JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2002, **35,** 105–123 NUMBER 2 (SUMMER 2002)

# *THE EFFECTS OF CUMULATIVE PRACTICE ON MATHEMATICS PROBLEM SOLVING*

KRISTIN H. MAYFIELD AND PHILIP N. CHASE

WEST VIRGINIA UNIVERSITY

This study compared three different methods of teaching five basic algebra rules to college students. All methods used the same procedures to teach the rules and included four 50-question review sessions interspersed among the training of the individual rules. The differences among methods involved the kinds of practice provided during the four review sessions. Participants who received cumulative practice answered 50 questions covering a mix of the rules learned prior to each review session. Participants who received a simple review answered 50 questions on one previously trained rule. Participants who received extra practice answered 50 extra questions on the rule they had just learned. Tests administered after each review included new questions for applying each rule (application items) and problems that required novel combinations of the rules (problem-solving items). On the final test, the cumulative group outscored the other groups on application and problem-solving items. In addition, the cumulative group solved the problem-solving items significantly faster than the other groups. These results suggest that cumulative practice of component skills is an effective method of training problem solving.

DESCRIPTORS: cumulative practice, problem solving, application, quantitative skills, algebra

---

Over 35 years of international comparisons of mathematics achievement have indicated problems with the performance of students from the United States. According to the latest international study, the average score of U.S. students was below the international average, and the top 10% of U.S. students performed at the level of the *average* student in Singapore, the world leader (Wingert, 1996). In addition, recent tests administered by the U.S. National Assessment of Educational Progress revealed that 70% of fourth graders could not do arithmetic with whole numbers and solve problems that required one manipulation. Moreover, 79% of eighth graders and 40% of 12th graders could not compute with decimals, fractions, and percentages, could not recognize geometric figures, and could not solve simple equations; and 93% of 12th graders failed to perform basic algebra manipulations and solve problems that required multiple manipulations (Campbell, Voelkl, & Donahue, 1997).

These statistics reveal students' deficits in the fundamental skills of mathematics as well as mathematical reasoning and problem solving. Indeed, poor problem-solving skills have been targeted by the National Council of Teachers of Mathematics (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1980; Carpenter, Kepner, Corbitt, Lindquist, & Reys, 1980; Kouba et al., 1988; National Council of Teachers of Mathematics, 1989, 2000). Thus, it seems appropriate that current behavior-analytic research in mathematics education should address problem-solving skills as well as basic mathematics skills (e.g., Wood, Frank, & Wacker, 1998).

---

This research was partially funded by dissertation grants from the Office of Academic Affairs and the Psychology Alumni Fund of West Virginia University. We thank Karrie Wagoner, Brittany Wintry, Jenny Mankins, Karen Dotson, Shane Taylor, Gilbert Miller, and Bridget Hayes for assisting in data collection. We also thank Barry Edelstein, Michael Perone, David Schaal, and Joseph Wilder for their helpful comments on an earlier draft of this paper.

Correspondence concerning this article should be addressed to Philip N. Chase, Department of Psychology, West Virginia University, P.O. Box 6040, Morgantown, West Virginia 26506-6040 (e-mail: pchase@wvu.edu).

According to Becker, Engelmann, and Thomas (1975), problem-solving tasks are those that demand a novel (untrained) synthesis of responses in the presence of a novel stimulus. Following this definition, problem solving can be distinguished from other kinds of novel behavior by requiring the presence of both novel stimuli and novel responses. Therefore, application of skills, stimulus generalization, extension (Skinner, 1957), and concept formation are different kinds of novel stimulus–response relations than problem solving because they are defined as engaging in a previously trained response in the presence of a new stimulus (Alessi, 1987).

Problem solving may take the form of a novel chain of previously learned responses (e.g., formulating a novel step-by-step geometric proof based on previously learned theorems; see Epstein, 1985, 1987) or a novel 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 needed to buy a bus ticket without direct instruction; see Johnson & Layng, 1992, 1994). Explicit instruction on the novel responses is never provided, and thus instruction from previous situations must be sufficient to produce a solution in the novel context.

This type of problem solving has been observed in various types of experimental and educational research, although it is often identified by different names, such as generalization (Streifel, Bryan, & Aikins, 1974; Streifel & Wetherby, 1973; Streifel, Wetherby, & Karlan, 1976), recombinative generalization (Goldstein, 1983; Goldstein, Angelo, & Wetherby, 1987; Goldstein & Mousetis, 1989; Melchiori, de Souza, & de Rose, 2000; Mueller, Olmi, & Saunders, 2000), contingency adduction (Andronis, Layng, & Goldiamond, 1997; Binder, 1996; Johnson & Layng, 1992, 1994), and productivity (Catania, 1980; Catania & Cerutti, 1986). Regardless of the terminology, however, this body of research suggests that one critical variable for producing problem solving is mastery of the component behaviors or skills. Although such mastery appears to be necessary, it is often not sufficient because in some cases participants have been successful and in others they have not (see discussion of negative results in Wetherby & Streifel, 1978). Thus, it appears critical to identify further the instructional variables that are important for successful problem solving. One such variable may be the procedure of having learners cumulatively practice previously mastered skills.

Cumulative practice is a procedure that has been used successfully in a variety of behavioral literatures, including direct instruction (Becker et al., 1975), miniature linguistic systems (Wetherby & Streifel, 1978), and stimulus equivalence (Sidman & Tailby, 1982). Cumulative practice begins by independently training two skills to criterion and then practicing them together, usually by mixing tasks for both skills within the same practice set. After a criterion is met on the cumulative practice set, a third skill is trained to criterion. Next, the new skill is added to the two previously trained skills in a cumulative set involving all three skills. This procedure is continued until all the skills in a sequence or hierarchy have been trained, with the mastered skills accumulating across the cumulative sets (Becker et al., 1975; Carnine, 1997; Carnine, Jones, & Dixon, 1994).

