P. N. (1997). A comparison of two behavioral training components: Rate building and component composite analysis. Poster presented at the Association for Behavior Analysis, Chicago.
- Ellenwood, D. W., Chase, P. N., & Madden, G. (1999). A behavior analytic model of verbal behavior: Combining synonyms, syntax, and parts of speech. Manuscript in preparation, West Virginia University.
- Englemann, S., Becker, W. C., Carnine, D., & Gersten, R. (1988). The direct instruction Follow Through model: Design and outcomes. Education and Treatment of Children, 11, 303-317.
- Englemann, S., & Carnine, D. (1982). Theory of instruction: Principles and applications. New York: Irvington Press.
- Epstein, R. (1985). The spontaneous interconnection of three repertoires. The Psychological Record, 35, 131-143.
- Epstein, R. (1987). The spontaneous interconnection of four repertoires of behavior in a pigeon (*Columba livia*). Journal of Comparative Psychology, 101, 197-201.
- Fink, W. T. & Brice-Gray, K. J. (1979). The effects of two teaching strategies on the acquisition and recall of an academic task by moderately and severely retarded preschool children. Mental Retardation, 17, 8-12.
- Finn, C. E. (1988, November). Math angles & Saxon. National Review, 40, 30-31.
- Frisch, S. A., & Schumaker, J. B. (1974). Training generalized receptive prepositions in retarded children. Journal of Applied Behavior Analysis, 7, 611-621.
- Geary, D. C. (1996, October). International differences in mathematical achievement: Their nature, causes, and consequences. Current Directions in Psychological Science, 5, 133-137.
- Gleason, M., Carnine, D., & Valia, N. (1991). Cumulative versus rapid introduction of new information. Exceptional Children, 57, 353-358.
- Goldstein, H. (1983). Training generative repertoires within agent-action-object miniature linguistic systems with children. Journal of Speech and Hearing Research, 26, 76-89.
- Goldstein, H., Angelo, D., & Wetherby, B. (1987). Effects of training method and word order on adults' acquisition of miniature linguistic systems. The Psychological Record, 37, 89-107.
- Goldstein, H., & Moussetis, L. (1989). Generalized language learning by children with severe mental retardation: Effects of peers' expressive modeling. Journal of Applied Behavior Analysis, 22, 245-259.
- Johnson, G., Gersten, R., & Carnine, D. (1987). Effects of instructional design variables on vocabulary acquisition of LD students: A study of computer-assisted instruction. Journal of Learning Disabilities, 20, 206-213.
- Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. American Psychologist, 47, 1475-1490.

- Johnson, K. R., & Layng, T. V. J. (1994). The Morningside model of generative instruction. In R. Gardner, D. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman, & T. A. Grossi (Eds.), Behavior analysis in education: Focus on measurably superior instruction (pp. 173-197). Pacific Grove, CA: Brooks/Cole.
- Keller, F. S. (1968). "Good-bye, teacher ..." Journal of Applied Behavior Analysis, 1, 78-79.
- Klingeale, W. E., & Reed, B. W. (1984, June). An examination of an incremental approach to mathematics. Phi Delta Kappan, 65, 712-713.
- Kouba, V. L., Brown, C. A., Carpenter, T. P., Lindquist, M. M., Silver, E. A., & Swafford, J. O. (1988). Results of the fourth NAEP assessment of mathematics: Number, operations, and word problems. Arithmetic Teacher, 35, 14-19.
- Kryzanowski, J., & Carnine, D. (1980). The effects of massed versus spaced formats in teaching sound-symbol correspondences to young children. Journal of Reading Behavior, 12, 225-229.
- Kulik, J. A., Jaska, P., & Kulik, C. C. (1978). Research on component features of Keller's Personalized System of Instruction. Journal of Personalized Instruction, 3, 129-141.
- Leitzel, J. R. (1989). Critical considerations for the future of algebra instruction, or, a reaction to: "Algebra: What should we teach and how should we teach it?" In S. Wagner & C. Keiran (Eds.), Research issues in the learning of algebra, 4, (pp. 25-32). Reston, VA: National Council of Teachers of Mathematics & Erlbaum.
- Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. Journal of Applied Behavior Analysis, 25, 51-57.
- Markle, S. M. (1983). Designs for instructional designers. Champaign, IL: Stipes.
- McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. Journal of Reading Behavior, 15, 3-18.
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. Journal of Verbal Learning and Verbal Behavior, 9, 596-606.
- Moore, J. (1986). Direct instruction: A model of instructional design. Educational Psychology, 6, 201-229.
- Moore, L. J., & Carnine, D. (1989). Evaluating curriculum design in the context of active teaching. Remedial and Special Education, 10, 28-36.
- National Council of Teachers of Mathematics. (1980). An agenda for action: Recommendations for school mathematics for the 1980's. Reston, VA: Author.
- National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.
- National Council of Teachers of Mathematics (1999). Principles and Standards for School Mathematics Electronic Version 1.0: Discussion Draft [On-line]. Available: <http://standards-e.nctm.org>
- Pyle, W. H. (1913). Economical learning. Journal of Educational Psychology, 3, 148-158.
- Rech, J. F., Juhler, S. M., & Johnson, H. L. (1995). The effectiveness of videotapes in the individualized instruction of intermediate algebra and college algebra. International Journal of Math Education and Scientific Technology, 26, 463-472.
- Reese, C., Miller, K. E., Mazzeo, J., & Dossey, J. A. (1997). NAEP 1996 mathematics report card for the nation and the states. Findings from the National Assessment of Educational Progress (Publication No. SE059966). Washington, DC: U.S. Government Printing Office.

