Effects of Contextualized Math Instruction on Problem Solving of Average and Below-Average Achieving Students

Brian A. Bottge, University of Wisconsin–Madison

The purpose of the study was to investigate the effect of contextualized math instruction on the problem-solving performance of 17 middle school students in one remedial class and 49 middle school average-achieving students in two prealgebra classes. The study employed experimental and quasi-experimental designs to compare the impact of word problem instruction and contextualized problem instruction on computation skills and problem-solving performance. Results showed that students in the contextualized problem remedial and prealgebra groups outperformed students in the word problem groups on a contextualized and a transfer problem. In an extended transfer activity, students in the remedial class applied what they had learned in order to plan and build two skateboard ramps. Results support the use of contextualized problems to enhance the problem-solving skills of students in general and remedial classes.

All students, including those with learning difficulties, need to be mathematically proficient to a level that will allow them to "figure out" math-related problems they encounter in the community and in future work situations. Unfortunately, evidence clearly shows that many students, not just those in special programs, do not have these skills. The National Research Council (NRC, 1989) warned that the mathematics skills of American children are woefully inadequate for the kinds of problem solving required in the workplace. This claim is supported by 1992 results of the National Assessment of Educational Progress (NAEP), which showed only about half (59%) of the 12th-grade students could solve problems beyond whole-number computation.

The outlook for students who have special difficulty in learning mathematics is even gloomier. Studies have reported that 16- and 17-year-old students with learning disabilities score at about the fifth-grade level in computation and application (Cawley, Kahn, & Tedesco, 1989; Cawley & Miller, 1989), can demonstrate only limited proficiency in tests of minimum competency at the end of secondary school (Algozine, O'Shea, Crews, & Stoddard, 1987), and function at least two grade levels below expectancy (Cawley, Fitzmaurice, Goodstein, Kahn, & Bates, 1979). Rather than catching up to the other students, these students have even larger age-in-grade discrepancy (Cawley, Parmar, Yan, & Miller, 1998), made worse by high dropout levels of students with learning and emotional disabilities (Phelps & Hanley-Maxwell, 1997).

The weaknesses in problem solving of students with disabilities can be traced to confusion over what constitutes problem solving and how to teach it. Some of the most intractable teaching practices in remedial classrooms involve withholding introduction of more complex and interesting content until easier material is mastered (Hiebert et al., 1996; Knapp & Turnbull, 1990) and emphasizing skill deficiencies rather than skill strengths (Means & Knapp, 1991). These practices are fostered by beliefs among many educators that (a) math is a set of rules that require memorization, (b) computation problems are always solved by using algorithms, (c) problems always have one correct answer, and (d) people who use mathematics are geniuses (Mtewà & Garofalo, 1989).

But real problems are often ill defined, and their solutions do not follow a linear, prescribed route (Polya, 1962). In mathematics, Schoenfeld (1989) described a problem as "a task (a) in which the student is interested and engaged and for which he wishes to obtain a resolution, and (b) for which the student does not have a readily accessible mathematical means by which to achieve that resolution" (p. 88). Notable scholars such as Wertheimer (1959), Bruner (1960), and Hiebert et al. (1996) have urged teachers to challenge students to find solutions to problems that interest them. In response to these suggestions, the National Council of Teachers of Mathematics (NCTM; 1989) recommended that teachers focus on tasks that encourage students "to explore, to guess, and even to make and correct errors so they gain confidence in their ability to solve complex problems" (p. 5). The challenge for
teachers, therefore, is to find problem-solving activities that are "authentic" and important to the learner and yet manageable in the school context (Brown, Collins, & Duguid, 1989; Dewey, 1926; Engler, Tarrant, & Mariage, 1992).

The theoretical underpinnings of mathematics problem solving derive from cognitive psychology. There are two distinct aspects of the problem solving: skill acquisition and generalization. For the most part, research in the area of special education has remained fixed on skill acquisition by teaching students strategies, or heuristics, when approaching school-based math problems. The logic of this approach is that students can solve such problems if they follow the cognitive trail of expert problem solvers and monitor their progress as they reach the problem goal. Results by Montague and her associates suggest that cognitive strategy instruction combined with metacognitive training for middle and secondary students with disabilities helps them solve one-, two-, and three-step mathematical word problems (Montague, 1997; Montague, 1992; Montague & Applegate, 1993; Montague & Bos, 1986).

Yet, there is little evidence to show that these gains are maintained in the experimental settings or successfully generalized to other situations (Ginsburg, 1997). This failure to maintain and generalize results has been a frustration throughout this century. Vygotsky (1978) and Whitehead (1929), for example, described this inability to use an appropriate math application as "fossilized behavior" and "inert knowledge." According to Bruner (1960), teachers can foster generalization by ensuring that students understand the underlying structure of the problem, by involving students in challenging and meaningful problems, by relying on students' intuition to arrive at a plausible yet tentative solution, and by stimulating in students a desire to learn. Quite simply, students should have a firm grasp on the nature of a problem, a sense that solving it is important, and an opportunity to develop intuitive, albeit imperfect, understanding about how to solve it.

