This is a pre-print of a paper subsequently published in Learning and Instruction in 1997. This paper does not correspond exactly to the one published.

There is More Than One Way to Solve a Problem: Evaluating a Learning Environment that Supports the Development of Children's Multiplication Skills

Shaaron Ainsworth, David Wood, and Claire O'Malley

ESRC Centre for Research in Development, Instruction and Training

University Of Nottingham.

Abstract

Interpretation of the nature of mathematical understanding has changed recently. These changes have prompted calls for different instructional methods in the primary classroom. COPPERS is a mathematical learning environment which explores how such goals should be implemented computationally. Two experiments have examined how system components have advanced children's understanding that multiplication problems can have many different correct solutions. Different numbers of decompositions, learner's choice of strategy and feedback in either tabular or place value representations were all found to significantly affect learning. Theoretical interpretations of these results are considered in terms of Vygotskian approaches to scaffolding learning and current research on external representations.

Introduction

The notion of what constitutes mathematics has changed many times this century. Recently, focus on the learning of formal procedures and accepted facts has been replaced with an emphasis on the mathematician as a flexible, insightful problem solver. Schoenfeld (1992) describes mathematics as a 'science of patterns' where the goal is to systematically study and explain the nature and principles of regularities in pure and applied systems. This approach, demands some understanding that mathematics involves experimentation, hypothesis testing and active seeking of solutions and contrasts strongly with children's beliefs about the nature of mathematics. For example, Baroody (1987) asserts that due to an overemphasis on 'the right answer', children commonly believe that all problems must have a correct answer, that there is only one correct way to solve a problem and that inexact answers (such as estimates) or procedures are undesirable. This may appear surprising given that pre-school children appear to naturally use multiple approaches to answer simple arithmetic problems (Fuson, 1992) and given their impressive pre-school competencies (*e.g.* Nunes & Bryant, 1996). These researchers conclude that the school and cultural environment shapes not simply the strategies children use but their overall beliefs about the nature of mathematics.

Cross cultural comparisons between teaching methods in Western classrooms and those in Russia and Asia have helped identify alternative approaches to mathematics instruction which might ameliorate such problems. For example, Fuson (1992) proposes that mathematics learning should involve (a) situations that are meaningful and interesting to children, (b) sustained engagement in mathematical situations, not on quickly finding answers, (c) alternative solutions and (d) analysis and acceptance of errors.

COPPERS is a prototype computer based learning environment that attempts to embody these ideas by placing emphasis on searching for many answers to a single problem. Many of the educational principles underlying its design were based on a system for teaching multiplication in the classroom described by Lampert (1986a,b). Two objectives of her scheme have been implemented in COPPERS. The first goal is to develop the understanding that there can be multiple routes to the solution of mathematical problems. COPPERS serves this objective by providing questions for which there is one right total, but requires this total to be decomposed in a number of different ways.

The second goal is the importance of allowing children's concrete and everyday knowledge to support the learning of other types of understanding such as principled and computational knowledge. However, in order to profit from children’s informal knowledge, it is not sufficient to simply present concrete manipulables (*e.g.* Schoenfeld, 1986). The domain chosen by Lampert and adapted for use in COPPERS is that of coin problems. It is commonly found that young children benefit from presenting problems in which numbers refer to meaningful situations (*e.g.* Hughes, 1986), even if these situations are imaginary. Money is familiar and meaningful to children. Furthermore, there is evidence that pre-literate children and unschooled adults can apply relevant concepts such as additive composition and place value when dealing with coin problems (Nunes & Bryant, 1996). Finally, computation is involved in dealing with money in everyday life. This should encourage children to frame the problem as a mathematical one and hence facilitate understanding (Kaput & Maxwell-West, 1994).

COPPERS is proposed as an environment to explore how to design computer systems that support instructional approaches such as Fuson’s. Children must consider many solutions to a single problem and no one answer is considered ‘better’ than another. A three coin solution is no more or less acceptable than a laboriously constructed 25 coin answer. Errors are recorded alongside right answers and, as discussed above, we propose that coin problems are relevant and meaningful to school children. These ideas form the basis for the design of COPPERS, but we also need to determine which system components facilitate this goal. Thus, we have conducted two experiments to examine aspects of the system which support understanding of multiple solutions to multiplication problems.