- Resnick, L. B., Wang, M. C., & Kaplan, J. O. (1993). Task analysis in curriculum design: A hierarchically sequenced introductory mathematics curriculum. Journal of Applied Behavior Analysis, 6, 679-710.
- Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. Journal of Educational Psychology, 55, 297-308.
- Rothkopf, E. Z., & Coke, E. U. (1966). Variations in phrasing, repetition intervals, and the recall of sentence material. Journal of Verbal Learning and Verbal Behavior, 5, 86-91.
- Saxon, J. (1982, March). Incremental development: A breakthrough in mathematics. Phi Delta Kappan, 63, 482-484.
- Skinner, B. F. (1954). The science of learning and the art of teaching. Harvard Educational Review, 24, 86-97.
- Skinner, B. F. (1966). An operant analysis of problem solving. In B. Kleinmuntz (Ed.), Problem solving: Research, method, and theory (pp. 159-62, 166-67). Englewood Cliffs, NJ: Prentice-Hall.
- Skinner, B. F. (1968). The technology of teaching. New York: Appleton-Century-Crofts.
- Streifel, S., Bryan, K. S., & Aikins, D. A. (1974). Transfer of stimulus control from motor to verbal stimuli. Journal of Applied Behavior Analysis, 7, 123-135.
- Streifel, S., & Wetherby, B. (1973). Instruction-following behavior of a retarded child and its controlling stimuli. Journal of Applied Behavior Analysis, 6, 663-670.
- Streifel, S., Wetherby, B., & Karlan, G. R. (1976). Establishing generalized verb-noun instruction-following skills in retarded children. Journal of Experimental Child Psychology, 22, 247-260.
- Tiemann, P. W., & Markle, S. M. (1990). Analyzing instructional content: A guide to instruction and evaluation (4th ed.). Champaign, IL: Stipes.
- Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning. Journal of Verbal Learning and Verbal Behavior, 9, 573-580.
- Vanneman, A. (1998, September). Long-term trends in student mathematics performance. NAEP Facts, 3, 1-5.
- Wetherby, B., & Streifel, S. (1978). Application of miniature linguistic system or matrix-training procedures. In R. L. Schiefelbusch (Ed.), Language Intervention Strategies (pp. 319-356). Baltimore, MD: University Park.
- Wilson, R., Majsterek, D., & Simmons, D. (1996). The effects of computer-assisted versus teacher-directed instruction on the multiplication performance of elementary students with learning disabilities. Journal of Learning Disabilities, 29, 382-390.
- Wingert, P. (1996, December 2). The sum of mediocrity. Newsweek, 128, 96.
- Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. Journal of Special Education, 32, 207-225.