The past failure to generalize problem-solving skills may have been exacerbated by researchers and educators who have inadvertently stifled student interest by smoothing out the curriculum into computational and simplistic formats, thereby limiting students' autonomous learning capabilities (Doyle, 1988). Rather than capitalizing on the insights and motivation that students bring to the classroom, schools may actually be wasting valuable time by withholding authentic problems until all "prerequisite" skills are acquired (Bruner, 1960).

One promising approach for improving generalization of mathematics problem solving is "anchored instruction," which enables students to explore semantically rich learning environments with the knowledge they bring to the learning situation (see Cognition and Technology Group at Vanderbilt, 1997). A series of 15- to 20-minute video-based vignettes called the Adventures of Jasper Woodbury are video-based problems in which the mathematics problems are not explicitly stated or well formulated. It is up to the students to determine what information is relevant and how this information can be used to solve the problem. The anchors help foster generalization by providing motivating and meaningful contexts in which students can develop their "intuitions" in combination with their computation and algorithmic skills to arrive at a plausible, yet tentative, solution.

Although the Jasper adventures have been used across all achievement levels, other anchors have been developed to improve problem-solving skills of students with learning disabilities in mathematics. Another series of video-based problems, not as difficult as the Jasper adventures, has been effective in helping students with disabilities solve meaningful problems (Bottge & Hasselbring, 1993). One of the adventures, called Bart's Pet Project, asks students to show how they can build a pet cage according to dimensions provided in the cage plan and with the money available to them. A central task in solving the problem is to add mixed numbers so that little lumber is wasted. In the Jasper study, more students in the video group were better at solving two transfer problems than students in the word problem group. An unexpected result was improved ability of students in the video group to solve word problems. This finding was important because it suggested that solving more interesting and complex problems may also improve performance of problems in more traditional problem formats, for example, word problems.

Whereas the Jasper study clearly showed that high school students with learning problems in mathematics were able to solve the video-based problem, it did not test whether students could actually apply what they had learned in an authentic task such as planning and building a wood project from schematic plans. The general goal of the present research was to replicate and extend previous findings by investigating students' performance in several problem-solving contexts and to test their ability to transfer this performance to an authentic task.

Method

Participants

Prior to the study, a math teacher with 8 years of experience and a special needs teacher in her second year shared the responsibilities for teaching the remedial math class. During intervention, it was decided that the special needs teacher would teach the remedial contextualized problem (CP) group because she wanted the opportunity to approach problem solving in a new way. The math teacher taught the remedial word problem (WP) and both prealgebra classes. A technology education teacher with more than 20 years of experience helped plan and supervise the construction of wooden skateboard ramps.

A total of 66 eighth-grade students participated in the study. All attended a middle school in a rural school district in the upper Midwest. Descriptive information on the students is provided in Table 1. Seventeen of the students attended a
remedial math class designed to improve their computation and functional math skills. Six of the students in this class met the Wisconsin criteria for receiving special services in at least one disability area: two in learning disabilities, one in learning disabilities and speech and language, one in speech and language, one in emotional disabilities, and one in other health impaired (attention deficit disorder). These students received special services an average of 1,098 minutes a week (range = 940 minutes to 1,360 minutes) that included general classroom instruction time when the special education teacher was present. Records indicated that these students had received from 1 to 10 years of special services. The other 12 students had been recommended for the remedial class by their math teacher the previous year.

Remedial math students were assigned to either the WP group or the CP group based on their performance on the computation pretest using a matching procedure recommended by Borg and Gall (1989). Test scores were ranked from highest to lowest, from which pairs of scores were identified. Students within each pair were then randomly assigned to either the WP or the CP group. Scheduling problems prevented assigning prealgebra students to instructional groups. Instead, each prealgebra class was randomly assigned to the CP group or the WP group. Thus, two WP groups and two CP groups were formed.

The Wisconsin Student Assessment System (WSAS) is a statewide test administered yearly to students in Grades 4, 8, and 10. A 2 × 2 analysis of variance (ANOVA) on WSAS scores was used to compare differences between classes (remedial and prealgebra) and instructional groups (WP and CP) prior to intervention. As expected, prealgebra students scored higher than the remedial students on the WSAS. Results showed a main effect for class, F(1, 62) = 12.92, p < .001, but not for instructional group, F(1, 62) = 0.224, p = .64. The interaction between class and instructional group was not significant, F(1, 62) = 3.50, p = .07.

The WSAS also indicated math achievement on a 4-point scale (1 = minimal performance, 2 = basic, 3 = proficient, and 4 = advanced). Chi-square analyses for class and instructional group on performance levels were conducted. Results indicated no significant difference between classes, χ²(3, N = 66) = 6.84, p = .08, or between instructional groups, χ²(3, N = 66) = 2.66, p = .45. However, visual inspection of the cells indicated lower levels of achievement by students in

<table>
<thead>
<tr>
<th>TABLE 1. Description of Participants in Each Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive variables</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Good at math</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Like math</td>
</tr>
<tr>
<td>All of the time</td>
</tr>
<tr>
<td>Most of the time</td>
</tr>
<tr>
<td>Some of the time</td>
</tr>
<tr>
<td>Not at all</td>
</tr>
<tr>
<td>State math assessment</td>
</tr>
<tr>
<td>Performance level</td>
</tr>
<tr>
<td>Minimal proficient</td>
</tr>
<tr>
<td>Basic</td>
</tr>
<tr>
<td>Proficient</td>
</tr>
<tr>
<td>Advanced</td>
</tr>
<tr>
<td>Mean Performanceb</td>
</tr>
</tbody>
</table>

Note: WP = Word Problem, CP = Contextualized Problem.

aOne student did not complete this item. bRepresented in Normal Curve Equivalents (NCE).
the remedial group. More than 50% of these students in both instructional groups scored at the lowest performance level.