System Description

The domain taught by COPPERS involves arithmetic problems such as ‘What is 3 × 20p + 4 × 10p ?’ Users must answer this question by providing alternative decompositions of this total. One way to answer the problem is to calculate the total to this sum (*i.e.* (3 × 20p = 60p) + (4 × 10p = 40p) = £1.00) and then provide multiple decompositions to this total (*e.g.* £1.00 = ‘20p + 20p + 10p + 50p’, or ‘10p + 2p + 2p + 1p + 5p + 10p + 10p + 5p + 5p + 50p’, *etc*.). An alternative is to decompose the sub-totals (*e.g.* 3 × 20p = ‘20p + 10p + 5p + 5p + 10p’, 4 × 10p = ‘5p + 5p + 5p + 5p + 20p’). To successfully solve these problems, children must demonstrate a number of skills and have certain conceptual knowledge. They must (at least) know the meaning of the symbols ‘x’ and ‘+’, and how to perform the operations they denote, they must know that the order in which the operations are performed is important, and they must know that there are multiple decompositions to these problems and how to calculate them.

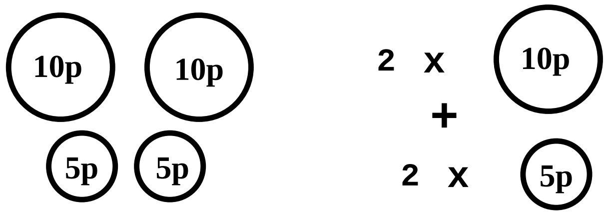


Figure 1. Typical problems at stage 2 and stage 4

Difficulty of problems is manipulated in two ways: abstraction of the question notation and the difficulty of the arithmetic operations with each notation type. There are six types of representation, ranging from problems that simply require addition of graphical components (*i.e.* pictures of coins on the screen, see figure 1), through multiplication problems with mixed text and graphics, to questions involving an algebraic notation (*e.g.* ‘What is 3t + 5k?’).

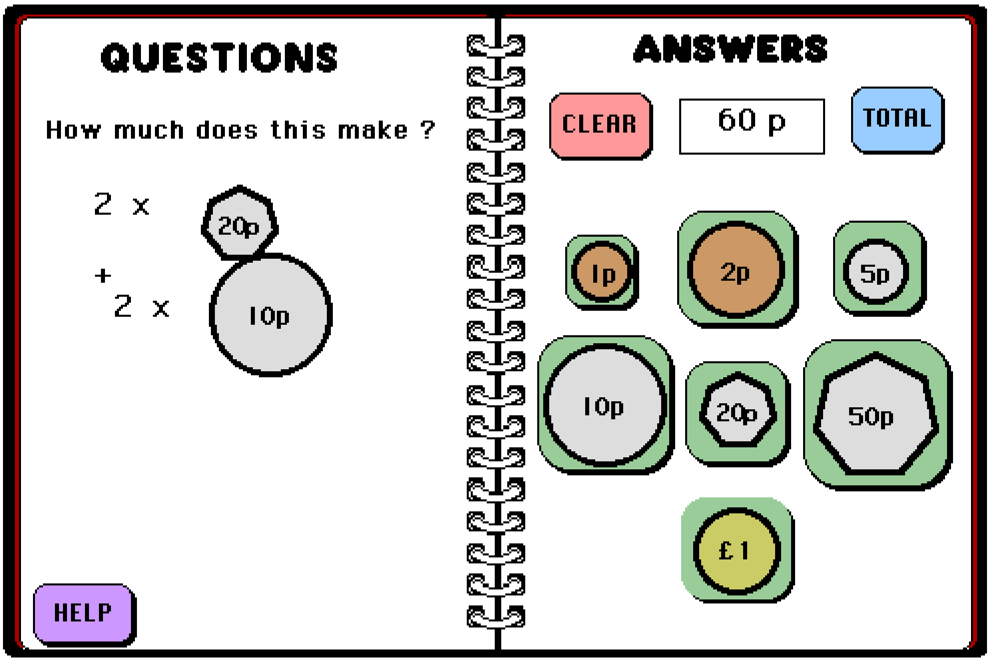
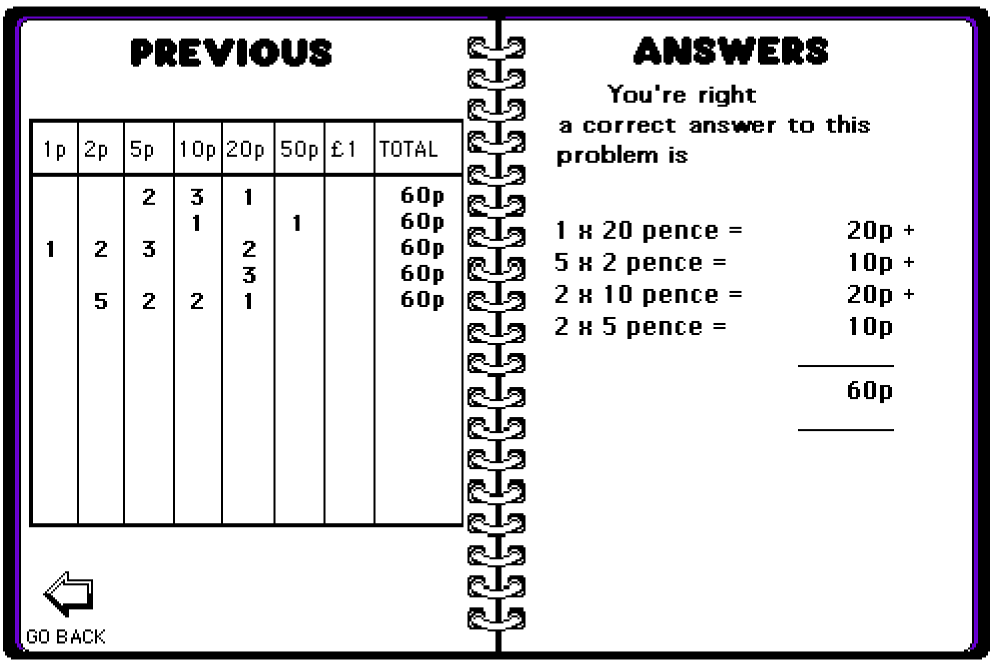