Appendix A. Subject Characteristics.

Group	Major	Math Classes				WVU/other
		9 th	10 th	11 th	12 th	
Cumulative	Undecided	Alg. I	Geom.	Alg. II	None	Workshop (non-credit)
	Political Science	Alg. I	Geom.	Alg. II	Alg. III	None
	Undecided	Alg. I	Geom.	Alg. II	None	Workshop
	Elementary Education	Geom.	Pre-Alg.	Alg. I	None	None
	Social Work	Alg. I	Geom.	Alg. II	College Review	23
	Interior Design	Pre-Alg.	Alg. I	Geom.	None	23
	Art	Geom.	Alg. II	None	None	None
	Child Development	None	None	None	None	Alg. I at a Learning Center; No math at WVU
	Advertising	Alg. I	Geom.	Alg. II	None	None
	Elementary Education	Alg. I	Geom.	Alg. II	Trig.	None
	Advertising	Alg. I	Alg. II	Geom.	Trig.	23
	Simple Review					
Social Work		None (Alg. in 8 th)	None	General Math	None	23

	Physical Education	Alg. II	Geom.	Trig.	None	None
	Business	Pre-Alg.	Alg. I	None	Stat. Elective	None
	Liberal Arts	None	Integration of Alg./Geom./Trig. Part I	Part II	Part III	None
	Social Work	Alg. I	Geom.	Trig.	None	23
	Communications	Alg. I	Alg. I	Alg. II	Geom.	23
	Social Work	Alg. I	Geom.	Alg. II	None	23
	Liberal Arts	Alg. I	Alg. II	Alg. for Artists	None	None
	Nursing	Pre-Alg.	Alg. I	Geom.	Alg. II	Workshop
	Psychology	Alg. I	Alg. II	Trig.	None	None
	General Studies	Alg. I	Geom.	Alg. II	None	None
Extra Practice						
	Elementary Education	Alg. I	Geom./Intermediate Alg.	Alg. II	None	Workshop at WV Northern Community College
	Speech Pathology	Pre-Alg.	Geom.	Alg. I	Alg. II	None
	Journalism	Alg. I	Geom.	Alg. II	None	23
	Undecided	Alg. I	None	None	Alg. II	None

Journalism	Integration Of Alg./ Geom./ Trig. Part I	Part II	Part III	Attempted Pre-Cal.; Received “D”	None
Landscape Architecture	Alg. I	Alg. II	Geom.	Geom.	None
Fashion Marketing	Alg. I	Alg. II	Trig./ Geom.	None	None
Journalism	Alg I Part I	Alg. I Part II	Alg. II	None	None
Sports Marketing	Pre-Alg.	Alg. I	Alg. II	None	None
Nursing	Alg. I	Geom.	Alg. II	Trig.	None
Marketing	Alg. I	Geom.	Alg. II	Trig.	Workshop

Appendix B. Sample Practice Worksheets

RULE 1

Version 1

To multiply variables with exponents, **ADD** the exponents.

Examples: $a^2 \cdot a^3 = (a \cdot a) \cdot (a \cdot a \cdot a) = a^5$ or $a^{2+3} = a^5$

$$x^5 \cdot x^9 = x^{14}$$

$$(g^3)(g) = g^4 \quad (\text{Remember: } g = g^1.)$$

If the variables have coefficients (numbers in front), **MULTIPLY** the coefficients as normal.

Examples: $3x^5 \cdot 7x^9 = 21x^{14}$
 $(t^3)(4t^5) = 4t^8$ (Remember $t^3 = 1t^3$.)

If the coefficients are the same number, you can **ADD** their exponents just like you do for the variables.