Students also completed a six-item questionnaire prior to the study. It was developed by the author to probe student perceptions about their math experiences. Chi-square analyses for class and instructional group were conducted on two questions. One question asked whether they liked math; no differences were found between classes, $\chi^2(3, N = 66) = 5.17$, $p = .16$, or between instructional groups, $\chi^2(3, N = 66) = 1.32$, $p = .72$. The second question asked whether they were good at math. Again, no significant differences were found between classes, $\chi^2(3, N = 66) = 4.14$, $p = .13$, or between instructional groups, $\chi^2(3, N = 66) = 5.25$, $p = .07$.

Instrumentation

Immediately before and after instruction, students were tested on three measures: a fractions computation test, a word problem test, and a contextualized problem test. A transfer test was administered 10 days after the posttests. The validity of the four measures was supported through logical analysis (Pedhazur & Schmelkin, 1991) as evidenced by previous successful hypothesis testing (see Bottge & Hasselbring, 1993). The items on the computation test and the word problem test were similar to those commonly found in intermediate and middle school textbooks. On each item, students were awarded one point for showing correct procedures and one point for the correct answer, in keeping with theory that math problems can be partitioned into procedural and conceptual knowledge (Goldman, Hasselbring, and the Cognition and Technology Group at Vanderbilt, 1997).

Fractions Computation Test. The 18-item fractions computation test measured students' ability to add and subtract fractions. The items were graduated in difficulty from simple addition of fractions with like denominators to subtraction of mixed numbers with unlike denominators. Students could earn 2 points per item, or a total of 36 points, by showing correct procedures and displaying the correct answer in simplified form. One point was awarded if students used correct procedures for solving a problem but arrived at an incorrect answer. This was often the case when students did not express their answers in simplified form. Cronbach's coefficient alpha for pretest scores of the study sample was .94. Interrater reliability on 25% of the protocols from a randomly selected sample of pretests and posttests was 98% (range = 78%–100%). It was calculated by dividing the number of agreements by the total number of agreements and disagreements and multiplying by 100 to yield percentage of agreement (Sulzer-Azaroff & Mayer, 1977).

Word Problem Test. An 18-item word problem test measured student performance in solving single-step and multistep word problems about linear measurement. The problems required students to add and subtract fractions in ways that paralleled the computation problems on the computation test. The text was written at or below the third-grade reading level. As with the computation test, students could obtain 36 possible points: one point for the correct procedure and 1 point for the correct answer. Cronbach's coefficient alpha for pretest scores of the study sample was .90. Interrater reliability on 25% of the protocols from a randomly selected sample of pretests and posttests was 99% (range = 91%–100%). Interrater reliability was calculated in the manner described previously.

Contextualized Problem Test. A 36-point contextualized problem test was based on an 8-minute video anchor, Bart's Pet Project, written for a previous study. This problem required a solution based on the calculation of several subproblems involving buying a small pet and building a home for it. To solve the problem, students had to add and subtract money, add and subtract fractions, and convert simple measurement equivalents, such as inches to feet. (These actions correspond to NCTM Standards #5 Numbers and Number Relationships, #7 Computations and Estimations, #10 Statistics, and #13 Measurement.)

After viewing the video, students showed their work in an open 6" × 6" work area on the problem response form. Ten procedures for solving the problem and the criteria for scoring them were developed in a previous study by asking ninth-grade advanced math students to try out the measure (Bottge & Hasselbring, 1993). Students could earn full or partial credit on each subproblem, depending on how clearly and accurately they represented a reasonable solution. For example, students had to show an economical way of cutting three pieces of wood for the pet cage. They could not afford to build the cage unless they cut the lengths of wood in combinations that wasted as little wood as possible. Students earned four points for representing each combination without making computation errors. They earned two points for the correct combination if they made minor computation errors. They did not earn any points for combinations that would not lead to a legitimate outcome. There were three possible ways to solve the problem.

Interrater agreement on 25% of the protocols from a randomly selected sample of pretests and posttests was 98% (range = 78%–100%). It was calculated as described previously under Fractions Computation Test.

Kite Transfer Task. The transfer task required students to show how they could afford to build a kite frame with a limited amount of money and materials. Students were furnished three pieces of information: the kite plan, a materials list required for building the kite, and the money and wood students had in their possession. By using the wood in the most economical way, they could build the kite frame. The overall problem required students to solve subproblems involving money, linear measurement, and building plans.
To solve the kite tasks, students used procedures similar to those used to solve the word problem and Bart’s Pet Project. The kite problem was similar to the word problem because it described the problem situation in text rather than in video format. It was like the contextualized problem because students figured out the problem from a drawing and a materials list. Calculations for the kite task were similar to those required for the word problem and the contextualized problem because students had to add and subtract fractions and money to arrive at the correct solution. There were two possible solutions to the problem.