Cumulative practice has been shown to produce greater efficiency in skill acquisition (e.g., Gleason, Carnine, & Valia, 1991) and posttest achievement (Klingele & Reed, 1984; Saxon, 1982). Additional benefits, such as the frequency of students completing high school algebra and higher scores on the college boards, have also been attributed to cumulative practice (Finn, 1988). Cumulative practice, then, has been associated with improvements in skill acquisition as well as high overall levels of posttest achievement.

Research on cumulative practice, however, has not isolated the critical elements of cumulative procedures that result in superior performance. Cumulative practice can be analyzed into at least three elements: review or maintenance training, mixing new items with previously learned items, and additional practice on the component skills. Any one of these elements may be sufficient to produce positive learning outcomes. For example, the benefits of cumulative practice may occur because of the extra practice that is a part of cumulative procedures. Much research has shown that extra practice facilitates performance (e.g., Beck, Perfetti, & McKeown, 1982; McKeown, Beck, Omanson, & Perfetti, 1983; Wilson, Majsterek, & Simmons, 1996). Alternatively, the positive effects may be due to the process of review that is embedded within cumulative practice procedures, where review is defined as practice distributed over time. Research has also shown the positive effects of review (e.g., Bahrick & Phelps, 1987; Dempster, 1988; Kryzanowski & Carnine, 1980; Melton, 1970; Pyle, 1913; Reynolds & Glaser, 1964; Rothkopf & Coke, 1966; Underwood, 1970). No controlled study has examined the effects of extra practice or review, however, compared to the effects of systematically mixing practice on new skills with practice on all previously mastered skills (cumulative practice). In addition, no study could be found that directly examined the effects of cumulative practice on problem solving.

The current study, therefore, compared a training method involving cumulative practice with methods involving extra practice and simple review practice. The three methods used the same mastery procedures for training each of five component algebraic skills and provided the same amount of practice beyond mastery through periodic review sessions. The methods varied, however, in terms of whether additional practice on the component skills provided in the review sessions was mixed with practice on all skills previously learned (cumulative practice), was provided directly after each skill was mastered (extra practice), or was provided on one component skill at a time (simple review). Both application tasks and problem-solving tasks involving novel combinations of the skills were used to evaluate the three training conditions.

## METHOD

### *Participants 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 for a course on finite math, which is the lowest level math class available for college credit. In addition, the participants had not received passing credit for either precalculus or calculus in high school. Participants also met the selection criteria during 2 days of pretesting on the exponents skills related to the study (described below).

Sessions were held in a room that contained four desks. Typically, only 1 or 2 participants reported to the experimental room at a given time.

### *Materials*

*Practice worksheets.* All participants received worksheets that provided training to mastery (see definition below) on each rule. Multiple versions of each worksheet were constructed by changing the numbers and letters (see the Appendix for an example of the worksheets; other worksheets are available from the authors upon request). On the first version of each worksheet for a particular rule, the rule was introduced along with an explanation and several examples of the

Table 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 1 | Rule 1 | Rule 2 |
| 10 | Test 3 | Test 3 | Test 3 |
| 11–13 | Rule 3 | Rule 3 | Rule 3 |
| 14 | Cumulative 2 | Rule 2 | Rule 3 |
| 15 | Test 4 | Test 4 | Test 4 |
| 16–18 | Rule 4 | Rule 4 | Rule 4 |
| 19 | Cumulative 3 | Rule 3 | Rule 4 |
| 20 | Test 5 | Test 5 | Test 5 |
| 21–23 | Rule 5 | Rule 5 | Rule 5 |
| 24 | Cumulative 4 | Rule 4 | Rule 5 |
| 25 | Test 6 | Test 6 | Test 6 |
| 26 (retention) | Test 7 | Test 7 | Test 7 |

rule. The five 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.

*Reviews.* The type of review presented to the three groups of participants was manipulated as the independent variable (review worksheets are available upon request from the authors). The cumulative practice group received a review that presented 50 practice items covering every rule that had been trained up to that review period (e.g., Rules 1, 2, and 3 were reviewed together after mastering Rule 3). The review items for each rule varied slightly across participants within the cumulative practice group, depending on which of the five versions of the cumulative review worksheets they received. The proportion of review items for each rule also varied across cumulative reviews. On the first review all participants in this group received 25 items on Rule 1 and 25 items on Rule 2. On the second review these participants received between 15 and 19 items on each of Rules 1, 2, and 3. On the third review these participants received 12 or 13 items on each of Rules 1 through 4. On the fourth review these participants received between 9 and 11 items on each of the five rules. Thus, the cumulative review participants received approximately 65 items that reviewed each of Rules 1 and 2, 40 items that reviewed Rule 3, 20 items that reviewed Rule 4, and 10 items that reviewed Rule 5. The simple review group practiced 50 items on one of the previously mastered rules for each review session (e.g., Rule 1 was reviewed after mastering Rule 2, Rule 2 was reviewed after mastering Rule 3, etc.). The extra practice group practiced another 50 problems of the rule they had just mastered (e.g., Rule 3 was reviewed after mastering Rule 3). Table 1 shows the distribution of review sessions for each group. By Test 6, all of the participants had received 200 practice items beyond mastery.

*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 item and the order of presentation of the application items (tests are available from the authors upon request).

The application items tested the participants' ability to apply the individual rules to solve problems that used different numbers and letters, and used two or three variables instead of only 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, the student might have been presented with an application item on a test such as $3a^6b^2 \cdot 5a^8b^2 = ?$.

Applications of all five rules were tested individually in the section containing application items. There were two items for each of the four exponent rules and four items 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 combinations of the trained rules could be solved correctly. For example, to solve the following: $[(3g^5 \cdot 8g^9)/4g^8]^2 = ?$, the student needed to combine four rules: multiplying variables and coefficients with exponents, dividing variables and coefficients with exponents, raising variables and coefficients with exponents to a power, and order of operations. The participants were never taught how to combine any of the rules; furthermore, they were shown problems that required combining the rules only on the tests.

