The Use of a Cognitive Tutoring System in the Improvement of the Abstract Reasoning Component of Word Problem Solving

Jenifer L. Wheeler

Mei Technology

J. Wesley Regian

Air Force Research Laboratory

Requests for reprints should be addressed to Jenifer Wheeler, Mei Technology, 8930 Fourwinds Drive, Suite 450, San Antonio, TX 78239. Phone: (210) 655-8911 E-mail: jenifer@meitx.com

Abstract

This study evaluated the Word Problem Solving (WPS) Tutor's ability to improve the abstract reasoning component of word problem solving. 632 ninth grade students were assigned to 3 groups: a control group exposed to traditional group instruction, a placebo group exposed to traditional group instruction plus a computerized word problem solving environment with no active tutoring, and an experimental group exposed to traditional group instruction and a computerized word problem solving environment with active tutoring. A paper and pencil test was administered to students at the beginning of the school year, at mid-term, and at the end of the school year and was used to compare the performance of the three groups on the concrete and abstract components of word problem solving. Results indicated that WPS-Tutored students improved significantly more than either the control group or the placebo group on both the abstract and concrete reasoning subtests. However, performance gains on the abstract subtest were not as large as the gains on the concrete subtest. Implications for mathematics education and suggestions for future research and development efforts are discussed.

Introduction

Research has shown that, compared to students in other countries, American students' performance in math is deficient (Sharma & Travaglini, 1988). In particular, American students tend to have difficulties with algebra (Wenger, 1987). However, mastering algebra is a critical step needed for learning other college-preparatory mathematics (Dean & Williams, 1993) and for working in technology-oriented fields. Therefore, it is important that educators and researchers determine what makes algebra difficult and adapt instruction accordingly.

Learning algebra in the context of word problems is especially difficult (Schoenfeld, 1987), perhaps because the central skill needed for solving algebra word problems is abstract reasoning. Algebra word problems integrate abstract thinking skills and procedural skills by providing concrete representations of numbers and then requiring students to use abstract representations (variables and equations), algebra, and arithmetic to solve them (Kallam & Kallam, 1996). Some related research has suggested that difficulties in algebra are caused by students' use and understanding of literal symbols (Booth, 1984; Kuchemann, 1981). Perhaps if educators focus more on teaching the abstract component of algebra word problem solving, students' overall performance in this area would improve.

As part of its effort to develop advanced instructional technologies, the Air Force Research Laboratory has developed a cognitive tutoring system (CTS) called the Word Problem Solving (WPS) Tutor. The WPS Tutor is designed to teach algebra word problem solving at the ninth grade level. An evaluation conducted by Regian and Wheeler (in press) indicated that students who used the WPS Tutor improved on word problem solving performance by 29%, whereas students who experienced only traditional classroom instruction improved by only 19%. In this paper we examine the WPS Tutor's effectiveness more closely by evaluating its ability to improve students' performance on abstract reasoning tasks.

Cognitive Tutoring Systems

Cognitive tutoring systems are computer programs that emulate human tutors by providing individualized instruction to students. In addition, CTSs use principled instruction based on empirically supported learning theories to teach skills and subject matter. Research has shown that private tutoring (by humans) can increase learning by two standard deviations beyond group methods of instruction (Bloom, 1984). Using CTSs in conjunction with group teaching methods may be a way to provide individualized instruction for students while maintaining the benefits of the classroom.

Although CTSs are not as effective as human tutoring, several studies have found that they are more effective than traditional classroom instruction. In a review of three studies conducted to determine the effectiveness of three CTSs in the secondary school, higher education, and military education domains, Shute and Regian (1990) found an average reduction in learning time of 55% when the CTSs were used compared to when traditional instruction was used. In another review of CTS studies, Shute and Psotka (1996) found an average effect size of 1.00 standard deviation beyond the learning gains produced from traditional instruction.

Word Problem Solving Tutor

The WPS Tutor is a CTS that teaches students to analyze and solve word problems in five steps: 1) identify the goal of the problem, 2) identify the values necessary to solve the problem, 3) make an equation, 4) solve the equation, and 5) answer the question using the appropriate solution value and units of measure. The Tutor does not teach mathematical calculation but is designed to be used as a supplement to traditional classroom instruction that includes mathematical calculation.