Eight procedures for arriving at the solution and the criteria for scoring the protocols were identified and field tested in the same manner as for the contextualized problem test. Interobserver agreement on 25% of the protocols from a randomly selected sample of pretests and posttests was 94% (range = 50%–100%). It was calculated as described previously under Contextualized Problem Test.

Skateboard Ramp Problem. The skateboard problem consisted of a schematic drawing for constructing a skateboard ramp and a problem-solving worksheet. From several skateboard designs obtained from a local skateboarding store and through Web sites, the technology education teacher drafted a skateboard ramp and made a materials list to go with it.

The problem-solving worksheet indicated that one 10-foot-long 2 × 4 and six 8-foot-long 2 × 4s were needed for the ramp. In the materials section, the lengths of wood were represented by rectangular bars. From dimensions on the schematic drawing, students were to indicate on the bars how they would cut up the lengths of wood in the most advantageous way. There were two possible solutions. The ramp problem was field tested by the teachers in the study and by a research assistant to ensure that the directions were clear and the solutions were correct.

The technology education teacher and the math teacher presented the ramp problem in the technology education classroom. They told students that they could begin to build their ramp after they found a way to cut the wood in the most efficient way. They did not help students in any way after the initial directions were given. The groups were compared on the length of time they took to solve the problem. Incomplete or incorrect solutions were not accepted. The technology education teacher judged whether the solutions were acceptable or needed further explanation. The math teacher and the author monitored the students’ progress. The author recorded the amount of time students took to satisfactorily finish the problem.

Research Design and Hypotheses

One remedial math class and two prealgebra classes were involved in this research. Two groups of students from the remedial math class and the two prealgebra classes composed the four comparison groups. A pretest–posttest experimental design was used with the remedial math groups and a quasi-experimental, nonequivalent control group pretest–posttest design with the prealgebra classes to test the following hypotheses:

1. Contextualized video instruction improves student performance on computation problems, word problems, contextualized problems, and transfer tasks.

2. Word problem instruction improves student performance on computation problems, word problems, contextualized problems, and transfer tasks.

In addition, qualitative contextual information was gathered to provide additional explanation in support of hypothesis testing (see Cook & Campbell, 1979). Procedures and timelines are summarized in Table 2.

Experimental Study

On four successive school days prior to intervention, students in the remedial math class and both prealgebra classes took the computation, word problem, and contextualized problem tests, and completed the questionnaire. Following 10 school days of instruction, students again took the computation, word problem and contextualized problem posttests. Ten school days after the last day of intervention, students in all four groups attempted the kite frame task.

Contextualized Instruction (Experimental Group Treatment). Two video-based, contextualized math problems, The 8th Caller and Bart’s Pet Project, were used in the experimental groups. The 8th Caller was shown first to familiarize students with the nature of the contextualized problems and to acquaint them with videodisc technology. Bart’s Pet Project was the primary instructional video.

The design of the videos was based on the principles established by the Cognition and Technology Group at Vanderbilt University (1997) in making the Adventures of Jasper Woodbury. The videos employ an “embedded data design,” in which the problems are not explicitly stated or well formulated as they usually are in standard word problems. To arrive at a plausible solution, students must first identify pertinent information, formulate hypotheses about how this information can be used, and test their hypotheses. The problems are presented in a videodisc format to allow students quick access to the relevant information.

The 8th Caller was written by three sixth-grade students in Minnesota after they had solved several Jasper problems in their math class. The video was filmed on location in the teacher’s house, a party supply store, a grocery store, and a park. The authors of the problem were also the main actors in the production. The video’s challenge asks students if they can afford to buy pizza for their party from a pizza shop or if they
must settle for frozen pizza from a grocery store. The challenge usually takes middle school students about one class hour to solve.

Primary instruction was based on a second video called Bart’s Pet Project, which was written and then reviewed by faculty and staff at Vanderbilt University prior to filming it. The first scene opens with Bart reading a book on his bed and occasionally listening to the television. As the announcer on the television gives a list of sale prices on construction lumber and landscape timbers, he turns up the volume and listens more intently. Int the next scene, he is shown looking at the money that he has spread out on the bed. His friend Billy arrives and they discuss Bart’s plan to buy a pet. Bart is unsure whether he has enough money to buy both a pet and a cage to put it in. In the next scene, Bart and Billy are in a pet store looking at the prices of several pets and cages. As Bart is about to leave the store, he notices a brochure that shows how to build a pet cage. The last scene takes place in Bart’s garage, where he and his friend are measuring several lengths of wood for the pet cage. The challenge is to construct a cage out of 2 × 2 wood with the least amount of waste so there is enough money left over to buy a pet. When the problem was first conceived, it had two plausible solutions. A third solution was found during a previous study (Bottge & Hasselbring, 1993) by a student with a long history of learning and behavior disabilities. All three solutions require students to compare and add whole numbers and fractions.