There were 12 problem-solving items. Four items tested the combination of two exponent rules and the order of operations rule; four tested the combination of three exponent rules and the order of operations rule; and four 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*

*Participant assignment.* Students who met the participation requirements made an appointment for a pretest session. During the first pretest session, participants completed one version of the experimenter-designed test. All participants who correctly solved application problems on more than two of the exponent rules did not qualify to continue. Furthermore, participants who answered more than four (out of 24) parts correctly on the problem-solving section of the test did not qualify to continue.

Participants who did not qualify on the 1st day of pretesting were paid for their correct answers, given an extra-credit slip validating their participation, and dismissed. All participants 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. Participants whose performance on the 2nd day of pretesting exceeded the criteria were paid, given an extra-credit slip, and dismissed.

The overall test scores from both pretests were averaged. All participants whose scores fell in the qualifying range (10% to 39% correct) were asked to participate in the study. Participants were then further categorized based on their average pretest scores in order to match the skills of the participants across groups. Nine participants fell into the "low pretest" range (average score between 10% and 19% correct on Pretests 1 and 2); 15 participants scored in the "middle pretest" range (average score between 20% and 29% correct); and 9 participants performed in the "high pretest" range (average score between 30% and 39%). Within each range of scores, the participants were randomly assigned to each of the three groups. Each group began with 1 male and 2 females in the low pretest range and high pretest range. The cumulative and simple review groups had 3 females and 2 males in the middle pretest range, and the extra practice group had 4 females and 1 male in the middle pretest range. For the sake of description, particularly in Figure 2, each participant was assigned a number from S1 through S33. Each of these participants also took the Basic Algebra Exam published by the Mathematics Association of America and signed an informed consent form.

*Overview of practice sequences.* Table 1 provides the general layout of the practice sequences for each group. All participants were required to complete 25 problems of each rule with 100% accuracy for three sessions before progressing to the next step in the training. Participants were given unlimited attempts at meeting this mastery criterion during each of the three sessions. When this mastery criterion was met on 3 separate days, participants were given a review session (cumulative, simple review, or extra practice). A test was then administered during the next experimental session; and on the following day, the next component rule was introduced. Participants typically fulfilled the final mastery criterion (3 days of mastery performance) in three consecutive sessions. Therefore, the typical cycle of events was 3 days of practice on an individual rule, 1 day of review (experimental manipulation), and 1 day of testing.

*Procedure for daily sessions.* Individual sessions were conducted daily (Monday through Friday) for approximately 15 to 30 min per participant. In addition, makeup sessions were occasionally conducted on the weekends for participants who missed days during the week. During the final pretest session, general instructions about the study were provided for all participants (available from the experimenters on request). During the next session, training began on Rule 1. When Rule 1 and every subsequent rule were introduced, participants received Version 1 of the practice worksheets containing an explanation of the rule and appropriate examples (see example in the Appendix). If questions were asked, the relevant parts of the instructions were pointed out for the participant, or the participant was told that no further explanations could be given.

Participants 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 participants. Feedback included whether each response was correct or incorrect and an explanation of how to achieve the correct answer for all problems answered incorrectly. Cycles of practice and feedback occurred in this manner until a set of 25 problems was completed correctly. If a participant needed to leave before meeting the mastery criterion for that day, 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. All the exceptions involved such things as a prerequisite rule that a participant had to be told explicitly (e.g., $3^2 = 3 \cdot 3$) or a clarification of how two rules were not contradictory. No information about how to complete the problem-solving items was ever given.

On review days, participants were presented with the review worksheets for their group. 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, participants took a test. There were no specific time limits imposed on any of the tests, but the experimenter recorded the time taken to complete the section with application items and the time taken to complete the section with problem-solving items. Corrective feedback was not given for test performance.

After the sixth test, a retention interval occurred before the seventh test was administered to all participants who were available. The retention intervals were developed around the general availability of participants after the date they finished the main body of the study and were dictated by the semester and summer schedule of classes at WVU. Seven participants from each group were available for retention testing. Four participants from each group were tested after a short retention interval (4 to 6 weeks), and 3 from each group were tested after a longer interval (9 to 12 weeks).

*Reinforcement procedures.* The reinforcement procedures were described to the participants in the general instructions administered at the beginning of the study. Participants earned money for correct answers on all the tests and were not penalized for incorrect answers. During the session following a test, participants were presented with a record of their total earnings on the test. This was the only feedback provided concerning their performance on the tests.

Participants 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. Participants 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.* Agreement was calculated for scoring both the tests and practice worksheets. All tests were scored on a separate coding sheet so as to leave the participants' answers intact with no marks from the scorer. Approximately 20% of the tests administered during the main body of the study were rescored by a second grader without seeing the original scores, 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 practice and review sessions were also randomly checked. If the second observer disagreed with any grading by the first observer (e.g., one part of a problem was marked incorrect by one but not the other observer), the whole session of 25 or 50 items was counted as a disagreement. This conservative measure was necessary because one disagreement in the grading of a practice item could potentially have led to the participant reaching the mastery criterion when he or she should have been given more practice. Therefore, using this stringent measure of interobserver agreement, the total agreements for all the participants was divided by the total number of practice sessions rescored (number of agreements plus disagreements), and the quotient was multiplied by 100%. This formula resulted in an interobserver agreement score of 89% for the practice and review sessions.