Figure 2. The ‘coin calculator’ and typical question

Students answer questions by pressing buttons on a ‘coin’ calculator whose buttons are British coins. The coin calculator is designed to provide a simple way of interacting with the system which is both easy and fun to use and which reduces the burden of number facts (see figure 2). After each question, users receive feedback on their solutions which reveals whether they were correct, and describes the elements of their answers in detail (figure 3). They are also shown a ‘trace’ of all their previous answers on that question.



Empirical

Figure 3. Feedback provided after each question

Introduction

Two experiments have been conducted using this system. In each we investigated children's ability to produce multiple solutions and examined whether a computer environment can support the learning of such skills. First, we examined whether children could easily produce multiple answers to coin problems. If the concerns of researchers such as Lampert and Baroody apply to English primary school children, we would expect that such an apparently simple task will prove difficult. Secondly, we wanted to treat COPPERS as a testbed to examine what features of a computer environment would help teach children to consider multiple solutions. Hence, our aim was not to compare a computer environment with others forms of learning (*e.g.* teachers, pen and paper, real coins). Instead, we conducted detailed within system evaluations to identify which features contribute most to learning.

If we wish to encourage multiple answers, we need to consider how many solutions should be given for each answer. Hence, the first feature evaluated was the consequence of requiring multiple correct answers per question rather than just a single answer. Pilot work suggested that most users could be persuaded to give four answers per question, so we contrasted users giving a single answer per question with those giving four.

The second aspect of the system that was evaluated was the representation of feedback. There is abundant evidence that the way information is presented affects how people reason, (*e.g.* Larkin & Simon, 1987). In addition, it has been claimed that presenting multiple linked representations results in flexible and insightful learning (*e.g.* Kaput, 1989). COPPERS tells users whether their answers are correct and displays answers in partial products (Figure 3). This is performed in two ways and, by highlighting, the system encourages students to map between the different representations. The first representation is a standard place value notation . The user is reminded of each type of coin they used and the number of coins for each type. The second form of representation is the table representation similar to the one described by Lampert. The answers are displayed in columns. To see how the values in the table correspond to the final total, the student must multiply the number in each column with the number of pence at the top and then add them together. The table also displays previous answers to the question. This allows students to compare their answers with those already given and (hopefully) prompts pattern seeking and reflection. However, it is known that young children have difficulty in working with tabular representations and often fail to use them successfully (*e.g.* Underwood & Underwood, 1987). Hence, performance will only be enhanced if learners can successfully meet the learning demands inherent in the addition of the second representation.