Before beginning instruction, both teachers attended a series of 2-hour meetings to learn the content of the video-based problems, to learn how to operate the videodisc equipment, and to review the daily lesson plans provided by the investigator. Two methods guided development of the daily lesson plans, which were partially scripted and included step-by-step instructions. The methodology of Rosenshine and Stevens (1986) was used as a general guide: review (check previous day’s work and reteach, if necessary); presentation of new content/skill; guided student practices (and check if necessary); feedback and correctives (and reteach, if necessary); independent student practice; and weekly and monthly reviews. The teachers also introduced and practiced a modified version of Montague’s cognitive strategy training model (see Montague, 1997). The model was condensed from seven to five steps because the original model was intended for text-based problems. The model included the following proce-
dures: paraphrase the problem, hypothesize, estimate, compute, check.

On the first day of instruction, the teacher explained to students that they were going to learn how to solve important problems. Then, she assigned them to groups of two or three. The teacher introduced students to the five-step problem-solving strategy and modeled how it could be used to solve a problem. Next, the teacher played *The 8th Caller* one time with no interruption and asked students to describe the challenge problem presented by the video. After students explained the problem situation, each group was given a problem-solving folder that contained several sheets of paper with large open areas organized into the following headings: information, frame number (videodisc), calculations. Following Noddings’ (1985) suggestion, students were encouraged to provide a written and an oral account of how they went about solving the problem. The teacher explained how to operate the video controller and access individual frame numbers. During the last 5 minutes of class, the students discussed possible ways to solve the problem and practiced using the video controller.

On Days 2 and 3, students reviewed the five steps for solving problems. Then they spent half of the class period discussing how to solve the video problem in their groups. When the teacher was sure that each group had reached a plausible solution, she asked representatives to explain the group’s solution to the class. During this early stage of instruction, the teacher assisted students in clarifying and paraphrasing their findings.

On the following 6 days, students reviewed the five-step strategy frequently and worked on solving *Bart’s Pet Project*. The teacher told students they would earn more credit if they included a detailed account of how they arrived at their solution. Groups shared the videodisc controller, searched the video, discussed ideas, and recorded their procedures on a recording form. The teacher answered questions to help clarify obvious misconceptions about the problem but did not provide specific ways to solve it. When the teacher noticed that students had reached a reasonable solution, she asked groups to describe their findings. Then they were encouraged to find the other two solutions.

On the last day of instruction, when students had successfully solved the problem in at least two ways, the teacher asked a series of “what if” questions. For example, students were asked, What if Ben had not spent $3 at the arcade? or, What if the cage plan had called for sides that were 3/4 inch longer?

**Word Problem Instruction (Control Group Treatment).** A series of standard single- and multistep word problems that paralleled the content of *Bart’s Pet Project* was written by the investigator. The problems looked like typical word problems found in many basal mathematics textbooks. To solve them, students needed to add and subtract money, calculate linear measurements involving fractions, and read tables of materials. The problems were copied to overhead transparencies.

Daily teaching plans guided the 10 days of instruction. These plans were based on the procedures of Rosenshine and Stevens (1986) and by the seven steps of Montague’s (1997) cognitive strategy training model: read, paraphrase, visualize, hypothesize, estimate, compute, and check. Many of the word problems were identical to the ones used in a previous study (Bottge & Hasselbring, 1993), but more complex problems were added because the intervention spanned more days. The word problems were similar in content and required many of the same procedures as the contextualized problems, although none of them made reference to pets, parties, or cages.

One of the problems involved planting rows of carrots, beans, and radishes. Each seed package indicated how long a row could be planted. The price of each package was also provided. Students were asked how much money it would cost to plant the garden if each row was 8 feet long. Another problem asked students to figure out how to cut lengths of 2 × 4 lumber in the most efficient way to build a raft. Most prealgebra students took more than 10 minutes to solve the raft problem.

On the first day of instruction, the teacher explained that word problems are common ways to test what students know about mathematics. Then she described the seven-step strategy for solving word problems, led students through several memorization activities, and then applied the strategy to simple word problems. When the teacher was reasonably sure that students had memorized the steps, she asked individuals to demonstrate each step.

On the second day, the teacher handed out file folders in which multiple sheets of problem-solving forms were stapled. A double line framed a large rectangular work area where students were to show all of their work. Each day, the teacher practiced the problem-solving strategy with the students the same way: group recitation, individual recitation, individual work on problems. At the end of the period, the teacher gave students up to five word problems to solve independently.

Classes on the next 8 days followed the same sequence of procedures. The teacher encouraged students to volunteer ways to solve each problem and to discuss their merits. When she was sure that students knew how to solve a problem, she modified it by asking “What if” questions. For example, after solving the raft problem, she asked: What if Jan only had enough money for one 2 × 4 length of lumber? How would that fact change the size of the raft?

On two occasions, students made up their own word problems in small groups. After the groups created their problems, the teacher selected students from each group to present their problems to the whole group. When students were sure of their answers, they took turns presenting their answers to the class.

**Qualitative Inquiry (Skateboard Ramp)**

Findings of experimental studies are enhanced and further validated when important discoveries within the context of the
study are described (see Cook & Campbell, 1979). To that end, 12 school days after the last posttest was administered, the five students from the remedial math CP group and five students from the remedial math WP group who obtained the highest summed scores on the computation, word problem, and contextualized problem tests were chosen for the extended transfer task. The purpose of the task was to find out whether the students could use their skills to plan and build a skateboard ramp.