Examples: $(2^4g^3)(2^5g) = 2^9g^4$
 $5^3y^3 \cdot 5y^4 = 5^4y^7$ (Remember $5 = 5^1$.)

$$2d^5 \cdot 5d^4 =$$

$$8h^5 \cdot 5h^9 =$$

$$(9r)(6r^3) =$$

$$3^7w^4 \cdot 3^8w^4 =$$

$$6g^4 \cdot g =$$

$$(2^5s^7)(2^9s^9) =$$

$$(8^3n^1)(8^7n^8) =$$

$$5c^4 \cdot 6c^2 =$$

$$7h^6 \cdot 8h^0 =$$

$$4^5y^3 \cdot 4^6y^6 =$$

$$(7k)(4k^3) =$$

$$9x^9 \cdot 8x^6 =$$

$$6b^8 \cdot 6b^3 =$$

$$(3t^9)(2t^4) =$$

$$5^6f^3 \cdot 5^8f^8 =$$

$$5j^8 \cdot 7j^5 =$$

$$9e^9 \cdot 5e^8 =$$

$$8v^7 \cdot 3v^4 =$$

$$(4r)(7r^7) =$$

$$4u^4 \cdot 5u^8 =$$

$$2^0g^2 \cdot 2^9g =$$

$$(1y^9)(y^9) =$$

$$(6n^4)(8n^8) =$$

$$8d^9 \cdot 4d^3 =$$

$$8^7t \cdot 8t^4 =$$

$$9w^9 \cdot 7w^3 =$$

$$(2r)(5r^2) =$$

$$3^2x^8 \cdot 3^8x^5 =$$

Cumulative 4 (Rules 1, 2, 3, and 4)**Version 1**

$$6c^8 \cdot 2c^4 =$$

$$\frac{16x^{12}}{2x^5} =$$

$$(10^3n^9)^5 =$$

$$\sqrt[5]{5^{25}r^{40}} =$$

$$\sqrt[6]{6^{36}d^{42}} =$$

$$\frac{32g^{14}}{8g^6} =$$

$$(7^8t^5)^4 =$$

$$7h^8 \cdot 3h^7 =$$

$$(7d^4)^6 =$$

$$8^3t^9 \cdot 8^5t^8 =$$

$$\sqrt{6^6y^2} =$$

$$\frac{42x^{15}}{6x^9} =$$

$$(5^5r^9)^8 =$$

$$\sqrt[3]{9^{21}h^{27}} =$$

$$(k^4)(9k^8) =$$

$$\frac{7^{17}x^9}{7^8x^3} =$$

$$5^4s^7 \cdot 5^6s^5 =$$

$$\frac{49x^{12}}{7x^5} =$$

$$(10^4w^6)^8 =$$

$$\sqrt[4]{10^{32}b^{36}} =$$

$$\frac{40x^{15}}{5x^7} =$$

$$\sqrt[7]{7^{49}r^{28}} =$$

$$(7^8h^1)^3 =$$

$$6j^5 \cdot 4j^2 =$$

Appendix C. Sample Test

Test 1

Part A: [Application Items]

1. $4^3 s^2 t^6 \cdot 4^9 s^8 t^7 =$ _____

2. $\sqrt{5^6 a^{18} b^{10}} =$ _____

3. $(20^6 x^3 y^7)^8 =$ _____

4. $\frac{7^{14} g^{13} h^{15}}{7^7 g^6 h^9} =$ _____

5. $(15 - 9)^4 =$ _____

6. $\sqrt[6]{6^{12} f^{42} g^{48} h^6} =$ _____

7. $\sqrt{\frac{18}{6}} + (2^2 - 3) =$ _____

8. $(14k^9 m^4 n^8)^3 =$ _____

9. $\frac{4+8}{12-8} =$ _____

10. $9t^5 u^7 v^8 \cdot 7t^4 u^9 v^5 =$ _____

11. $\frac{36q^{18} p^{12} r^{13}}{9q^9 p^7 r^7} =$ _____

12. $\sqrt{(3 \cdot 3) + 7} =$ _____

Part B: [Problem-Solving Items]

13. $\left(\frac{21b^{11}}{3b^4}\right)^2 =$ _____

14. $(8x^2 \cdot 7x^7)^9 =$ _____

15. $\frac{\sqrt{8^2 k^6}}{4k} =$ _____

16. $\sqrt[7]{5^9 d^5 \cdot 5^5 d^9} =$ _____

17. $\left(\frac{4a^3 \cdot 5a^5}{2a^4}\right)^5 =$ _____

18. $\left(\frac{\sqrt[3]{4^6 y^{12}}}{4y^2}\right)^7 =$ _____

19. $\sqrt[4]{(9h^2 \cdot 6h^1)^8} = \underline{\hspace{2cm}}$

20. $\frac{\sqrt[5]{18^5 f^{25}}}{2f^2 \cdot 3f^2} = \underline{\hspace{2cm}}$

21. $\left(\frac{\sqrt[4]{36^4 g^{12}}}{3g^1 \cdot 3g} \right)^9 = \underline{\hspace{2cm}}$