The impact of these features (multiple answers and table feedback) was examined by producing several forms of the program, varying the presence of these elements. Performance was measured by examining subjects' pre-test to post-test changes on pen and paper questions similar to those set by the computer. We examined the number of novel correct multiple solutions which measures both accuracy of the initial calculation and skill at producing accurate decompositions of this total. In addition, we measured the total number of solutions (irrespective of accuracy) which simply examines the number of decompositions. It was suggested that children's performance will improve if they give multiple rather than single answers on the computer. Furthermore, it was proposed that they would perform better if given feedback with the tabular in addition to the place value representation.

Experiment One

Method

Design

A three factor mixed design was used. The first factor was the presence of a summary table in addition to the place value notation (table, notable). The second was the number of answers required for each question (multiple, single). Half the children were required to give four correct novel answers to four questions and half to give a single answer to 16 questions. The third factor was a within groups measure, time (pre-test, post-test, delayed-test). This resulted in four experimental groups, with ten subjects in each group. Subjects were assigned to conditions using a randomised block design by mathematical ability. Each had the same number of boys and girls and the mean age of the subjects did not differ.

Subjects

Forty mixed ability year two pupils from a state infant school took part in the experiment. They ranged in age from 6 years 10 months to 7 years 9 months; the mean age was 7 years 3 months. All children had some experience with both calculators and computers.

Materials

A general test of mathematical concepts and skills for seven to eleven year olds was given to all the subjects (Basic Number Screening Test - Gilham and Hesse 1976).

Pre-test and Post-test Material: These tasks examined children's ability to give multiple solutions to the sorts of coin problems generated by COPPERS. The tests consisted of three coin problems; there was between 28 and 4900 possible correct solutions for each problem. In order to answer the question, the children were given blank pieces of paper and instructed to draw coins that would make the same total as the one in the question (although they did not have to write this total). They were encouraged to produce as many solutions as they could for each question before moving onto the next one. Children could approach this task in two ways - either by first calculating the total and then producing different decompositions to the whole amount, or by calculating sub-totals and producing different decompositions for each of the these.

Procedure

1) Pre-tests: An experimenter gave children mathematics ability and multiple solutions tests in groups of five. The importance of working individually was stressed and reassurance was given about the tests.

2) Computer Intervention: Subjects used the computer individually in a quiet space with an experimenter present to help explain the instructions. To ensure sufficient practice with the system, each child used COPPERS twice (the total time spent on the computer was between 60 and 90 minutes), separated by approximately two weeks.

3) Post-test: Two further multiple solutions pen and paper task were administered to the subjects within a) ten days of their second computer trial and b) six weeks after that.

Results

To examine the effects of the intervention, a number of [2 by 2 by 3] ANOVAs were carried out on the pre-test, post-test and delayed post-test data. The design for the analyses was 2 (table, no-table) by 2 (multiple, single) by 3 (pre-test, post-test, delayed test). The first two factors were between groups and the third a within group repeated measure. The results from one subject have been dropped. He was an extreme outlier scoring nearly six standard deviations above the mean at pre-test. Four children were unavailable to take the delayed post-test. The first measure examined was correct multiple solutions. Note that all the data is presented per test, the average number of solutions per question can be found by dividing this total by three.

