Sources of Difficulty in Multi-Step Geometry Area Problems

Yvonne S. Kao (ykao@andrew.cmu.edu)
Department of Psychology, 5000 Forbes Ave.
Pittsburgh, PA 15213

Ido Roll (idoroll@cmu.edu)
Human-Computer Interaction Institute, 5000 Forbes Ave.
Pittsburgh, PA 15213

Kenneth R. Koedinger (koedinger@cmu.edu)
Human-Computer Interaction Institute, 5000 Forbes Ave.
Pittsburgh, PA 15213

Abstract

Although U.S. students often perform well on basic, single-step math problems, they often struggle with extended, multi-step free-response problems. This study examines the sources of difficulty in multi-step geometry area problems. We found that the presence of distracters creates significant difficulty for students solving geometry area problems, but that practice on composite area problems improves students' ability to ignore distracters. In addition, this study found some support for the hypothesis that the increased figural analysis requirements of a complex diagram can negatively impact performance and could be a source of a reverse composition effect in geometry. More practice with challenging single-step problems (e.g., by incorporating a more complex figure) may improve students' performance on multi-step problems as well.

Keywords: education; geometry; mathematics; problem-solving; spatial processing

Introduction

Although much work has been done to improve students' math achievement in the United States, geometry achievement appears to be stagnant. While the 2003 TIMSS found significant gains in U.S. eighth-graders' algebra performance between 1999 and 2003, it did not find a significant improvement on geometry items during the same time frame (Gonzales et al., 2004). Furthermore, of the five mathematics content areas assessed by TIMSS, geometry was the weakest for U.S. eighth-graders (Mullis, Martin, Gonzales, & Chrostowski, 2004). While students have often demonstrated reasonable skill in "basic, one-step problems," (Wilson & Blank, 1999, p. 41) they often struggle with extended, multi-step problems in which they have to construct a free response. In this paper, we will explore two possible reasons for poorer performance on multi-step geometry area problems: composition effects and poor operator knowledge, and their implications for instruction.

Multi-step problems should be more difficult than single-step problems—in essence, there are more opportunities to make a mistake. However, researchers have found that multi-step problems can be more difficult than the sum of their single-step components. In Heffernan and Koedinger's (1997) study of multi-step algebra story problems, they found the probability of answering a multi-step problem (e.g., 800-40x) correctly to be less than the product of the probabilities of answering each of the single-step components (e.g., 800-y *and* 40x) correctly. They called this the "composition effect." They suggested the composition effect occurred because students lacked a key skill unique to multi-step problems—composing algebraic subexpressions (e.g., producing the expression that results from substituting 40x for y in 800-y). Thus a task analysis based simply on the surface forms, which does not include this knowledge component, underestimates the difficulty of the multi-step problems.

Our study uses multi-step geometry area problems like those shown in Figure 1b. In these problems, students must find the areas of an outer figure and an inner figure, and then subtract the two areas to find the area of a shaded region—a sort of "shape algebra." It is reasonable to think that we would find a composition effect in this task. The literature suggests two skills that could be the source of a composition effect in this task: subgoaling and figural analysis.

Subgoaling

A problem-solver must do a certain amount of planning and subgoaling to find an efficient solution to a multi-step problem. According to some researchers (e.g., Sweller, 1988), this subgoaling presents a significant cognitive load and detracts from the problem-solving performance.

Figural Analysis

Lean and Clements stated in their 1981 review of spatial ability and mathematics, "perceptual analysis and synthesis of mathematical information presented implicitly in a diagram often make greater demands on a pupil than any other aspect of the problem (p. 277)." Bishop (1983) has identified the ability to interpret figural information as one of the two key spatial skills that support mathematics achievement. In addition, Koedinger and Anderson (1990) found that a hallmark of geometry expertise was the ability to parse a complex diagram into perceptual chunks that could be used to guide a search of problem-solving schemas. Geometry novices most likely are not able to parse complex diagrams into meaningful perceptual chunks quickly or efficiently and thus the increased diagram complexity often seen with multi-step problems could result in increased problem difficulty.

We would also like to suggest an interesting possibility. If figural analysis is the source of a composition effect in our geometry area problems, then it is possible we will see a *reverse* composition effect. Students pay the cognitive cost of figural analysis once per problem, whether the problem is single- or multi-step. Thus, the product of the probabilities of success on each of the single-step problems might actually be greater than the probability of success on a multi-step problem; the former calculation might over-estimate the load of performing a figural analysis.