The WPS Tutor consists of 24 independent modules covering pre-algebra, algebra, and geometry. Each module begins with a tutorial that uses animation, graphics, and text to review the appropriate subject material. In addition to this review, the tutorial contains a worked example of a sample problem, questions to test students' knowledge of the material, and a summary page. After successful completion of the tutorial, students are presented with a set of word problems organized by level of difficulty. The problems in the first level of difficulty are isomorphic to the example problem in the tutorial. This equation is also the simplest version of the equation for that module. Each subsequent level is only slightly more difficult than the previous level, and the source of additional difficulty is explicitly taught in the Tutor. Diagnostic algorithms in the Tutor determine the type of feedback students receive. In addition, based on their knowledge of students' skills and abilities, teachers can set the number of problems students are required to solve per module, the number of levels of difficulty they are required to achieve per module, and the number of times they are allowed to ask for help per problem.

The WPS Tutor is pedagogically based on five cognitive theoretical foundations, including learning by practice (Anderson & Fincham, 1994; Blessing & Anderson, 1996; Lovett & Anderson, 1994), elaboration (Reigeluth, 1987, 1992), categorization (Rosch, 1988a, 1988b), mastery (Kulik, Kulik, & Bangert-Drowns, 1990a, 1990b; Slavin, 1990), and induction (Reed & Bolstad, 1991; Reed, Willis, & Guarino, 1994; Ward & Sweller, 1990). These theories are well supported by empirical data and are described briefly in Table 1.

Insert Table 1 Here

The current study evaluated the WPS Tutor's ability to improve the abstract reasoning component of word problem solving. A placebo tutor allowed students to solve the same set of computerized word problems and receive computer help statements while doing so. However, the theoretical foundations described above were not integrated into the placebo tutor. In other words, problem solving was not preceded by a tutorial with a sample problem, difficulty differences were developed by subjective teacher ratings, and problems were presented at random, rather than by level of difficulty. All students were also exposed to traditional lecture-based instruction, which covered the same material presented by the WPS Tutor and the placebo tutor. Teachers conducted their lecture-based classes using their own methods of instruction.

Method

Participants

Participants were 632 students from seven high schools in Texas, New Mexico, and Ohio during the 1992-93 school year. Fifty percent were male. The median age of the students was 15 years. All students were enrolled in elementary algebra classes. The class sizes ranged from 20 to 30 students.

Test

A paper-and-pencil test was used to measure each student's ability to solve algebra word problems. The test items encompassed the mathematics curriculum common to all of the classes. During the first semester, the test consisted of eight word problems. Each word problem had five multiple-choice questions associated with it. The first question asked students to find the goal statement of the problem, the second question asked students to identify relevant information, the third question asked students to identify the correct equation needed to solve the problem, the fourth question asked students to identify the correct numerical answer, and the fifth question asked students to identify the correct units for the answer. Each question contained three distractors and one correct answer. Because several modules were added to the WPS Tutor during the second semester, six word problems were added to the test. Therefore, the test during the second semester consisted of 70 questions.

The test was also designed to compare abstract and concrete reasoning skills. The first three questions for each word problem had two levels of abstraction (concrete and abstract). For the first question, students were asked to identify a paraphrase of the original goal statement (concrete level) or a conversion of the goal statement into one that included the major mathematical operation (abstract level). The second question asked students to differentiate between relevant and irrelevant information. The options either included the actual numbers with their description (concrete level) or the description only (abstract level). On the third question, students were asked to select a correct equation for solving the problem. Options for these questions were equations with the variables numerically represented (concrete level) or equations with variables symbolically represented (abstract level). The last two question types (solve the equation and provide the correct units) did not vary on the abstractness dimension. Table 2 provides descriptions of the levels of abstraction for each question type. The Appendix contains an example of each question type.

Insert Table 2 Here

Equipment

Students used 80386-based computers running Windows 3.1. The software was developed in Asymmetrix Toolbook version 2.0.

Procedure

Eight classes (84 students) were randomly assigned to the control group. They were taught the ninth grade mathematics curriculum using traditional classroom instruction only. Twelve classes (139 students) were randomly assigned to the placebo group. They received traditional classroom instruction, but one session per week (about 50 minutes) was replaced with the placebo tutor. Thirty-two classes (409 students) were randomly assigned to the treatment group. They received traditional classroom instruction, but one session per week (about 50 minutes) was replaced with the WPS Tutor. All students did the same homework assignments. Classes within the three treatment groups were distributed across locations.