Table 1

Mean Number of Correct Novel Multiple Solutions by Condition and Time

	Table				NoTable			
	Multiple		Single		Multiple		Single	
Time	M	SD	M	SD	M	SD	M	SD
Pre-test	2.60	2.12	1.78	1.20	1.88	2.03	4.38	3.2
Post-test	14.70	5.54	10.22	6.20	8.13	6.31	8.38	4.87
Delayed	8.50	5.21	8.44	5.88	7.88	7.02	9.63	10.10

There was a significant main effect of time (F(2,62) = 30.69, $p<0.001$). There was no significant interaction between multiple and time, but the interaction between table and time was significant (F(2,62) = 3.70, $p<0.030$). Simple main effects showed the only significant difference between the conditions was at post-test, where the table group produced significantly more correct solutions. Tukey's tests showed that the groups improved over time from pre-test to post-test and from pre-test to delayed-test: table (q = 10.53, $p<0.01$ & q = 6.36, $p<0.01$); and no-table (q = 4.78, $p<0.01$ & q = 5.24, $p<0.01$). However, the table group's scores also decreased significantly from immediate to the delayed post-test (q = 4.17, $p<0.05$) although remaining significantly above pre-test performance.

Children's pre-test performance differed widely. There was a great deal of variability in the number of solutions (both correct and incorrect) given at the pre-test (2 to 37), the median being 5. We wished to perform an aptitude by treatment analysis, but due to small cell sizes it was not possible to split the data by the median. It was decided to look at those subjects who had most to learn and so higher performers at pre-test were removed from the sample and the results re-analysed. Twenty-nine of the subjects gave seven or less answers on the pre-test, which is displayed in table 2, and the remaining eleven gave nine or more answers.

Table 2

Mean Number of Correct Novel Multiple Solutions for Subjects Scoring Less than 3 Answers per Question at Pre-test.

	Table				NoTable			
	Multiple		Single		Multiple		Single	
Time	M	SD	M	SD	M	SD	M	SD
Pre-test	2.43	1.92	1.87	1.25	2.00	1.94	3.00	2.45
Post-test	14.00	6.30	8.50	5.78	8.11	4.65	5.00	1.87

There were insufficient numbers of subjects at the delayed post-test, so only the pre-test and post-test scores were examined. When an [2 by 2 by 2] ANOVA was performed, the significant interaction between table and time remained ($F(1,25) = 5.50$ $p<0.03$). However, in contrast to the analysis of practice and time with **all** subjects, the interaction between multiple and time was significant ($F(1,25) = 4.44$, $p<0.045$). Analysis showed that the groups only differed at post-test ($F(1,50) = 5.912$, $p<0.019$), and that both conditions improved significantly: multiple ($q = 8.51$, $p<0.01$) and single ($q = 4.37$, $p<0.01$).

The second measure of performance is the number of solutions given irrespective of accuracy. This measure includes incorrect answers created either because the initial calculation was in error (the majority were of this form) or because of a mistake in the decomposition (often a slip such as writing 41 rather than 42 pence coins). Table 3 shows the results for all subjects expressed as the mean number of novel solutions for the three questions per test.

Table 3

Mean Number of All Novel Multiple Solutions by Condition and Time

	Table				NoTable			
	Multiple		Single		Multiple		Single	
Time	M	SD	M	SD	M	SD	M	SD
Pre-test	7.60	4.62	6.67	5.05	4.50	1.42	7.63	3.34
Post-test	15.30	5.81	12.44	7.95	9.50	4.96	9.63	6.30
Delayed	10.80	4.29	9.33	6.97	10.25	6.39	11.00	9.39

There was a main effect of time (F(2,62) = 16.20, p<.001), and a significant interaction between table and time (F(2,62) = 3.54, p <0.035). Again, a simple main effects analysis revealed a single significant difference between the groups which occurred at post-test (F=3.542 p<0.035). Both groups also improved significantly from pre-test to post-test: table (q = 7.63, p<0.01) and no-table (q = 4.71, p<0.05). However, the table group's performance dropped significantly from post-test to delayed test (q = 4.31, p<0.01).

Finally, we can look at percentage accuracy (total correct solutions / (total correct solutions + total errors)). Accuracy improved over time (F(2,62) = 30.31, p<0.001), from 43% to 87% by post-test and remained stable at delayed post-test with 80%.

Types of solution

We performed some preliminary analysis of the types of solutions children produced by examining both the number of coins per solution, and the number of different types of coin per solution at both pre-test and post-test. In both cases there were main effects of time. The number of coins per answer increased from an average of 5.76 at pre-test to 10.94 at post-test (F(1,35)=22.71, p<0.001). At pre-test, there were 2.09 types of coins per question which had increased significantly to 2.67 by post-test (F(1,35)=31.28, p<0.001).