An alternate hypothesis is that we will see no composition effect in our task. This would support what we call the Operator Sufficiency Hypothesis: that knowledge of the basic operators is both necessary and sufficient for success on a multi-step task—there are no skills unique to multi-step problems. According to the Operator Sufficiency Hypothesis, difficulty on multi-step problems is entirely due to insufficient understanding of basic operators. In our task, these basic operators include how to find the area of a triangle and how to subtract a smaller area from a larger area. Multi-step problems may inherently contain distracters that expose students' shallow understanding of basic operators.

It is worth mentioning that these two hypotheses—the Composition Effect Hypothesis and the Operator Sufficiency Hypothesis—are not necessarily in opposition. For example, in Heffernan and Koedinger's (1997) algebra study, if a deeper cognitive task analysis identifies subexpression composition as a key operator and instruction is designed to address this operator (e.g., give students substitution exercises), then learning the identified operators may be sufficient. A composition effect may still exist with respect to a surface or behavioral analysis may—but not with respect to a deeper cognitive analysis. More generally, if researchers or educators have performed a complete and accurate analysis of the basic operators, then knowledge of the basic operators should indeed be sufficient for success on multi-step problems and no composition effect should be found.

To summarize our primary research questions:

1. Why are multi-step geometry problems so difficult for students? Is it because multi-step problems require additional skills or because students haven't mastered the basic operators?
2. What types of instruction will best improve students' performance on multi-step geometry problems?

In order to answer our questions about multi-step problems, we needed to first assess the difficulty of single-step area problems. Koedinger and Cross (2000) found that the presence of distracter numbers on parallelogram problems—the length of a side was given in addition to the lengths of the height and base—significantly increased the difficulty of the problems due to students' shallow application of area knowledge. In particular, students seemed to have over-generalized a procedure for finding the area of rectangles—multiplying the lengths of adjacent sides—to parallelograms. In addition, Koedinger and Cross conjectured that non-standard orientations for shapes—non-horizontal bases and non-vertical heights—would also expose students' shallow knowledge. Given that a multi-step area problem necessarily contains one or more numbers necessary only for a particular sub-problem and often features shapes that are rotated from their standard orientations, it will be important for us to follow Koedinger and Cross's lead and get a baseline measure of how distracters and orientation affect performance on single-step area problems. Then we will study how combining single-step area problems into the typical, "find the shaded area" composite area problem described above affects performance.

Method

Participants

We collected data from 98 10th- and 11th-grade high school students in six geometry classes at a public rural vocational high school. All classes had the same instructor. We performed analyses only on the data from the 66 students who completed both the mid-test and post-test.

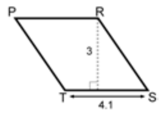
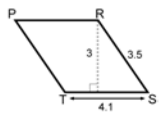
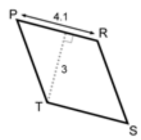
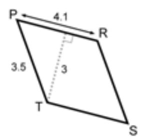
Design

This study contains two components. The first component is a within-subjects 2 (distracters absent vs. present) × 2 (standard orientation vs. rotated) difficulty factors analysis of typical simple-diagram area problems involving parallelograms, regular pentagons, triangles, and trapezoids. Distracters took the form of additional, unnecessary segment lengths marked on the problem diagram. For example, a side of a triangle may be labeled in addition to the base and height. Standard-orientation problems contained shapes with horizontal bases and vertical heights and, in the cases of pentagons, trapezoids, and triangles, the apex of the shape pointing towards the top of the page; Rotated problems were rotated from the standard orientation by an arbitrary number of degrees that varied between diagrams. In these Simple problems, we asked students to calculate the area of the single shape that appeared on the diagram. Example Simple problems are presented in Figure 1a.

The second component is a within-subjects analysis of the skills required to solve a multi-step area problem. In these problems, students were given a complex diagram that consisted of a large parallelogram, triangle, trapezoid, or regular pentagon and a small interior rectangle with the area between the large shape and the rectangle shaded. Students were instructed to either complete the entire multi-step problem and find the area of the shaded region or to perform only one of the three required sub-problems: find the area of the outer shape, find the area of the inner rectangle, or find the area of the shaded region when both the areas of the large shape and the interior rectangle were given—

essentially, subtract two given numbers. Example Complex problems are presented in Figure 1b.

a)

Condition	Instructions	Diagram
Basic	All Simple conditions received the same instructions: PRST is a parallelogram. Find the area of PRST	
Distracters		
Rotated		
Distracters + Rotated (DR)		

b)

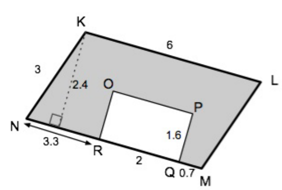
Condition	Instructions	Diagram
Outer	KLMN is a parallelogram. Find the area of KLMN.	All Complex conditions received the same diagram: 
Inner	OPQR is a rectangle. Find the area of OPQR.	
Subtract	The area of parallelogram KLMN is 14.4; The area of rectangle OPQR is 3.2 Find the area of the shaded region.	
Multi-Step	KLMN is a parallelogram OPQR is a rectangle. Find the area of the shaded region	

Figure 1: Example area problems. a) Simple problems, b) Complex problems.