There were two parts to the ramp task. First, students had to decide what lengths to cut the $2 \times 4$s in order to make the best use of the available wood. They used a planning form to show their work. Then they built the ramp according to their plan and the schematic drawing. Students used power tools such as skill saws and hand saws, sanders, and drills. They also used hand tools such as tape measures, chalklines, T and framing squares, and a hand saw. The students completed the projects in six class periods. On the last day, several students brought rollerblades to school and tested the ramps.

Thus, two groups of students built two skateboard ramps in the industrial technology room. In final form, the ramps measured 5 feet in length, 4 feet in width, and 2 feet in height. The ramp surface was $\frac{3}{4}$-inch treated plywood.

**Fidelity of Treatment Implementation**

Several methods helped to ensure fidelity of treatment implementation during the 10 days of instruction. First, the experimenter provided the teachers with daily lesson plans for each instructional group. Parts of directions were scripted to help ensure that directions were explained in the same way to each group. Twice each day, before and after instruction, we reviewed details of the plan to clarify classroom procedures for the following day. Second, the author observed 26 out of 40 (65%) total class periods and checked off each procedure as it was completed. All of the procedures on the lesson plans were followed. Third, student work folders gave a clear account of what students did in class each day. On days when a class was not observed, the author selected student folders at random and matched the student work to the procedures in the lesson plans. The combination of lesson plans and student work provided a clear account of what happened in class when the observer was not present. Finally, six class sessions were videotaped for a permanent record of class activities.

**Data Analysis**

**Experimental Study.** Prior to analyzing the research hypotheses of the experimental study, a preanalysis of statistical power estimation was conducted to determine the alpha level for hypothesis testing with this sample size. Effect sizes were converted to a common metric, $r$, as recommended by Rosenthal (1991) by a generic formula that computes $r$ for $F$ values and degrees of freedom:

$$r = \sqrt{\frac{F(1, \ldots)}{F(1, \ldots) + df_{error}}}$$

According to Cohen (1977), using the $r$ metric, .10 is a small effect, .30 is a medium effect, and .50 is a large effect. In my previous research, which involved similar treatment but shorter durations, effect sizes on the contextualized posttest and transfer task were .50 and .37, respectively. Using the sample power procedures suggested by Lipsey (1990) and considering available sample size of the remedial math class ($n = 17$), an alpha level of .10 was chosen to detect a moderate effect size of (.40) with adequate statistical power.

Three separate two-way analyses of covariance (ANCOVA) were conducted on the computation, word problem, and contextualized posttests, with each pretest serving as the covariate. One factor was the type of Instruction (WP or CP), and the other factor was the Class (remedial or prealgebra). The obtained means were adjusted statistically to take into account the initial differences among the pretest means and represent the best possible estimate as to what the groups would have scored on the posttest if both groups had started with identical means. A two-way analysis of variance (ANOVA) (Instruction $\times$ Class) was used to test the difference in means of the groups on the kite transfer task.

**Qualitative Inquiry.** The industrial technology teacher, the math teacher, the special needs teacher, and the author observed students every day as they planned and then measured, cut, and assembled the skateboards. We periodically asked students to show us fractions on the tape measure such as sixteenths, eighths, quarters, and halves. We also asked students several procedural questions, such as, "How do you know that you are cutting this piece to the correct length?" Five of the class periods were videotaped, and the two ramps remain at the school as permanent products of the students' work.

After each class, the teachers and the author discussed their observations and perceptions. The author summarized the meeting notes and, after viewing the videotapes, drafted tentative conclusions. The teachers were asked to confirm or suggest modifications of the draft, which were incorporated into the final summary.

**Results**

**Experimental Findings**

The obtained and adjusted means and standard deviations of the four experimental groups are reported in Table 3. Note that the highest possible score on the computation, word problem, and the contextualized problem tests was 36. A perfect score on the kite frame transfer test was 30.
### TABLE 3. Obtained and Adjusted Means and Standard Deviations for WP and CP Groups

<table>
<thead>
<tr>
<th>Test data</th>
<th>Remedial group</th>
<th>Prealgebra group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WP</td>
<td>CP</td>
</tr>
<tr>
<td>Number of students</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Computation test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>27.11</td>
<td>24.37</td>
</tr>
<tr>
<td>$SD$</td>
<td>7.6</td>
<td>11.7</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtained $M$</td>
<td>26.78</td>
<td>21.00</td>
</tr>
<tr>
<td>$SD$</td>
<td>9.1</td>
<td>14.6</td>
</tr>
<tr>
<td>Adjusted $M$</td>
<td>26.34</td>
<td>22.65</td>
</tr>
<tr>
<td>Word problem test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>17.56</td>
<td>18.63</td>
</tr>
<tr>
<td>$SD$</td>
<td>9.3</td>
<td>9.0</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtained $M$</td>
<td>21.44</td>
<td>23.87</td>
</tr>
<tr>
<td>$SD$</td>
<td>8.0</td>
<td>9.7</td>
</tr>
<tr>
<td>Adjusted $M$</td>
<td>23.81</td>
<td>25.63</td>
</tr>
<tr>
<td>Contextualized test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>2.78</td>
<td>6.62</td>
</tr>
<tr>
<td>$SD$</td>
<td>2.3</td>
<td>3.9</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtained $M$</td>
<td>2.8</td>
<td>19.0</td>
</tr>
<tr>
<td>$SD$</td>
<td>2.0</td>
<td>13.3</td>
</tr>
<tr>
<td>Adjusted $M$</td>
<td>3.38</td>
<td>18.30</td>
</tr>
</tbody>
</table>

Note. CP = Contextualized Problem. WP = Word Problem.