Results

Test scores were converted to percentages by dividing each raw score by the 95th percentile score (21 on the concrete subtest and 17 on the abstract subtest) of each posttest. Teachers graded their students on a curve and considered the 95th percentile score a cutoff point that could be reasonably achieved. On the concrete pretest, treatment group students (those who would use the WPS Tutor) scored an average of 33%, placebo group students scored 26%, and control group students scored 29%. Although not large, these differences were significant [$F(2,629) = 11.39$, $p < .001$]. On the abstract pretest, treatment group students scored 40%, placebo group students scored 33%, and control group students scored 39%. These differences were also significant [$F(2,629) = 9.53$, $p < .001$].

On the concrete posttest, treatment group students scored an average of 65%, placebo group students scored 45%, and control group students scored 51%. These differences were significant [$F(2,629) = 42.58$, $p < .001$]. Finally, on the abstract posttest, treatment group students scored 60%, placebo group students scored 49%, and control group students scored 50%. Again, these differences were significant [$F(2,629) = 16.86$, $p < .001$]. The mean percent change scores for each group by subtest are shown in Table 3.

Insert Table 3 Here

On the concrete subtest, a test of multiple comparisons indicated that the mean change score of the WPS-Tutored group (31%) was significantly different ($p = .05$) than the change scores of both the placebo group (19%) and the control group (22%). There was no difference between the placebo group and the control group. On the abstract subtest, the mean change score of the WPS-Tutored group (20%) was significantly different ($p = .05$) than the change scores of the control group (11%). Neither of the other differences was significant.

Because the scores on the pretest were significantly different, path analysis, a method that examines the direct and indirect effects between variables, was used as a method of analysis of covariance in analyzing the effects of the Tutor on concrete and abstract posttest scores. Path analysis is especially useful in conducting analyses of covariance because it allows the covariates to have measurement error, something that is almost always unavoidable. It is also more flexible and comprehensive in isolating commonly accepted causal variables than analysis of variance (Hoyle, 1995).

The model tested proposed that abstract and concrete scores on the pretest were measures of a latent variable, pre-treatment math skill. In addition, abstract and concrete scores on the posttest were measures of the latent variable, post-treatment math skill. Post-treatment math skill was suggested to be effected by pre-treatment math skill and treatment type. Pre-treatment math skill was used as a covariate. Because there was no difference between the placebo group and the control group on either subtest, only the control group and the treatment group were compared in the path analysis.

Maximum likelihood was used as a method of estimating free parameters. The chi-square statistic was used as the fit index, and the Tucker-Lewis coefficient (TLI) and comparative fit index (CFI) were used as incremental indices. The path model and the results of the analysis, including standardized regression weights and squared multiple correlations are shown in Figure 1.

Insert Figure1 Here

Results of the path analysis [x^2 (3, N = 632) = 5.61, p = .13, TLI = 1.00, CFI = 1.00] indicated that the model was a good fit of the data. All regression coefficients were significant ($p < .01$). As expected, pre-treatment math skill had a large effect on post-treatment math skill. However, the WPS Tutor significantly added (b = .16) to post-treatment math skill. In addition, post-treatment math skill (the skill acquired after receiving treatment) significantly added to scores on both the abstract (b = .69) and the concrete (b = .86) subtests; however, students improved less on the abstract subtest.

To further evaluate the correctness of the path model in Figure 1, an alternative model was tested. The second model was exactly like the first except that a constraint was added so that post-treatment math skill did not depend on treatment. In other words, the regression weight of treatment on post-treatment math skill was fixed at 0. A test of this model indicated that it was not a good fit [x^2 (4, N = 632) = 15.03, p = .005]. Therefore, it was concluded that the WPS Tutor had a significant effect on post-treatment math skill.

Discussion

The WPS Tutor significantly improved ninth graders' performance on both the abstract and concrete word problem solving subtests. However, the gain in performance on the abstract subtest was much less than the gain on the concrete subtest. The path analysis indicated that pre-treatment math skill contributed almost equally to the abstract pretest (b = .59) and the concrete pretest (b = .57). However, the skill acquired after being exposed to the WPS Tutor (post-treatment math skill) contributed much more to the concrete posttest (b = .86) than it did to the abstract posttest (b = .69).