The study reported in this paper was part of a large-scale research effort to evaluate learning and performance in the Area unit of the Geometry Cognitive Tutor curriculum. In addition to the manipulations discussed below, the study also contained a small, worked-example instructional intervention and a test of a new algorithm to monitor students' mastery of different skills in the tutor. Neither of these additional components affected results, so we will not discuss them further.

Materials

Test items were generated and counterbalanced across paper-and-pencil test forms. There were two forms of the pre-test, which contained eight Simple problems—two problems from each Simple condition—and no Complex problems. There were four different forms of the mid-test and the post-test, both of which contained four Simple problems and four Complex problems—one problem from each condition.

The students in this study used the Geometry Cognitive Tutor curriculum, a highly effective commercial intelligent tutoring system (Koedinger, Anderson, Hadley, & Mark, 1997), in which students engaged in self-paced problem-solving two days a week on a computer. This study spanned the first seven sections of the curriculum. In the first six sections, students practiced calculating the perimeter and area of different shapes in simple, single-step problems. Each of these six sections focused on a single class of shapes: squares and rectangles, parallelograms, triangles, trapezoids, regular polygons, and circles. In the seventh section, Area Composition, students learned and practiced how to calculate the area of a third shape by adding or subtracting the areas of two other shapes.

Procedure

At the beginning of the semester and prior to any instruction, students took the pretest during their regular class time. Students then progressed through the tutor curriculum at their own pace. Students practiced the single-step problems in the first six sections of the curriculum for 4-7 weeks, depending on their individual pace. The mid-test was administered in an as-you-go fashion following completion of these sections of the curriculum. Students then completed the seventh section and took the post-test immediately upon completion of this section. All tests were administered by the classroom instructor; test forms were randomly distributed to students. Students were allotted 30 minutes for all tests.

Scoring

Responses were initially given a score of 1 for correct and 0 for incorrect. If a student did not read the directions carefully for a Complex problem and completed the entire problem when only a sub-problem, e.g., Subtract, was required, the student was given credit for the response if the sub-problem was computed correctly.

Results and Conclusions

An alpha value of .05 was used for all statistical tests.

Scores on the pre-test were at floor, ranging from 0 to 50% correct (M = 14.94%, SD = 13.61%). Pre-test scores did not correlate significantly with either mid-test scores or post-test scores. Thus we did not analyze the pre-test further.

Table 1: Mean performance on mid-test and post-test by diagram type[1]

Diagram	Mid-test (%)		Post-test (%)		Gain (%)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Simple	68.6	30.8	79.2	23.0	10.6**	27.8
Complex	61.7	30.1	71.2	31.1	9.5*	33.3
Simple-Complex	6.8*	25.0	8.0*	27.4		

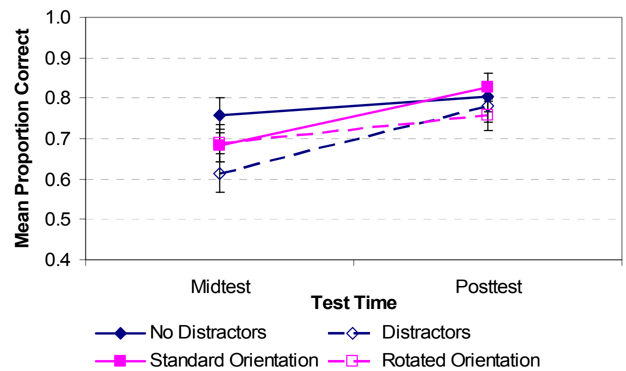
Mid-test and Post-test Performance Analysis

Performance on Simple problems was significantly better than performance on Complex problems at both mid-test, $t(65) = 2.214$, $p = .030$, 2-tailed, and at post-test, $t(65) = 2.355$, $p = .022$. These data are presented in Table 1. A repeated-measures ANOVA on Simple problems found a main effect of test time on accuracy, $F(1, 65) = 9.637$, $p = .003$, with students performing significantly better on post-test (M = 79.17%, SD = 23.03%) than on mid-test (M = 68.56%, SD = 30.79%). In addition, we found a main effect of distracters on accuracy, $F(1, 65) = 5.630$, $p = .021$, with students performing significantly better on no-distracter problems (M = 78.03%, SD = 26.13%) than on distracter problems (M = 69.70%, SD = 28.59%). There was no main effect of orientation. We did find a marginal distracters × test time interaction, $F(1, 65) = 2.861$, $p = .096$, but no other interactions approached significance. These results are presented in Figure 2a.