## RESULTS

### *Pretests, Practice, and Review Performance*

There were five pretest measures of mathematics performance administered to all participants. Table 2 presents a summary of descriptive and inferential statistics for all five

Table 2
Summary of Pretest Results

| | Cumulative practice | | Simple review | | Extra practice | | |
|---|---|---|---|---|---|---|---|
| Measure | *M* | *SE* | *M* | *SE* | *M* | *SE* | *F* (2, 30) |
| Application % correct | | | | | | | |
| Test 1 | 38.1 | 4.92 | 36.7 | 5.00 | 36.8 | 4.07 | 0.027 |
| Test 2 | 44.0 | 3.34 | 43.2 | 3.54 | 43.0 | 4.86 | 0.018 |
| Application rate[a] | | | | | | | |
| Test 1 | 1.72 | 0.26 | 1.58 | 0.22 | 1.61 | 0.27 | 0.079 |
| Test 2 | 3.12 | 0.59 | 2.20 | 0.37 | 2.46 | 0.45 | 0.974 |
| Problem-solving % correct | | | | | | | |
| Test 1 | 4.1 | 1.27 | 3.7 | 1.64 | 4.0 | 0.76 | 0.022 |
| Test 2 | 3.4 | 1.59 | 2.3 | 1.55 | 1.5 | 0.81 | 0.493 |
| Problem-solving rate[a] | | | | | | | |
| Test 1 | 0.13 | 0.04 | 0.09 | 0.04 | 0.12 | 0.03 | 0.341 |
| Test 2 | 0.08 | 0.05 | 0.05 | 0.04 | 0.04 | 0.02 | 0.191 |
| General algebra skills % correct | | | | | | | |
| | 18.9 | 3.54 | 15.3 | 3.40 | 15.6 | 2.38 | 0.404 |

*Note.* None of the *F* values were significant at the $p < .05$ level. *M* = mean; *SE* = standard error.
[a] Measured in number of correct responses per minute.

measures. The first four performance measures were assessed using Test 1 and Test 2 of the experimenter-designed tests. These tests covered skills specifically related to the current study and were administered on two consecutive daily sessions. The four measures evaluated were percentage correct on application items, rate of correct application items, percentage correct on problem-solving items, and rate of correct problem-solving items. Although the aggregate score of the cumulative group appeared to be slightly higher on these measures, the numerical differences were small and the individual variability within groups was large, such that differences between groups were not 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

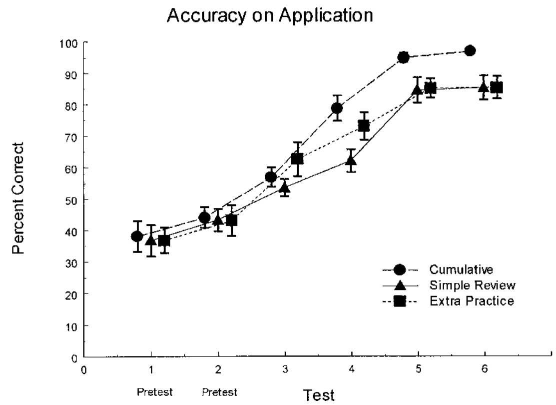


Figure 1. Mean percentage correct on application items for all three groups across tests. Tests were administered to all groups at the same time, although data points are spaced apart for legibility. Error bars represent ±1 standard error.

were no statistical differences among groups on any of the pretest measures. There also were no group differences on the number of attempts needed to master each rule or the accuracy of review session performance.

*Application*

Figure 1 shows the mean percentage correct performance and standard errors on application items for all groups across the six tests. The performance of all groups improved across tests, but systematic differences were found between the cumulative practice group and the other two groups on Test 6. The cumulative group's average score was 97% correct (*SE* = 0.90%); the simple review group's mean score was 85.4% correct (*SE* = 3.8%); and the extra practice group's mean score was 85.5% correct (*SE* = 3.6%). In addition, these mean differences appear to be representative because the standard errors are so small. To statistically analyze the mean differences in application accuracy, a one-way analysis of variance (ANOVA) was conducted on Test 6. The overall difference in mean scores was statistically significant on Test 6, $F(2, 30) = 4.72$, $p < .05$. Moreover, comparisons revealed that the cumulative

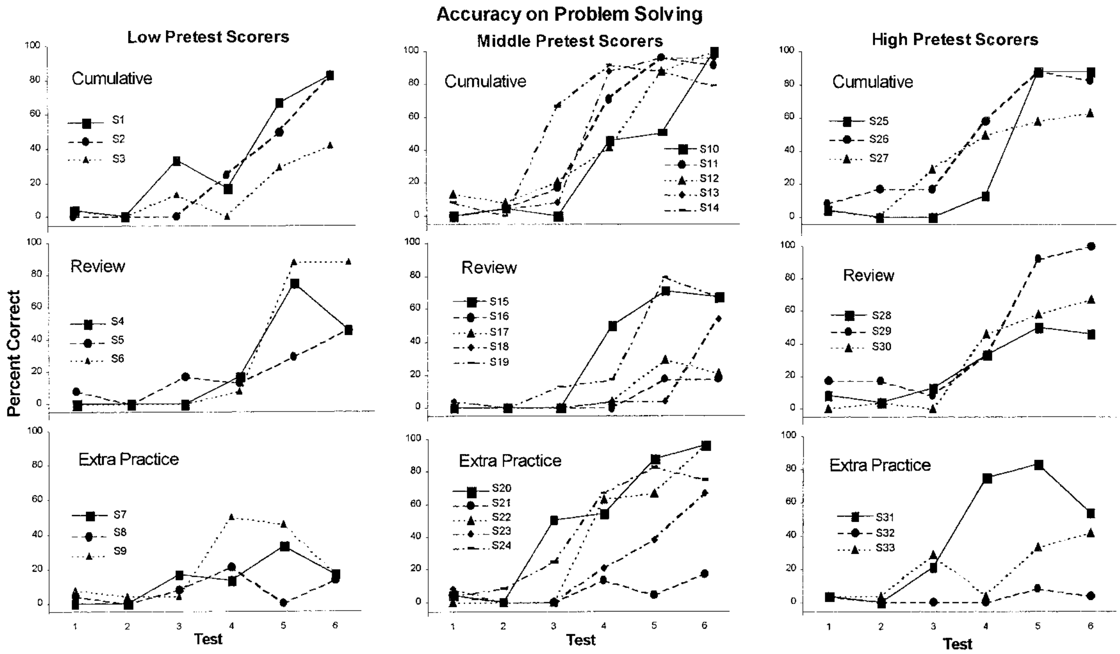


Figure 2. Percentage correct on problem-solving items across tests for low pretest scorers (S1 through S9), middle pretest scorers (S10 through S24), and high pretest scorers (S25 through S33). Cumulative, review, and extra practice participants are shown in the top, middle, and bottom rows, respectively.

group outperformed the simple review group, $t(11.11) = 2.95$, $p < .05$, and the extra practice group, $t(11.26) = 3.12$, $p < .05$. [The degrees of freedom for the $t$ tests are appropriate for groups with unequal variances, as detected by Levine's test for homogeneity of variance, Levine statistic(2, 30) = 5.65, $p < .01$.]