Interaction Strategies

Subjects in all conditions had to calculate the problem total regardless of whether they produced multiple depositions of this total. They commonly employed two types of strategy to reach this total. The first strategy was to use the 'coin calculator' to copy the format of the question. For example, if the question asked "what is 2 x 20p + 3 x 10p ?" they would enter two 20 pence's and three 10 pence's. A second strategy was to calculate either part or whole of the sum and then press different coins to reach this total. For example, users might say 2 x 20p = 40p, and then press four 10 pence's to make the total. The percentage of times that subjects used the 'copying' strategy was significantly negatively correlated with both the number of correct solutions at the post-test (r = -0.3561, p < 0.05) and the total number of solutions (r = -0.3472, p<0.05). The more times a student used this strategy on the computer, the poorer their subsequent performance on the pen and paper tests

General Mathematical Ability

Subjects had been given a general mathematics test (Basic Number Screening Test) at the beginning of the study. No significant correlation was found between the number of correct solutions and the BNST at pre-test (r = 0.130). However, the correlation between BNST and the total number of solutions was significant (r = 0.334, p < 0.05). The BNST scores did not correlate with any measure taken at the post-test. The only other measure that correlates with BNST was the subject's strategy for producing the initial total (*i.e.* percentage of the time the question format was copied) (r

= -0.46, $p < 0.01$). The higher the BNST score the less likely a subject was to use the copying strategy.

Experiment Two

A second experiment was designed to address issues raised by the first experiment. The system was re-implemented before a second experiment and a number of changes were made.

The coin calculator was replaced by money tubes. One tube is empty and corresponds to part of the question to prevent the execution of the ‘copying’ strategy. The table is now visible continuously and updated as each coin is selected. Finally, a ‘score’ indicator was added which consisted of a pointer and a numerical score.

The system was used to further examine the issue of how to set children the ‘right’ number of solutions per question. This was approached in two ways. Firstly, we allowed children to choose how many solutions to give per question. There is intuitive appeal for this idea; children could move at their own pace and it is hoped become more interested and motivated. However, the complicated literature on learner control (*e.g.* Steinberg, 1989) suggests that giving learners this sort of control will only be effective to the extent that they chose a successful learning strategy. Results from the previous study indicated that an effective strategy would maximise the number of answers per question. Hence, we proposed to examine whether children would develop such strategies.

Secondly, the previous study had found only lower performing subjects benefited from giving four answers per question. It was hypothesised to improve high performers, they needed to give more answers per question. Hence, an eight answer condition was included.

Method

Design

A two factor mixed design was used. There were four levels to first factor (condition) which varied the number of answers children were required to give to each problem they were given on the computer. The first group were required to give four answers to four question (four), the second group eight answers to two questions (eight), the third group could choose how many answers they gave per question (limited to 16 answers) (autonomous) and the fourth group were a no-treatment control group. This resulted in four experimental groups, with ten subjects in three groups and 20 in the autonomous group. The second factor, time, was within groups. Each group had similar number of boys and girls. The mean age of the subjects and their scores on a maths test did not differ significantly. Dependent variables were identical to those used in the previous experiment.

Subjects

Fifty mixed ability, year four pupils, from a state junior school took part in the experiment. They ranged in age from 8 years 4 months to 9 years 2 months; the mean age was 8 years 9 months. All children were familiar with calculators and computers.

Materials

A general test of mathematical concepts and skills for seven and eight year olds was given to all the subjects; the Y1 (Young, 1979)

Pre and Post-test Material: These were identical to those used in Experiment One.

Procedure

This was similar to the first experiment. The only major difference was that a different mathematics test, the Y1, was administered. This was given to the children in groups of ten and a second experimenter was present. The computer intervention followed the same procedure as before. The autonomous group were told that they had to give 16 answers to the computer, but that they could choose how many answers to give per question and the scoring mechanism was explained to them. A 'no treatment' control group were included who simply took the pen and paper tests.

Results

To examine the effects of the intervention a number of [4,3] ANOVAs were carried out on the pre-test, post-test and delayed post-test data. The design of the analyses was 4 (control, four answers, eight answer, autonomous) by 3 (pre-test, post-test delayed post-test). The first factor is between groups and the second a within group measure. Two subjects have been excluded from the analysis. The first child was recognised as having special educational needs and the second scored significantly higher than the other subjects at pre-test.