Results of the ANCOVA were mixed across the three posttests. On the computation posttest, there was a significant main effect for class, $F(1, 61) = 4.86, p = .03, r = .31$, but not for instruction, $F(1, 61) = 1.50, p = .23, r = .15$, or for class by instruction interaction, $F(1, 61) = 0.90, p = .35, r = .12$. On the word problem posttest, there was no main effect for class, $F(1, 61) = 0.14, p = .71, r = .05$, for instruction, $F(1,61) = 0.05, p = .82, r = .03$, or for class by instruction, $F(1, 61) = 1.05, p = .31, r = .13$. On the contextualized problem posttest, there was a significant main effect for instruction, $F(1, 61) = 27.79, p < .001, r = .56$, but not for class, $F(1, 61) = 0.11, p = .74, r = .04$, or for class by instruction, $F(1, 61) = 1.21, p = .28, r = .14$.

On the kite transfer task, the ANOVA revealed a main effect for class, $F(1, 58) = 6.99, p = .01, r = .33$, and for instruction, $F(1, 58) = 8.99, p < .001, r = .37$, but not for class by instruction, $F(1, 58) = 0.049, p = .82, r = .03$. The means and standard deviations of the remedial and prealggebra CP groups were 11.75 ($SD = 3.4$) and 19.52 ($SD = 10.6$) compared to the remedial and prealggebra WP groups, 4.22 ($SD = 2.7$) and 10.79 ($SD = 9.9$), respectively.

### Qualitative Findings

On the first day, confusion over how to plan construction of the ramp resulted in heated discussions in both groups. Students in the WP group hastily added several combinations of wood and compared the total amount to the available amount of lumber. They repeated this procedure for 30 minutes until one student figured out the correct way to solve the problem and explained it to the other students in her group. Their final solution was accepted by the technology education teacher after they had worked on it for 35 minutes. They began construction of the ramp the next day.

Students in the CP group did not want to work together, which hampered their progress. In contrast to the cooperative atmosphere in the WP group, students in the CP group argued for the first half of the class period and then quit working on the problem. The next day they began to work on the problem in earnest when they noticed that WP students had begun to construct their ramp in the workshop. In 5 minutes, the CP group solved the problem to the teacher’s satisfaction.
The following informal conclusions can be drawn from observing how students used what they had learned during instruction to plan and build two skateboard ramps. First, students in the CP groups quickly solved the ramp problem once they realized that formulating a workable plan would enable them to build the project. Because the planning form looked like schoolwork, several students refused to work on it until after they realized how the plan fit into the total project. Second, motivating characteristics of the construction quickly eliminated most of the student conflicts. In fact, once students began construction, they worked well together and only a few students needed disciplining. Finally, students were proud of their accomplishments. The quality of the finished products attested to the care they took in constructing them.

**Discussion**

The purpose of this study was to test the hypothesis that instruction with contextualized problems (CP) improves student performance on computation problems, word problems, contextualized problems, and on transfer tasks compared to instruction with word problems (WP). The results of this study support the practice of situating problems in a meaningful context for improving the math problem-solving skills of low-and average-achieving students. Statistically significant differences were found on the contextualized problem test and on the transfer task for CP students in both the remedial and prealgebra classes. Differences in computation and on word problems were not significant.

The highest-achieving students in the CP and WP groups were able to use what they learned to plan and build two skateboard ramps. It took the WP group one class period to solve the ramp problem. The CP group did not solve it during the first class period because they could not work together and quarreled much of the time. They solved the problem on the second day within 5 minutes when they saw WP students building the ramp. They likely would have solved the problem quickly the first day if they had been able to put their personal differences aside.

**Limitations**

There are several limitations of the study. The most obvious limitation of this study was its length. Although it spanned more than 2 months, from pretesting to the end of the ramp construction, there was not time to assess exactly how the contextualized problems helped students plan and build the ramp or whether the construction project contributed to more positive classroom experiences. Nonsignificant findings of the groups on computation and word problem posttests also may have been due to the short instruction period.

Second, questions related to readiness remain unanswered. Pretest scores indicated that several students in the Math 8 class had weak fractions computation skills. Posttest scores showed no improvement for either WP or CP instruction, yet most of the students seemed to have had sufficient skills to build the skateboard ramps. Should we enable students to become involved in complicated projects before they know how to compute fluently? Do some students learn computation by working on significant problems? Perhaps the answer is yes for some students and no for others.