Clark (1983) has argued that the instructional design of a system, rather than the medium, affects learning. In this study, we did not control for the design of the lecture-based instruction. However, the WPS Tutor was designed to supplement traditional instruction by acting as a substitute for private human tutoring. The Tutor's positive effect on learning in this study was most likely due to its capability to provide individualized instruction. The benefit of the computer medium is that it can provide automated, individualized instruction to students, allowing teachers to focus on other classroom issues. Although the placebo tutor also provided automated individualized instruction, it did not integrate an active pedagogy for teaching word problem solving. Therefore, we used the placebo tutor to help us rule out the possibility of novelty effects.

As industries become more technology-oriented, employers will require workers who are competent in math and science. Educators and researchers must begin developing new instructional strategies for teaching these subjects so that students are prepared for the workforce. The results found in this study indicated that students experienced more difficulties with the abstract component of algebra word problem solving than they did with the concrete component. This difficulty may contribute to the national low scores in algebra word problem solving. Research has, in fact, shown that abstract reasoning is more difficult than concrete reasoning, particularly for 13-16 year olds (Markovits & Vachon, 1990), the age group evaluated in this study. Perhaps if mathematics curricula were designed to focus more on abstracting information from word problems, American students could improve their math skills and become more employable.

An effort to redesign the instruction of algebra word problem solving has already begun. The Air Force Research Laboratory is developing a new CTS called Mathematical Abstract Reasoning Tutor (MARTHA), in which students are taught to generate a concrete expression for solving a word problem and then convert this expression to an abstract expression. Eventually, MARTHA students are required to directly generate the abstract expression without first generating a concrete expression. This tutor is still in the development stages and has not yet been evaluated.

Abstract reasoning in algebra word problem solving is just one element of mathematics that may be causing difficulties for students. Other research has shown that parent, teacher, and student attitudes toward math and perceptions of successful performance in math are quite different in other countries (Stevenson, Lee, Chen, Lummis, Stigler, Fan, & Ge, 1990). Perhaps, in addition to focusing on abstraction, the

American culture needs to raise expectations of performance and educate people on the importance of math for future success in college and careers.

References

Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 1322-1340.

Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 576-598.

Bloom, B. S. (1984). The 2 Sigma Problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 41, 4-17.

Booth, L. (1984). Misconceptions leading to error in elementary algebra. Journal of Structural Learning, 8, 125-138.

Clark, R. E. (1983). Reconsidering research on learning from media. Review of Educational Research, 53, 445-459.

Dean, J. W., & Williams, H. G. (1993). Project PRIME--Meeting great expectations. NASSP, 76, 1-8.

Hoyle, R. H. (1995). Structural equation modeling. Thousand Oaks, CA: Sage Publications.

Kallam, L. G., & Kallam, M. (1996). An investigation into a problem solving strategy for indefinite integration and its effect on test scores of general calculus students (Report No. SE057954). Hays, KS: Fort Hays State University. (ERIC Document Reproduction Service No. ED393674).

Kuchemann, D. (1981). Algebra. In K. Hart (Ed.) Children's understanding of mathematics. (pp. 79-93). London: John Murray.

Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990a). Effectiveness of mastery learning programs: A meta-analysis. Review of Educational Research, 60, 265-299.

Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990b). Is there better evidence on mastery learning? A response to Slavin. Review of Educational Research, 60, 303-307.

Lovett, M. C., & Anderson, J. R. (1994). Effects of solving related proofs on memory and transfer in geometry problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 366-378.

Markovits, H., & Vachon, R. (1990). Conditional reasoning, representation, and level of abstraction. Developmental Psychology, 26(6), 942-951.

Reed, S. K., & Bolstad, C. A. (1991). Use of examples and procedures in problem solving. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 753-766.

Reed, S. K., Willis, D., & Guarino, J. (1994). Selecting examples for solving word problems. Journal of Educational Psychology, 86, 380-388.

Regian, J. W., & Wheeler, J. L. (in press). The Effects of a Cognitive Tutoring System on Ninth Graders' Performance in Word Problem Solving (Air Force Technical Report). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Reigeluth, C. M. (1987). Lesson blueprints based on the Elaboration Theory of Instruction. In C. M. Reigeluth (Ed.), Instructional theories in action (pp. 245-288). Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M. (1992). Elaborating the elaboration theory. Educational Technology, Research & Development, 40, 80-86.