A large amount of variability in the individual rates of correct responses on the application items contributed to a lack of statistical differences among groups. Therefore, data on rate of application items are not presented.

### *Problem Solving*

Figure 2 displays accuracy of performance on problem-solving items on the experimenter-designed tests. Individual-participant data are presented on this measure because problem-solving accuracy was the primary dependent variable. The data are grouped by participants who had low, middle, and high pretest scores, as well as by experimental group. As shown across the graphs, the performance of most participants generally improved across tests, but those in the cumulative group improved the most. Eight of the 11 cumulative participants scored over 80% correct on the final test. In contrast, only 2 of the 11 simple review participants and 2 of the 11 extra practice participants scored over 80% correct on the final test. The low pretest participants in the extra practice group did particularly poorly, with all 3 scoring below 20% correct on the final test.

The top panel of Figure 3 shows the mean percentage correct and standard errors on problem-solving items across tests for all groups. To analyze the mean differences in problem-solving accuracy, a two-way mixed measures ANOVA with a between-participants factor of group and a repeated within-participant factor of test was conducted. Because the Group × Test interaction was significant, $F(10, 150) = 3.09$, $p < .01$, simple

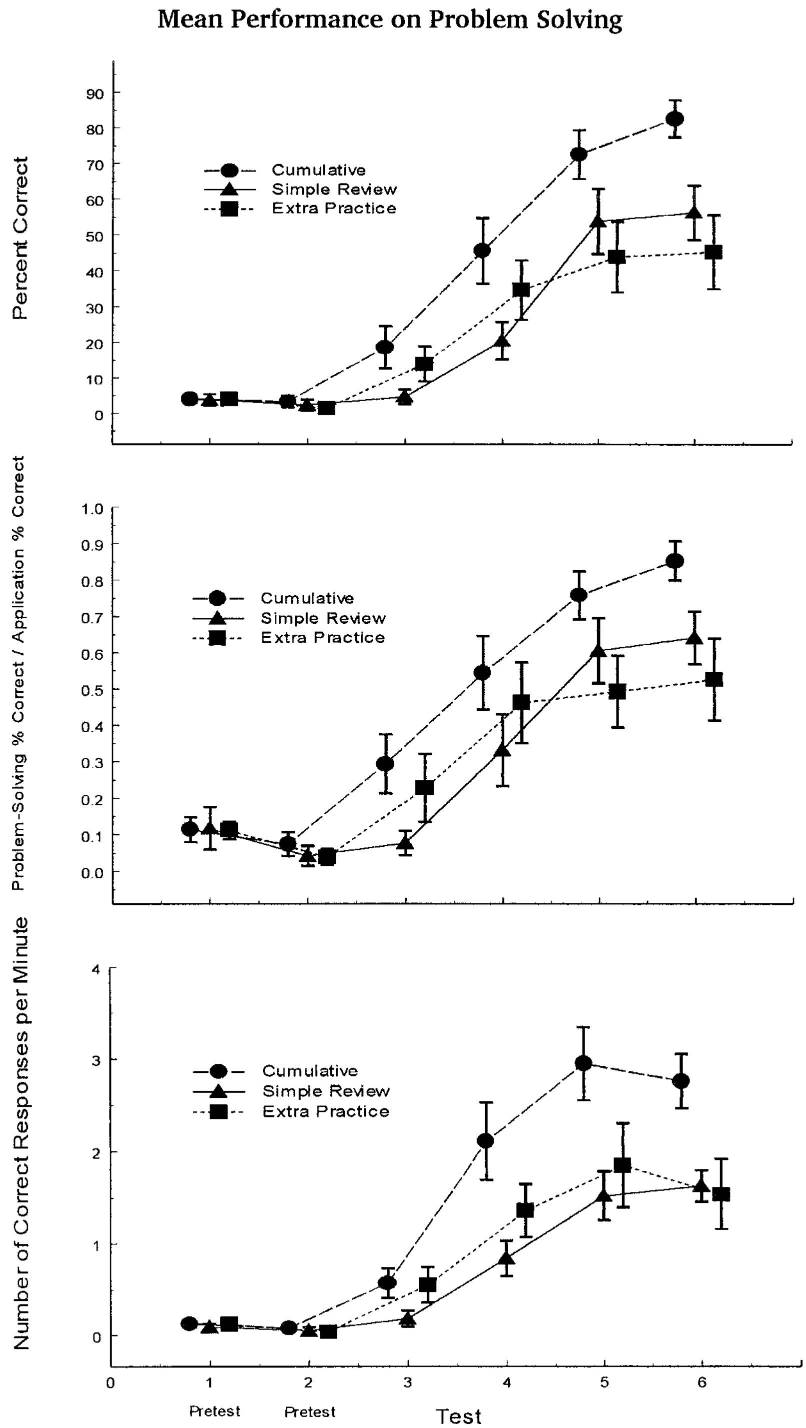


Figure 3. Mean group performance on problem solving. The top panel displays mean percentage correct on problem-solving items for all three groups across tests. The middle panel displays mean proportional problem-solving scores for all three groups across tests. The bottom panel displays mean number of correct problem-solving items per minute for all three groups across tests. Tests were administered to all groups at the same time, although data points are spaced apart for legibility. Error bars represent ±1 standard error.

effects of group at each test were examined. There were no significant differences in performance on Test 1, Test 2, Test 3, Test 4, or Test 5. There was a significant difference on Test 6, however, $F(2, 30) = 5.78$, $p < .01$. The cumulative group ($M = 82.6\%$, $SE = 5.2\%$) outperformed the simple review group ($M = 56.3\%$, $SE = 7.6\%$), $t(17.62) = 2.86$, $p < .05$, and the extra practice group ($M = 45.3\%$, $SE = 10.3\%$), $t(14.73) = 3.23$, $p < .01$. [The degrees of freedom for the $t$ tests are appropriate for unequal variances, Levine statistic(2, 30) = 4.47, $p < .05$.] There was no significant difference between the means of the simple review group and the extra practice group, however.