22. $\sqrt[3]{\frac{(3h^1 \cdot 3h^2)^6}{h^9}} = \underline{\hspace{2cm}}$

23. $\frac{\sqrt{2^4 t^{10}} \cdot (2t^4)^2}{2t^5} = \underline{\hspace{2cm}}$

24. $\frac{(3^4 m^6)^3}{m^9} \cdot \sqrt[5]{1^5 m^{45}} = \underline{\hspace{2cm}}$

Appendix D. Informed Consent Form

TITLE OF RESEARCH: Acquisition of Mathematics Problem-solving Skills

INTRODUCTION: I, _____, have been invited to participate in this research study which has been explained to me by either Dr. Philip N. Chase, Kristin Hazlett, or one of their assistants. This study is part of Kristin Hazlett's dissertation research, which is partially funded by the Department of Psychology Alumni Fund and/or the Office of Academic Affairs.

PURPOSE OF RESEARCH: I understand that the purpose of this study is to measure mathematics performance on both basic and complex math skills. I understand that data from my participation may be used to partially fulfill the requirements for a doctoral dissertation. Approximately 25 subjects will participate in this study.

PROCEDURE: I understand that this project may require up to 9 hours of my time. I willingly consent to allow the principal investigator to have access to my academic records, including my West Virginia University Math Placement Test score(s) and grade(s) in the Pre-College Algebra Workshop and/or College Algebra class.

I understand that I will participate in one session per day (Monday through Friday). I understand that for each session I will receive some training on rules of exponents for 10-15 minutes or will be given a paper-and-pencil test on those skills, along with some more complex problems involving rules of exponents for 10-15 minutes. On test days, I will earn money (up to \$3.00 per test) based on the number of questions I answer correctly. In addition, on training days I will earn between \$1.50 and \$2.00 per day for attendance and completion of training procedures.

I understand that if I miss two or more sessions, or if I do not call in advance of missing a session, I may be dropped from the experiment. I understand that if I am dropped from the experiment or choose not to complete it, I will still receive any money I have earned up to that point. I understand that I will be paid my earnings after completing half of the experiment and then again at the end of the experiment. I understand that the experimenters will keep track of my earnings. I understand that I will learn the math skills at my own pace; thus the length of the study may vary from approximately 4 weeks to 6 weeks.

BENEFITS: I understand that this study will not necessarily be of direct benefit to me, but I am likely to improve my academic skills. The knowledge gained from the experiment may be of benefit to others. I understand that I will earn money for my training and test performance. My test earnings will grow throughout the study as I learn the math skills. Thus, I may earn up to \$9.50 per week for my attendance and performance by the end of the experiment. There will be no monetary costs to me as a subject. I understand that I have the right to ask questions before signing this consent form, and I may contact a principal investigator if I have any additional questions about the research or my rights.

RISKS AND DISCOMFORTS: I understand that there are no known or anticipated risks involved in serving as a subject in this study. I am aware that there may be unforeseeable risks in participating in any experiment.

CONTACT PERSONS: For more information about this research, I can contact Dr. Philip N. Chase at 293-2001 ext. 626 or Kristin Hazlett at 293-2001 ext. 855. For more information about my rights as a subject, I may contact the Executive Secretary of the Institutional Review Board at (304) 293-7073.

CONFIDENTIALITY: I understand that any information about me obtained because of my participation will be kept as confidential as legally possible. I understand that my research records, just like hospital records, may be subpoenaed by court order or may be inspected by federal regulatory authorities. In any publications that result from this research, neither my name nor any information from which I might be identified will be published.