Third, there were preexisting differences in math achievement and attitude about math between the comparison groups. For example, there was a 9-point mean difference on Wisconsin state tests in favor of the CP prealgebra group prior to intervention. Although only approaching significance, the disparity raises questions about the equivalence of the prealgebra comparison groups. The remedial math CP and WP groups also differed in self-efficacy. More than 75% of the WP students did not think they were good at math, whereas 75% of the CP students thought they were good at math.

Finally, results of this study cannot be generalized easily beyond this sample. It is important to note that the process of doing an experiment often increases internal validity while decreasing external validity (Cook & Campbell, 1979). Although the experimental design of this study allowed a direct connection between intervention and outcome, external validity was compromised by a restricted student sample exposed to a treatment of relatively short duration, by the small numbers of teachers involved, and by measures that appeared to sample the content reliably but for which no psychometric information was available.

**Relationship to Previous Research**

Despite these limitations, findings of this study seem to support efforts to incorporate meaningful and challenging problems into math instruction for low-achieving students and are consistent with previous research (Bottge & Hasselbring, 1993). Most of the students in the CP groups were better at noticing critical features of problems than were students who learned with standard types of word problems. Both groups of CP students used this skill to score significantly better than comparison WP groups on the contextualized posttest and the transfer task.

The practical importance of these results is demonstrated by the magnitude and consistency of effect sizes in this study and the earlier study. For example, effect sizes were .56 for CP students in this study compared to .50 for CP students in the previous study. On the transfer task, effect sizes for CP students were .37 in both studies. According to Cohen (1977), an effect size of .30 is a medium effect and .50 is a large effect. The slightly larger effect sizes obtained in this study may be attributed to the longer intervention period. These findings suggest that the treatment resulted in significant and important differences in student performance.

Several students in this study were able to transfer what they learned in the instructional setting to two transfer tasks.
This is encouraging in light of consistent findings that show that transfer of skills is exceedingly difficult to achieve among students with learning difficulties (see Cooper & Sweller, 1987; Meltzer, 1994; Stone & Michaels, 1986). The ability of students in the CP groups to transfer skills may be explained, in part, by those who urge educators to help students recognize common structures of problems, as opposed to rushing them into quick conclusions (Bruner, 1960; Resnick & Ford, 1981). Hasty conclusions may inhibit students’ ability to recognize common patterns in problems, thereby limiting their ability to remember and transfer information. In recognizing the structure of a problem, students should be able to identify analogous schemata in related problems, thus reducing working memory load and freeing the cognitive resources for further planning (Cooper & Sweller, 1987).

Other factors such as group dynamics and motivation may also help explain these results. For example, the second part of the study asked students in the remedial class to figure out how to construct a skateboard ramp from a set of schematic plans and a materials sheet. Both CP and WP groups solved the problem in about one class period. The five students in the WP group began to work on the problem almost immediately. Although the discussions were heated, the WP group found a solution within 30 minutes on the first day. Students in the CP group argued and gave up on the first day. However, the arguing quickly subsided on the second day as they watched students in the other group beginning to build their ramp. Less than 5 minutes after the class began, CP students generated two workable plans.

The relationship of clarity of purpose to student motivation has been a popular topic in preservice and inservice teacher education programs, especially in promoting authentic instruction and assessment. But the relationship, although easy to understand on the surface, is still elusive in this controlled study. Even when we “think” we are teaching and assessing in authentic ways, we may not be. It takes a skillful teacher to embed the problem, the solution, and the cognition involved in easily recognizable and motivating contexts (Brown, Collins, & Duguid, 1989). These connections must appear obvious and unmistakable, especially with students who have a history of learning and behavior problems.

The influence of motivating, contextualized tasks on learning was spotlighted by a girl in the remedial group who had a long history of math underachievement. She was the only student in the study, including prealgebra students, who obtained a perfect score on the contextualized posttest and the kite transfer task. She also led the CP group to a swift solution of the ramp problem the second day. Justifiably proud of her achievement but also a little embarrassed by it, she exclaimed, “Don’t tell my parents about this. They will faint.” When asked how she showed such ability to solve difficult math problems, she explained that school math did not make sense to her and served no useful purpose. For her, the contextualized and applied problems did not look like math, which she regarded as a set of exercises unrelated to events in her life. The surprising achievements of this girl replicated a similar finding in the previous research.

The present study highlights important strengths that many students, especially those who experience difficulty learning mathematics, bring to the classroom. Unfortunately, instruction for these students usually targets identified weaknesses rather than strengths, a situation that may contribute to a sense of hopelessness and concomitant behavior problems. This is a recurring theme in education and continues despite emphasis on standards, authentic learning, and performance-based assessment. To make an impact on middle school and high school students’ lives, we must move beyond the typical into contexts that promote and motivate student thinking. The power of technology matched to suitable learning tasks can lead to positive results for many disenchanted students. It is time to use our resources to make learning for all students more meaningful and attainable.

AUTHOR’S NOTE

I thank teachers and administrators who participated in the study, especially Robin Crow, Amy Lobeck, Rich Hagens, Anne Thompson, and Karen Mathis. I am also grateful to Elizabeth Watson for her assistance in data analysis and to Edna Szymanski, who reviewed drafts of the manuscript.

REFERENCES