A further analysis was conducted on the mean problem-solving performance at Test 6 to see if the differences in accuracy among groups were due solely to differences in the direct acquisition of the component rules (as measured by application performance). To conduct this analysis, a proportional problem-solving score was calculated for each participant by dividing the percentage correct on the problem-solving items on Test 6 by the percentage correct on the application items on Test 6. For example, using this relative analysis, a participant who scored 100% on the problem-solving section and 100% on the application section would receive the same score (1.0) as a participant who scored 65% on both sections.

The middle panel of Figure 3 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. [The degrees of freedom are appropriate for unequal variances, Levine statistic(2, 30) = 7.36, $p < .01$.] Once again, however, there was no significant difference between the means of the simple review and extra practice groups. Therefore, the same relation among the means was found on this relative measure of problem-solving performance. On average, then, the cumulative participants gained more in problem-solving accuracy than the simple review and extra practice participants, even after controlling for individual levels of rule acquisition, as reflected in the application scores.

The rate of correct responses on the problem-solving items was also analyzed. The bottom panel of Figure 3 displays the mean number of correct responses per minute and standard errors across tests for all groups. To analyze the mean differences in the rate of problem-solving performance, a two-way mixed ANOVA, with a between-participants factor of group and a repeated within-participant factor of test, was used. A significant interaction of Group × Test was found, $F(10, 150) = 3.26$, $p < .01$. There were no differences among groups on Test 1, Test 2, or Test 3. There was a difference in performance on Test 4, however, $F(2, 30) = 4.20$, $p < .05$. The cumulative group ($M = 2.11$, $SE = 0.42$) performed significantly faster than the simple review group ($M = 0.84$, $SE = 0.19$), $t(14.09) = 2.78$, $p < .05$. [The group variances were unequal, Levine statistic(2, 30) = 5.36, $p < .05$.] On Test 5, the cumulative group ($M = 2.95$, $SE = 0.40$) again outperformed the simple review group ($M = 1.52$, $SE = 0.26$), $t(30) = 2.66$, $p < .05$. On Test 6, the cumulative group ($M = 2.76$, $SE = 0.30$) responded faster than both the simple review group ($M = 1.63$, $SE = 0.17$) and the extra practice group ($M = 1.54$, $SE = 0.38$), $t(30) = 2.70$, $p < .05$, and $t(30) = 2.93$, $p < .01$, respectively. There was no significant difference between

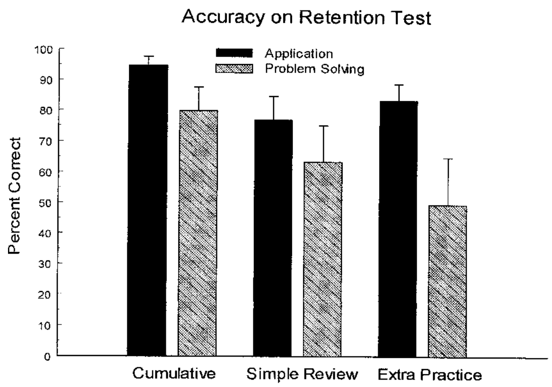


Figure 4. Mean percentage correct on application and problem-solving items for all three groups on the retention test. Error bars represent $\pm 1$ standard error.

the mean rates of the simple review group and the extra practice group, however.

*Retention*

Accuracy of performance on both application and problem-solving items is displayed in Figure 4 for the 21 participants who completed Test 7 (the retention test). The data were analyzed using one-way ANOVAs with a between-participants variable of group. The analyses revealed no significant differences on any of the measures.

## DISCUSSION

The current study compared the effects of cumulative practice, simple review practice, and extra practice on problem-solving and application skills. Three groups of college students with poor mathematics skills learned to use 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 cumulative group received reviews that intermixed component skills; the simple review group received reviews on one component skill at a time; and the extra practice group received one review session of additional practice on the component skills.

The pretests showed that all three groups of participants started 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 or on each group's performance on the review sessions. Results from the final test administered to all groups, however, indicated that the cumulative participants significantly outperformed the other participants on accuracy of application and problem-solving skills as well as rate of correct performance on problem-solving skills.

The effects of cumulative review were also clinically significant. The cumulative group's average performance of 97% on application items was more than 10% higher than the average performance of the other groups. In a college course this would represent a difference in accuracy of more than a letter grade. In addition, the cumulative group's average performance on the final test of problem-solving skills was 82.6% correct. Not only was this better than the average performance of the other two groups (at 56% for the simple review and 45% for the extra practice groups), it was a vast improvement over the average pretest score of approximately 4% correct for the cumulative group. Students who answered an average of approximately one 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. In addition, the average retention score for the cumulative group was 95% correct on application items and 80% on problem-solving items. Considering the many weeks of no practice, this level of performance is clinically important despite the overall finding of no significant differences among groups on the retention test.