VOLUNTARY PARTICIPATION: Participation in this study is voluntary. I realize that I am free to withdraw my consent to participate in this study at any time. Refusal to participate or withdrawal will involve no penalty or loss of benefits and will not affect any of my grades or class standing. I have been given the opportunity to ask questions about the research, and I have received answers concerning areas I did not understand.

Upon signing this form, I will receive a copy.

I willingly consent to participate in this study.

Signature of Subject or Subject's Representative

Date

Investigator or Investigator's Representative

Date

Appendix E. Instructions

Thanks for agreeing to participate in this mathematics study!

You should already have done the following:

- Taken the pretests
- Scheduled a 15-minute time slot to meet every day (Monday – Friday)
- Read, signed, and received a copy of the Informed Consent Form

From now on, during each session of the study, you will do the following:

- **practice** a math skill
- OR
- Complete a **Test**

Instructions for the Practice Sessions:

During this study you will learn five math rules. Every time a new rule is introduced, you will read a worksheet that explains the rule and presents some examples that have been worked out for you. After you have read the explanation of the rule and have studied the examples, you will begin practicing the rule. Your goal is to complete 25 problems correctly. After you have attempted the first 25, the experimenter will grade them and tell you which ones you missed. If you didn't miss any, the session will be over and you will earn **\$1.50** for your performance. If you miss one or more problems, you will attempt an additional set of 25 problems. Then the experimenter will grade them and tell you which ones you missed. You will be allowed to continue attempting sets of 25 problems until you get them all correct. Thus, you will always earn \$1.50 if you show up for the session and continue working on the practice problems until you have met your goal.

You must meet your goal of 25 correct problems on three different days before you are allowed to move ahead to the next step in the training.

On special practice days, the experimenter will tell you that your goal for that day is to complete 50 problems on the worksheet she gives you. You will only be allowed one attempt to complete all 50 problems correctly, but if you do, you will earn **\$2.00** that day.

Instructions for the Test Sessions:

Approximately every five days, you will complete a test of the math skills you are learning. There will be 24 problems on each test, divided into 2 parts, with 12 problems on each part. Your goal is to answer as many of the problems correctly as you can. You will be learning how to do the problems throughout the study, so you may not be able to answer all of the questions at first. Do the best you can. If your answer contains a number raised to an exponent (like 3^4 or 13^2), you can leave your answer in that form. You do not have to figure out what the number raised to the exponent equals. If you know the answer, though, (like $2^2 = 4$), you can write what it equals. Either answer will be counted as correct.

There will be no specific time limit on the test, but the experimenter will record how long it takes you to complete each part of the test. Therefore, when you have answered all the

questions you can on the first part of the test, tell the experimenter that you are done so she can record your time. Do the same after you have completed all you can on the second part of the test.

You will earn money for whatever correct responses you make. You will not lose money for incorrect responses. Therefore, it is to your advantage to attempt every problem. You also will receive partial credit for your answers, so you can earn money for an answer even if only part of it is correct. You will earn **3¢** for every part of an answer that you complete correctly on Part A of the test. You will earn **8.5¢** for every part of an answer that you complete correctly on Part B of the test. Thus, if you get the entire test correct, you will earn **\$3.00**.

Instructions for Payment and Extra Credit Procedures:

Each day you will receive a slip of paper telling you how much you earned during the previous day's 15-minute session. You will be paid your earnings halfway through the study and then again at the end of the study. You will also receive a slip verifying your participation at the end of the study. You can turn this slip into your psychology instructor (if you are in a psychology class) and you might receive extra credit, depending on the policy of your instructor.

Instructions for the Follow-up Test next Fall:

The experimenter will need your local phone number for next fall (or home phone number if you do not know your local phone number yet). She will call you next fall to schedule one 15-minute session during which you will complete a follow-up test. You will have the same opportunity to earn money on the follow-up test as you do during the main part of the study. You can also earn extra credit for your participation if you are taking a psychology course next fall.

Other Important Instructions:

It is very important that you do not discuss anything about the skills you are learning in the study with anyone else. Moreover, please do not refer to mathematics textbooks or other sources of math instruction because that would interfere with the results of the study. Remember, the study is based on the fact that you are only receiving math instruction during your daily participation in the study! Thanks for your cooperation.