A repeated-measures ANOVA on Complex problems also found a main effect of test time on accuracy, $F(1, 65) = 5.328$, $p = .024$, with students performing significantly better on post-test (M = 71.21%, SD = 31.08%) than on mid-test (M = 61.74%, SD = 30.13%). We also found a main effect of sub-problem condition, $F(1, 65) = 12.890$, $p < .001$. Pairwise comparisons found that Subtract and Inner Shape problems did not differ significantly from each other and that Outer Shape and Multi-Step problems did not differ significantly from each other. All other comparisons were reliable at $p < .005$, with students performing significantly better on the Subtract and Inner sub-problems than on the Outer and Multi-Step problems. These results are presented in Figure 2b.

We conclude the following from these preliminary analyses: Complex problems are more difficult than Simple problems. Area Composition instruction significantly improved performance on both Simple and Complex problems. For Simple problems, much of this improvement seems to result from students' increased ability to cope with distracters—while distracters presented a significant difficulty at mid-test, they only presented a slight difficulty at post-test. Presumably this improvement in coping with distracters would improve performance on Complex problems as well. We did not find any effects of orientation; however we did find evidence that some students simply rotated the test paper until they were viewing the figure in a standard orientation, effectively negating our manipulation. The presence of this behavior suggests that orientation does present a significant difficulty for some students, and that a study with the appropriate controls might reveal it. For Complex problems, success on Multi-Step problems seemed to be largely predicted by success on the Outer shape sub-problem, perhaps because the Inner Shape and Subtract calculations were relatively easy. A forward step-wise binary logistic regression on the Complex problem data supports this conclusion. We used the sub-problem conditions as predictors (i.e., whether the problem required an Outer calculation, an Inner calculation, and/or a Subtraction) of Complex problem accuracy. At both mid-test and post-test, only Outer Required was predictive of success, $B = -.815$ and $p = .002$ at mid-test, and $B = -1.329$ and $p < .001$ at post-test.

a)



b)

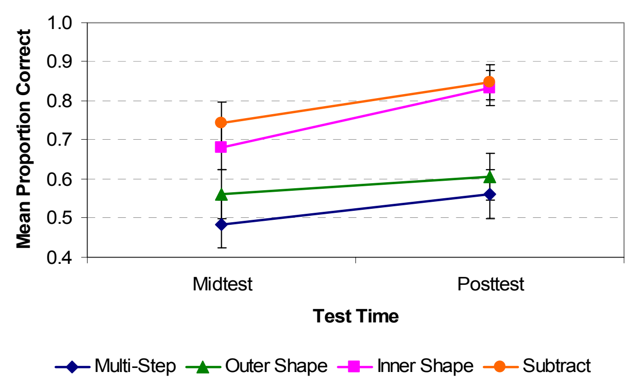


Figure 2: Mid- to post-test gains by condition. a) Simple problems, b) Complex problems. Error bars represent standard error.

[1] *p < .05, **p < .01.

Composition Effect Analysis

In order to determine whether there is a composition effect in this task, we tried to predict accuracy on Multi-Step problems using accuracy on the sub-problems. If there is no composition effect, then Accuracy(Multi-Step) = Accuracy(Outer) × Accuracy(Inner) × Accuracy(Subtract) for each student. However, the predicted accuracies differed significantly from the actual data at mid-test, $t(65) = 2.193$, $p = .032$, 2-tailed. Students performed significantly better on the Multi-Step problems than was predicted by the model above—a reverse composition effect. The difference between the model and the data was not significant at post-test. These results are presented in Table 2.