Two of the participants in the cumulative group scored poorly on the problem-solving items of Test 6 (S3, 43%; S27, 62%). Conversely, levels of performance comparable to the majority of the cumulative group were found for 2 participants in the simple review group (S6, 88%; S29, 100%) and 2 participants in the extra practice group (S20, 96%; S22, 96%). These data indicate that some participants perform poorly and others perform well, regardless of the method of review implemented. However, given that there was so little overlap between the cumulative group and the other two groups, the overall results suggest that the majority of the students may suffer when cumulative review is not used.

The cumulative program also led to an improved ratio of learned skills to novel behavior. It may be important to educators to measure the ratio of trained skills to untrained combinations of skills. That is, for every skill that a student can demonstrate an "understanding" of (i.e., through some application task), one might measure whether the student can solve a novel problem that uses that skill (i.e., assuming the other skills required to complete the novel problem are also understood). Thus, the proportional problem-solving analysis of the current experiment served as a measure of this ratio. This analysis revealed that, on average, for every correct application item, the cumulative participants were able to complete about .85 novel problem-solving items. This ratio of "novel learning" was significantly better than that of the other two groups (only .64 and .53 novel items per application item for the simple review and extra practice groups, respectively). Thus, the cumulative program led not only to a better understanding of the trained skills (as seen in the differential performance on the application items) but also to a greater gain in the ratio of novel behavior per understood skill, suggesting the greater effectiveness and efficiency of the training.

*Retention Results*

Statistical analyses of retention performance did not find any significant results. Although this outcome suggests that the long-term effects of cumulative review should be viewed with caution, there are a number of factors that are likely to have contributed to this lack of significance. Probably the most critical were an increase in variability within each group coupled with a decrease in statistical power resulting from fewer participants participating in the retention tests. The increase in variability may have been a function of the different retention intervals, because participants were tested between 4 and 12 weeks after taking Test 6. The decrease in power resulted from only 7 of the 11 participants being available for retention testing. Clearly future research should address questions of retention and should standardize the retention interval as well as ensure that a higher proportion of the participants complete the retention tests.

*Reasons for the Effectiveness of Cumulative Practice*

There seem to be at least three explanations of the current findings that could be explored further. First, it is possible that the differences were due to the fact that the cumulative group reviewed all the rules and the other groups did not. Although the study controlled for the number of review sessions (four) and the number of review items (200), there was differential exposure to the individual rules; the extra practice participants never reviewed Rule 1, and the simple review participants never reviewed Rule 5. Thus, one potential reason for the group differences was that the cumulative participants had reviewed all the rules and the other participants had not.

A related reason for the group differences is that on the session before Test 6 the cumulative participants reviewed and received differential feedback on all of the rules they had previously learned. Differential feedback generally can have two operant functions, a discriminative function and a reinforcement function. If participants in the cumulative group had difficulty with any of the previously learned rules, they practiced the application of those rules in a condition in which correct responding was cued and reinforced via differential feedback on the session before the tests. By contrast, participants in the other two groups received differential feedback on just one rule on the day prior to Test 6; thus, that feedback may have affected performance only on that rule on during the test.

If the differences in reviewing and receiving feedback on individual rules accounted for the differences, then one would expect that there would also be differential maintenance of the application items for rules reviewed versus those not reviewed. The data indicate that this was not the case. The proportion of errors on application items for the participants in the simple review and extra practice groups was examined as a function of the rules tested in the application items. Each test consisted of 12 application items that were divided into 32 parts that could be correct or incorrect. There were seven parts testing each of the first four rules (28 total parts) and four parts testing the fifth rule. If the errors were evenly distributed across rules according to the relative proportion of items representing each rule, then approximately 22% of the errors would occur on the parts testing each of first four rules and 12.5% of the errors would occur on the parts testing Rule 5. Data from Test 6 indicate that the extra practice participants performed as well on application items related to Rule 1 as on the items testing the other rules, with 21% of the errors occurring on Rule 1 items. Similarly, when the application items for the simple review participants on Test 6 were examined, only 13% of the errors occurred on Rule 5 items. These data suggest that the participants in these groups maintained performance on rules for which they did not explicitly review and receive differential feedback as well as on those they did review.

In addition, even if an explanation based on review and feedback accounted for differences in performance on application items, such an explanation is not sufficient to account for the observed difference on the problem-solving items. These items were always novel; the participants never received practice or differential feedback for these items. Although all three groups improved on the problem-solving items over the course of the experiment, the cumulative group performed at a level approaching mastery on items for which they never received training ($M$ = 83%), whereas the other two groups never came close to this level of performance.

A second interpretation of the results is in terms of other behavioral concepts involved in the review procedures. The contingencies involved in practice and mastery of each of the individual skills ensured discrimination. Practice on each skill across a range of examples also ensured generalization and extension (Skinner, 1957). Rules were also taught that made it likely that the skills would occur in the presence of their relevant signs. The effect of rules like this can probably best be understood in terms of equivalence (Chase & Danforth, 1991). These behavioral concepts, however, were evident in the procedures for all three groups. Thus, other behavioral concepts should be examined to interpret the differences between the cumulative group and the other groups.

It is probable that the concepts from the literature on behavioral variability and problem solving are most critical for understanding the differences among the conditions. As

Skinner (1966) and others (e.g., Page & Neuringer, 1985) have speculated, providing individuals with contingencies for varying their behavior may be an important determinant of the variability that is required to solve novel problems. Variability in behavior was repeatedly induced across reviews for the cumulative participants by the juxtaposition of two, three, four, and five different kinds of problems within the same review sessions. This juxtaposition of different kinds of problems may have facilitated learning to discriminate when to apply each rule and also when to switch between rules. This multiple discrimination then may have facilitated the novel combinations of these rules in the presence of the problem-solving items. Switching between the use of different rules on successive problems in the review sessions for the cumulative participants certainly more closely approximated the combining of these rules than the reviews given the other participants. During the review sessions received by the simple review and extra practice groups, participants always answered the same kind of item, and they never explicitly had to discriminate between rules. Thus, similarities between the stimulus conditions of cumulative practice and the problem-solving items during testing may have facilitated the transfer of stimulus control from the former to the latter.