These instructions will be available to you at any time for your reference. In addition, you will be provided with a copy to keep with you. On your copy, please fill in a record of your weekly schedule and who to contact if you have an emergency and cannot make the session. It is very important that you contact the appropriate person. Remember, if you miss two or more sessions without informing the experimenter, you can be dropped from the study. You will only receive whatever money you have earned up to that point, and you will miss out on the extra credit slip that is given out at the end of the study. If you come every day, the study should only take about 4 or 5 weeks, so please make every effort to attend daily!

A RECORD OF YOUR WEEKLY SCHEDULE:

Monday:	Experimenter to contact:
Tuesday:	Experimenter to contact:
Wednesday:	Experimenter to contact:
Thursday:	Experimenter to contact:
Friday:	Experimenter to contact:

Appendix F. Feedback Exceptions

1. One subject from each group was told that $3^2 = 3 \times 3$ (as opposed to 3×2).
2. One subject in the cumulative group was told that anything times 1 equals itself. This was in reference to feedback on the problem $8^7t \bullet t^7 = ?$. The problem was rewritten for the subject as $8^7t \bullet 1t^7 = ?$, and then the extra feedback stated above was provided.
3. Two subjects in the extra practice group and one subject in the simple review group were provided with clarification on how Rule 4 (which said to divide the exponents when dealing with roots) fit in with Rule 5 (which was a simplification rule that sometimes used examples involving square roots as part of the expressions to be simplified). The clarification was made using one of the examples listed in the instructional frame for Rule 5. It was explained that $\sqrt{16} = \sqrt[2]{4^2} = 4^{2 \div 2} = 4^1 = 4$; and thus, the rules were not contradictory.

Curriculum Vita

Kristin Hobbs Hazlett

645 Springdale Avenue, Morgantown, WV 26505 (304) 598-8672 khazlett@wvu.edu

Present Position

Graduate Student and Research Assistant
Department of Psychology
West Virginia University

Education

B.A., Psychology and Mathematics, 1996
G.P.A. = 4.0
West Virginia University
Morgantown, WV

M.A., Psychology, 1998
G.P.A. = 4.0
Program Area: Behavior Analysis (Learning)
Emphases: Mathematics Education and Instructional Design
West Virginia University
Morgantown, WV

Ph.D., Psychology, requirements completed 1/00
Degree officially awarded (graduation ceremony) 5/00
Expected G.P.A. = 4.0
Program Area: Behavior Analysis (Learning)
Emphases: Mathematics Education and Instructional Design
West Virginia University
Morgantown, WV

Awards and Scholarships

Swiger Fellowship for Graduate Education (1996 – 1999)
A three-year academic fellowship given to top graduate students at the University.
Foundation Scholarship for Undergraduate Education (1992 - 1996)
A four-year academic scholarship given to top 4-5 undergraduate students at the University.
Order of Augusta (1996)
An academic and well-rounded citizenship award given to top 8 students in the graduating class.
Outstanding Senior in the Psychology Department (1996)
Psychology Department Quin Curtis Award (1996)
Phi Beta Kappa (1996)
University Honors Scholar (1996)

Instructional Design and Teaching Experience

- 7/96 - 8/96 Educational Internship and Teaching Practicum at Morningside Academy, Seattle, WA
Trained to use Morningside's model of Generative Instruction; taught algebra to high school students using generative instruction; attended classes on curriculum-based evaluation and behavioral interventions for family therapy.
- 8/97 - 12/97 Instructor for WVU's Pre-College Algebra Workshop
Taught two sections of developmental college algebra (each met four times weekly); used generative instruction for one section; designed lectures, practice worksheets, and tests.
- 1/98 – 8/99 Mathematics Education Research Assistant
Assisted in writing a National Science Foundation Grant proposal for computerizing the developmental mathematics curriculum at West Virginia University. Served as a research assistant for developing this individualized algebra review course for entering WVU freshmen during Fall 1999. Duties included assisting in design of course curriculum, procedures, and web site; also helped oversee implementation process.
- 9/98 – 10/98 Instructional Design Consultant for Performance Management Course produced by the Virtual Education Corporation and Aubrey Daniels and Associates
Designed generalization and fluency exercises for 8 lessons.
- 11/98 – 5/99 Instructional Design Consultant for Professional Engineering Exam Review Course produced by the Virtual Education Corporation
Designed module on retaining walls, including conceptual background, representative problems, scripts for instructor, graphics, and instructional models.
- 8/99 – 12/99 Intern for the Continuous Learning Group (a behavioral consulting firm).
Duties involved activities related to instructional design projects such as writing and editing scripts for computerized, multimedia presentations; developing graphics to accompany scripts; researching and writing about online training; and developing guidelines for effective multimedia and online training.