The first measure examined was the number of novel correct solutions given for the three item pen and paper tests (see Figure 4 and table 4).

Table 4

Mean Number of Correct Novel Solutions by Time and Condition

	Control		Autonomous		Four Answers		Eight Answers	
Time	M	SD	M	SD	M	SD	M	SD
Pre-test	3.20	3.23	4.32	3.85	3.22	3.114	3.90	3.67
Post-test	3.10	3.70	8.05	3.10	6.67	2.915	11.80	3.52
Delayed	3.50	4.12	7.42	4.21	7.44	3.844	12.00	3.56

There were significant main effects of time ($F(2,88) = 33.03$, $p<0.001$) and condition ($F(3,44) = 9.08$, $p<0.001$). The interaction between condition and time was also significant ($F(6,88)=3.75$, $p<.002$). Simple main effects showed significant differences at post-test ($F(3,132) = 5.14$, $p<0.002$) and delayed post-test ($F(3,132) = 5.44$, $p<0.002$). All but one of the experimental groups showed significant increase in performance from pre-test to post-test and from pre-test to delayed post-test: four ($q = 3.32$ & $q = 3.37$, $p<0.05$); autonomous ($q = 5.23$, $p<0.01$ & $q = 4.34$, $p<0.01$) and; eight ($q = 8.02$, $p<0.01$ & $q = 8.22$, $p<0.01$). There was no significant change from post-test to delayed post-test in any condition.

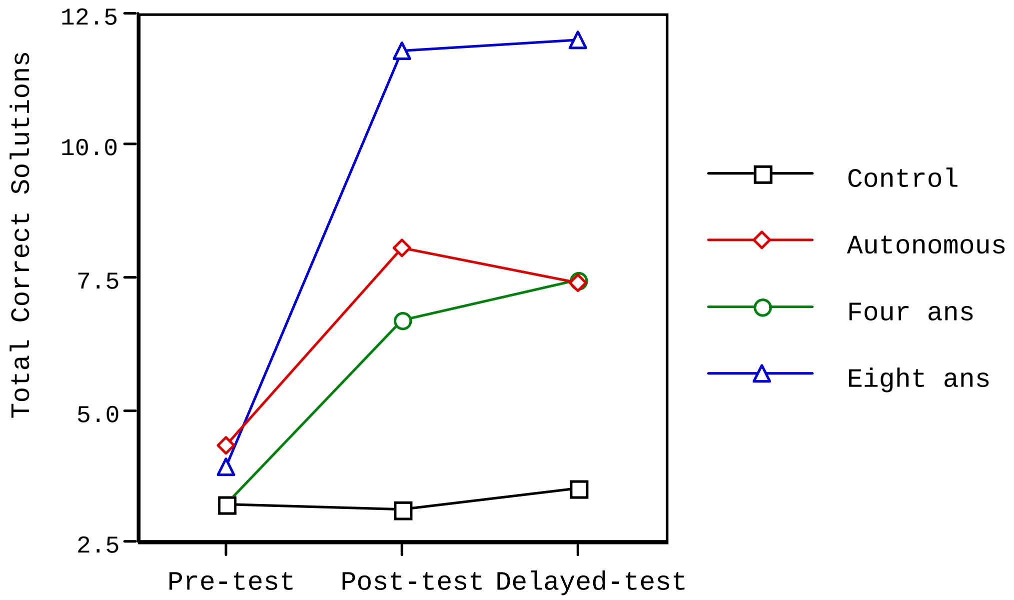


Figure 4. Correct novel multiple solutions by condition and time

Another measure of performance is the number of solutions irrespective of accuracy (Table 5 and Figure 5).