Table 2: Comparisons of mathematical models of Multi-Step performance to actual data[2]

	Mid-test (%)		Post-test (%)	
Problem Type	*M*	*SD*	*M*	*SD*
	Simple			
Distracters + Rotated	62.1	48.9	71.2	45.6
	Complex			
Outer	56.1	50.0	60.6	49.2
Inner	68.2	46.9	83.3	37.6
Subtract	74.2	44.1	84.9	36.1
Multi-Step	48.5	50.4	56.1	50.0
	Models of Multi-Step Performance			
Outer × Inner × Subtract	33.3*	47.5	53.0	50.3
D+R × Inner × Subtract	40.9	49.5	60.6	49.2

As suggested in the introduction, the presence of the reverse composition effect at mid-test may be due to an overestimation of the difficulty imposed by the complex figural analysis in the Multi-Step problems. While Multi-Step problems require the student to perform complex figural analysis only once, performing the three sub-problems separately requires the student to do it twice—for the Outer and the Inner sub-problems. In order to correct for this overestimate, we substituted the accuracy on the Simple Distracters + Rotation problem for Accuracy(Outer) above. These two problem types are mathematically equivalent, and the figural analysis requirements for the Simple D+R problem are substantially lower, although not zero. The predictions of this new model, Accuracy(Multi-Step) = Accuracy(D+R) × Accuracy(Inner) × Accuracy(Subtract), did not differ significantly from the actual Multi-Step performance data at either mid-test or post-test. This model is also presented in Table 2. However, it is worth mentioning that performance on Simple D+R problems and Outer problems did not differ significantly on either mid-test or post-test, even with the difference in figural analysis requirements. Thus, our conjecture that figural analysis is a key skill that is particularly important for Multi-Step problem success is only partially supported.

[2] *Model is significantly different from data at $p < .05$.

Discussion

We will return to our original research questions to begin the discussion.

1. Why are multi-step geometry problems so difficult for students? Is it because multi-step problems require additional skills or because students haven't mastered the basic operators?

It seems that both were true for our students. We did find a reverse composition effect in geometry area problems—the probability of success on multi-step problems could not be predicted by simply multiplying the probabilities of success for the associated sub-problems. We found some evidence that figural analysis presents significant difficulty for students, that this difficulty is greatest on multi-step problems with complex diagrams, and that this may be the source of the reverse composition effect we found. However, a figural analysis explanation does not tell the whole story and other skills (e.g., subgoaling) may have an effect as well. In addition, we found that specific practice on multi-step area problems improved performance on one-step problems as well as multi-step problems, suggesting that students may not have fully mastered the basic skills before beginning the Area Composition unit, but that practice with Area Composition provides continued practice with the basic skills as well as instruction on multi-step problems.

2. What types of instruction will best improve students' performance on multi-step geometry problems?

The greatest improvement after the Area Composition unit seemed to be in dealing with distracters. This makes a great deal of intuitive sense if you consider composite area problems to inherently contain distracters. Thus, instruction on how to recognize and extract important problem information from the figure while ignoring distracting information may improve students' performance on multi-step geometry problems.

As a final thought, it seems that the challenge of multi-step problems is not the act of composing single-step problems, but that the nature of multi-step problems renders the single steps more difficult. This is in line with the Operator Sufficiency Hypothesis. Mathematics curriculums should be sure to include challenging single-step problems—problems with distracters, in the case of geometry area, or problems that require manipulation of a subexpression, in the case of algebra—in order to help students develop a more advanced understanding of the basic operators and better prepare students for problems that require the composition of several skills. It may be the case that more emphasis on the challenging aspects of single skills may be more efficient at preparing students for multi-step problems than practicing specific instances of multi-step problems.

Acknowledgments

This work was supported in part by a Graduate Training Grant awarded to Carnegie Mellon University by the Department of Education (#R305B040063) and by the Pittsburgh Science of Learning Center, which is funded by the National Science Foundation (award #SBE-0354420). We would like to thank Ms. Colleen Conko and Hao Cen for their help in carrying out this study.

References

Bishop, A. J. (1983). Space and geometry. In R. Lesh & M. Landau (Eds.), *Acquisition of Mathematics Concepts and Processes*. New York: Academic Press.

Gonzales, P., Guzman, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). *Highlights From the Trends in International Mathematics and Science Study: TIMSS 2003.* Washington, DC: National Center for Education Statistics.

Heffernan, N. T. & Koedinger, K. R. (1997). The composition effect in symbolizing: The role of symbol production vs. text comprehension. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science, 14*, 511-550.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30-43.

Koedinger, K. R., & Cross, K. (2000). Making informed decisions in educational technology design: Toward meta-cognitive support in a cognitive tutor for geometry. In *Proceedings of the Annual Meeting of the American Educational Research Association (AERA)*, New Orleans, LA.

Lean, G. & Clements, M. A. (1981). Spatial ability, visual imagery, and mathematical performance. *Educational Studies in Mathematics, 12*, 267-299.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Wilson, L. D. & Blank, R. K. (1999). *Improving mathematics education using results from NAEP and TIMSS.* Washington, DC: Council of Chief State School Officers, State Education Assessment Center.