Research Experience

- "The Effects of Generative Instruction on Mathematics Performance"
Undergraduate Research Project conducted with Philip N. Chase
Tested the effects of building high rates of basic math skills on performance of higher-level skills.

“The Effects of Fluency-Building of Arithmetic Skills on the Performance of Complex Algebra Skills”

First-year Graduate Research Project conducted with Philip N. Chase

Tested the effects of building high rates of math facts and fractions on the performance of complex problems involving exponents; developed sequential fractions curriculum to accompany whole number arithmetic curriculum.

“Teaching Algebra through Cumulative Mastery”

Thesis conducted with Philip N. Chase

Compared traditional method of algebra instruction with innovative method involving mastery of basic skills and other fundamental algebra skills through increased amounts of practice with rate criteria; developed rate-building curriculum on factoring, solving systems of equations, and manipulations with exponents.

“The Effects of Cumulative Practice on Mathematics Problem-Solving Behavior”

Dissertation conducted with Philip N. Chase (currently)

Taught five algebraic rules concerning manipulations of exponents and order of operations. Tested for the emergence of novel behavior both in the form of simple applications of the rules and complex combinations of the rules. Compared the accuracy and rate of novel, or problem-solving, behavior for students in each of three types of instructional conditions: additional practice trials per rule (beyond mastery), alternation of new items with review items, and cumulative review of all rules in a mixed format.

“Resistance to Change Revealed by Residual History Effects of VR and DRL Performance”

Animal Research Project conducted with Sergio Cirino, Hiroto Okouchi, Adam Doughty, and Kennon A. Lattal

Tested the residual history effects of high- versus low-rates of responding (using a multiple VR DRL reinforcement schedule) on subsequent performance on the same reinforcement schedule (multiple VI VI schedule). Prefeeding and extinction were used as disrupters during the VI VI schedule. Compared the resistance to change due to the differential reinforcement histories.

Presentations

Hazlett, K. H., Chase, P. N., & Munson, K. J. (1998, May). Potential Costs and Benefits of Computerizing Precision Teaching in Mathematics Education. Paper presented at the Association for Behavior Analysis, Orlando, FL.

Hazlett, K. H., & Chase, P. N. (1999, May). The Effect of Cumulative Practice on Mathematics Problem-Solving Behavior. Poster presentation at the Association for Behavior Analysis, Chicago, IL.

Hazlett, K. H. (1999, May). Extensions of Skinner’s Naturalistic Theory of Values. Paper presentation at the Association for Behavior Analysis, Chicago, IL.

Hazlett, K. H., & Chase, P.N. (2000, May). The Effect of Cumulative Practice on Mathematics Problem-Solving Behavior. (Extension of Research). Paper to be presented at the Association for Behavior Analysis, Washington, DC.

Professional Affiliations

Association for Behavior Analysis

References

Philip N. Chase, Ph.D.
Professor of Psychology
West Virginia University
(304) 293-2001 x 626
pchase@wvu.edu

Michael Perone, Ph.D.
Professor and Chairperson of Psychology
West Virginia University
(304) 293-2001 x 604
mperone@wvu.edu

Kennon A. Lattal, Ph.D.
Centennial Professor of Arts and Sciences and
Coordinator of the Behavior Analysis Training Program
West Virginia University
(304) 293-2001 x 608
klattal@wvu.edu

Joseph Wilder, Ph.D.
Associate Professor of Mathematics
West Virginia University
(304) 293-2011 x 2344
wilder@math.wvu.edu