Previous research supports these interpretations. Fink and Brice-Gray (1979) showed that cumulative reviews involving many simultaneous discriminations produced superior performance on identifying five two-syllable words compared with cumulative reviews that required fewer simultaneous discriminations. Clark and Sherman (1975) found that when new concepts were introduced into a training program, responding to previously trained concepts decreased unless the new concepts were trained concurrently with the old concepts. Although these studies did not test performance on problem-solving items, they suggest that cumulative review is a form of simultaneous discrimination training that facilitates learning.

*Conclusion*

In summary, the results from the current 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 advanced mathematics 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 to be adequate to produce similar results, particularly on problem-solving skills.

Furthermore, based on the participant 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 novelty is simultaneously programmed through the use of cumulative review procedures. The end result is that students perform well on the component skills, extend the individual skills to novel situations (application), and synthesize the skills into novel solutions derived from combinations of the skills (problem solving). These are the results that mathematics educators seek.

One obvious advantage of adopting this approach to mathematics education is the built-in efficiency of the instruction. Instead of attempting to cover every skill and its application, teachers can simply teach a core set of skills and strategies for synthesizing them that will produce novel combinations and applications without further training. These findings suggest the importance of incorporating cumulative practice into future curriculum and instruction for mathematics education and possibly other areas (e.g., reading, music, etc.) that involve hierarchical sets of skills and sequential patterns of learning.

## 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.

Bahrick, 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.

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.

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. (1997). Instructional design in mathematics for students with learning disabilities. *Journal of Learning Disabilities, 30,* 130–141.

Carnine, D., Jones, E., & Dixon, R. (1994). Mathematics: Educational tools for diverse learners. *School Psychology Review, 23,* 406–427.

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Lindquist, M. M., & Reys, R. (1980). Results of the second NAEP mathematics assessment: Secondary school. *The Mathematics Teacher, 73,* 329–338.

Carpenter, T. P., Kepner, H., Corbitt, M. K., Lindquist, M. M., & Reys, R. E. (1980). Results and implications of the second NAEP mathematics assessment: Elementary school. *Arithmetic Teacher, 27,* 10–12, 44–47.

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: Erlbaum.

Chase, P. N., & Danforth, J. S. (1991). The role of rules in concept formation. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 205–225). Reno, NV: Context Press.

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.

Dempster, F. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43,* 627–634.

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.

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., & Mousetis, 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, 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.

Klingele, 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.

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.

Melchiori, L. E., de Souza, D. G., & de Rose, J. C. (2000). Reading, equivalence, and recombination of units: A replication with students with different learning histories. *Journal of Applied Behavior Analysis, 33,* 97–100.

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.

Mueller, M. M., Olmi, D. J., & Saunders, K. J. (2000). Recombinative generalization of within-syllable units in prereading children. *Journal of Applied Behavior Analysis, 33,* 515–531.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Retrieved from http://standards.nctm.org/previous/CurrEvStds

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Retrieved from http://standards.nctm.org/protoFINAL/cover.html

Page, S., & Neuringer, A. (1985). Variability is an operant. *Journal of Experimental Psychology: Animal Behavioral Processes, 11,* 429–452.

Pyle, W. H. (1913). Economical learning. *Journal of Educational Psychology, 3,* 148–158.

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.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Skinner, B. F. (1957). *Verbal behavior.* Englewood Cliffs, NJ: Prentice Hall.

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.

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.

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.

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: 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.

Wood, D. K., Frank, A. R., & Wacker, D. P. (1998). Teaching mathematics facts to students with learning disabilities. *Journal of Applied Behavior Analysis, 31,* 323–338.

*Received July 30, 2001*
*Final acceptance February 20, 2002*
*Action Editor, Robert Stromer*

APPENDIX

SAMPLE PRACTICE WORKSHEET

## 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 $a^{2+3} = a^5$

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

$(g^3)(g) = g^4$ (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 $5 = 5^1$.)

---

| | | | |
|---|---|---|---|
| $2d^5 \cdot 5d^4 =$ | $8h^5 \cdot 5h^9 =$ | $(9r)(6r^3) =$ | $3^7w^4 \cdot 3^8w^4 =$ |
| $6g^4 \cdot g =$ | $(2^5s^7)(2^9s^9) =$ | $(8^3n^1)(8^7n^8) =$ | $5c^4 \cdot 6c^2 =$ |
| $7h^6 \cdot 8h^0 =$ | $4^5y^3 \cdot 4^6y^6 =$ | $(7k)(4k^3) =$ | $9x^9 \cdot 8x^6 =$ |
| $6b^8 \cdot 6b^3 =$ | $(3t^9)(2t^4) =$ | $5^6f^3 \cdot 5^8f^8 =$ | $5j^8 \cdot 7j^5 =$ |
| $9e^9 \cdot 5e^8 =$ | $8v^7 \cdot 3v^4 =$ | $(4r)(7r^7) =$ | $4u^4 \cdot 5u^8 =$ |
| $2^0g^2 \cdot 2^9g =$ | $(1y^9)(y^9) =$ | $(6n^4)(8n^8) =$ | $8d^9 \cdot 4d^3 =$ |
| $8^7t \cdot 8t^4 =$ | $9w^9 \cdot 7w^3 =$ | $(2r)(5r^2) =$ | $3^2x^8 \cdot 3^8x^5 =$ |

## *STUDY QUESTIONS*

1. How did the authors define problem solving, and what did they consider to be one of its key components?

2. What five rules were trained via the practice sheets?

3. Describe the differences among the three review procedures.

4. What were the two general sections into which the tests were divided, and what was the purpose of each?

5. What was the criterion for mastery of a rule, and why was it important to the study?

6. What reinforcement contingencies were used in the study?

7. Summarize the results depicted in Figure 2.

8. What explanations did the authors provide to account for observed differences across groups?

Questions prepared by Stephen North and David Wilson, The University of Florida