Rosch, E. (1988a). Coherences and categorization: A historical view. In F. S. Kessel (Ed.), The development of language and language researchers: Essays in honor of Roger Brown (pp. 373-392). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E. (1988b). Principles of categorization. In A. Collins & E. E. Smith (Eds.), Readings in cognitive science (pp. 312-322). San Mateo, CA: M. Kaufmann Publishers.

Schoenfeld, A. H. (1987). Cognitive science and mathematics education. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sharma, M. C., & Travaglini, L. E. (1988). Math notebook (From theory to practice): Information for teachers/parents of children with learning problems in mathematics (Report No. SE050734). Framingham, MA: Center for Teaching/Learning Mathematics. (ERIC Document Reproduction Service No. ED342600).

Shute, V. J., & Psotka, J. (1996). Intelligent tutoring systems: Past, present, and future. In D. Jonassen (Ed.), Handbook of research on educational communications and technology. New York: Macmillan Publishing Company.

Shute, V. J., & Regian, J. W. (1990, June). Rose garden promises of intelligent tutoring systems: Blossom or thorn? Paper presented at the Space Operations, Applications and Research Symposium, Albuquerque, NM.

Slavin, R. E. (1990). Mastery learning re-reconsidered. Review of Educational Research, 60, 300-302.

Stevenson, H. W., Lee, S. L., Chen, C., Lummis, M., Stigler, J., Fan, L., & Ge, F. (1990). Mathematics achievement of children in China and the United States. Child Development, 61, 1053-1066.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. Cognition and Instruction, 7, 1-39.

Wenger, R. H. (1987). Cognitive science and algebra learning. In A. H. Schoenfeld (Ed.) Cognitive science and mathematics education. (pp. 26-52). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Table 1

WPS Tutor theoretical foundations

Learning Theory	Description of Theory	Application in WPS
Practice	Learning and skill acquisition occur when students are taught the required declarative knowledge and then are required to actively process the information or practice the skill.	Students learn declarative knowledge in the subject (e.g., area of shapes) and pass a test on that knowledge before they are allowed to practice solving problems in that domain.
Elaboration	Students begin by practicing the epitome of a skill, and then practice progressively and systematically more complex versions of the skill.	Students are first presented with the simplest possible version of the relevant formula. As students progress in the Tutor, steps are added to the formula, making it more difficult and complex.
Categorization	An appropriate way to organize the subject information is specified and then explicitly taught to students. Information is organized around prototype concepts, which become the center for conceptual categories.	Students are presented with a prototype formula in each module. As students practice their skills, they build expertise and (ideally) recognize that new problems require solution strategies that are minor variations on known strategies.
Mastery	When a complex skill is to be taught based on important prerequisite skills, the curriculum must be ordered appropriately. In addition, an appropriate level of mastery for each skill is insured before advancement is allowed.	Students are allowed to progress to a new level or module only when they have mastered the current level or module. For this experiment, mastery was achieved when students successfully completed two problems in a level without asking for help.
Induction	Students are shown how to solve examples of the kinds of problems they will encounter. Student can solve new isomorphic problems by mapping steps from the worked example onto the new problem.	Each module of the WPS Tutor presents worked examples of word problems. Then students are given isomorphic problems to solve.

Table 2

Descriptions of concrete and abstract answer formats for each test question type

Question Type	Concrete Answer Format	Abstract Answer Format
Identify the goal	paraphrase of the goal	conversion of the goal with major mathematical operation
Identify the values	actual numbers with descriptions	descriptions of values only
Make an equation	numerically represented variables	symbolically represented variables

Table 3

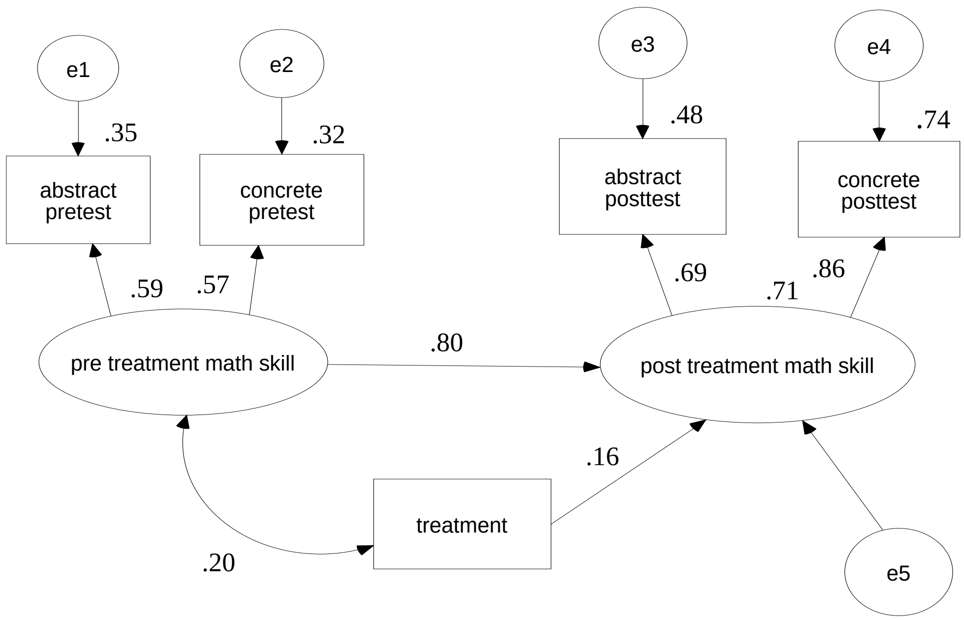
Mean change scores by treatment group

	Abstraction Level	
Treatment Group	Concrete	Abstract
WPS Tutor Group	31%	20%
Placebo Group	19%	15%
Control Group	22%	11%

Figure Caption

Figure 1. Path model comparing abstract and concrete scores, including standardized regression weights and squared multiple correlations.

Chi-square = 5.61
df = 3
p = .13



Appendix

Word Problem Examples

Word problem questions with abstract answers

A fund-raising group hopes to sell all the Christmas trees grown on their lot in the mountains. This plot of land yields three fir trees per every 15 square feet (the "sub-area"). If the length of their rectangular lot is 80 feet and the width is 120 feet, how many fir trees were grown on the lot?
* indicates the correct answer

1. What is the problem asking you to do?
    - a. Multiply the area of the lot by the sub-area, then divide by the number of trees in the sub-area
    - b. Divide the area of the lot by the number of trees in a sub-area, then, subtract the result from the lot area.
    - *c. Multiply the number of trees in a sub-area by the number of sub-areas in the lot
    - d. Divide the lot area by the number of trees in a sub-area, then divide by the number of sub-areas

2. To solve this problem, what information is <u>not</u> important to know?
    - *a. Number of trees they hope to sell
    - b. Number of trees per sub-area
    - c. Length of the lot
    - d. Width of the lot

3. Which of the following is a correct equation to solve this problem?
    - a. (L x W) - ((L x W) / S)
    - *b. ((L x W) / S) x T
    - c. ((L x W) x S) / T
    - d. ((L x W) / T) / ((L x W) / S)

4. What is the correct answer?
    - a. 480
    - b. 500
    - c. 1280
    - *d. 1920

5. What are the correct units for the answer?
    - a. Rectangles
    - *b. Trees
    - c. Feet
    - d. Square Feet

Word problem questions with concrete answers

A 30-foot long brick walkway lies next to the house. The plan is to lay out a small lawn in the shape of a half circle, with the straight side of the lawn alongside the walkway. A brick wall will edge the far side of the lawn. To decide how many pounds of grass seed is needed, what is your estimate of the area to be planted?
* indicates the correct answer

1. What is the problem asking you to do?
    a. Find how many pounds of grass seed is needed
    b. Find the size of the walkway next to the lawn
    *c. Find the area to be planted in grass seed
    d. Find the length of the edge around the new lawn

2. To solve this problem, what information is <u>not</u> important to know?
    *a. A brick wall will surround part of the lawn.
    b. The entire area is to be planted.
    c. The distance across the straight side is 30 feet.
    d. The area is one half a circle.

3. Which of the following is a correct equation to solve this problem?
    a. 30 x π x 2
    b. 30 x 30 / π x 2
    *c. 30 / 2 x 30 / 2 x π / 2
    d. 30 x 30 x π / 2

4. What is the correct answer?
    a. 573
    *b. 353
    c. 287
    d. 188

5. What are the correct units for the answer?
    a. Bricks
    b. Pounds
    c. Feet
    *d. Square feet