Table 5

Mean Number of All Novel Multiple Solutions by Condition and Time

	Control		Autonomous		Four Answers		Eight Answers	
Time	M	SD	M	SD	M	SD	M	SD
Pre-test	7.80	3.49	6.74	3.75	8.89	4.14	8.70	3.60
Post-test	6.10	2.84	8.48	2.92	8.56	5.22	13.30	2.41
Delayed	6.70	3.10	8.79	4.03	8.56	2.41	13.3	3.16

There was a trend for improvement over time (F(1,45) = 3.63, p <0.063) but there was a significant main effect of condition (F(3,44) = 7.23, p<0.001). The interaction between condition and time was also significant (F(6,88) = 3.08 p <0.009). Although the differences in the means seem large, the limited numbers of subjects coupled with high variances led to few significant differences: the control and eight group at post-test (q = 6.44, p <0.01), and four and eight groups (q = 3.80, p<0.05). At delayed post-test the only significant difference was between the eight and control groups (q=6.29, p<0.01).

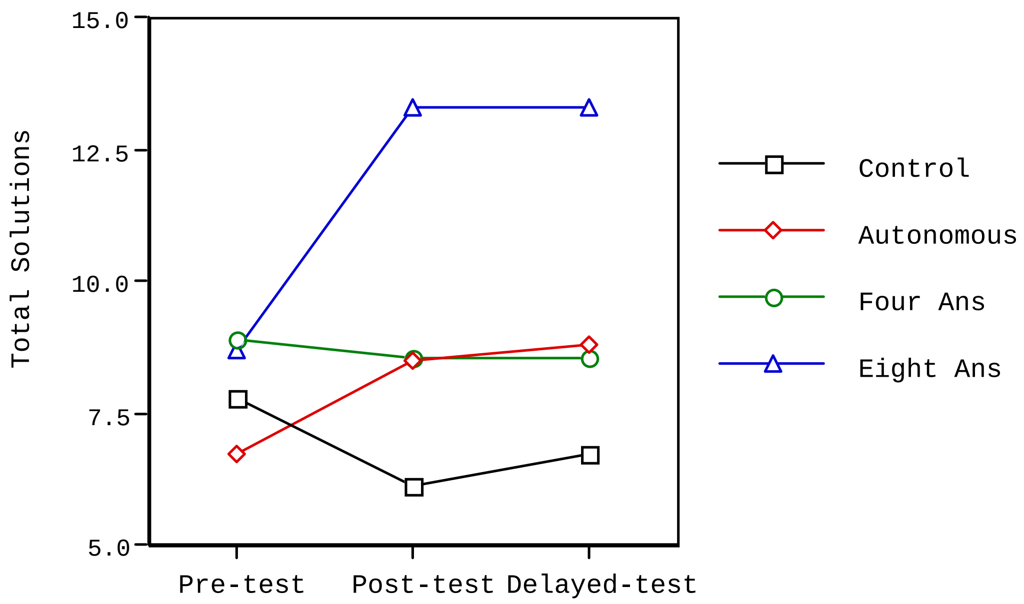


Figure 5. All novel multiple solutions by condition and time

The only significant differences between the conditions was at both post-test (F(3,32) = 8.02, p<0.001) and at delayed post-test (F(3,32)= 8.023, p< 0.001). Tukey's tests revealed that the number of solutions in the four, control and autonomous conditions did not differ on any occasion. Subjects in the eight condition produced significantly more solutions at post-test and delayed post-test than they had at pre-test (q = 5.23, p <0.01 & q = 5.23, p <0.01).

Finally, we can also look at percentage accuracy. At pre-test, the average percentage accuracy across all conditions was 48.2%. At post-test, the control group's scores remained low at 51.3%, whilst the average experimental groups score rose to 87.5%. This improved performance remained stable to delayed post-test, with the groups scoring 45.6% and 85.3% respectively.

Analysis demonstrated main effects of time (F(2,88) = 10.833, p<0.001) and of condition (F(3,44) = 6.176, p<0.002). Tukey's comparisons revealed that all experimental groups were significantly more accurate at post-test and at delayed post-test: autonomous (q = 5.24, p<0.05 & q = 3.38, p<0.05); four (q = 3.84, p<0.05 & q = 4.15, p<0.05) and; eight (q = 3.93, p<0.05 & q = 4.21, p<0.05). There were no significant changes for the control group.

Interaction Strategies

The autonomous group answered an average of 9.6 questions during their interaction with the computer and therefore produced an average of 1.66 answers per question. They were very consistent in their approach. The correlation between the number of questions answered on their first and second interaction was significant (r = 0.76, p<0.001). However, the variability was large with some children answering 16 different questions and one giving 16 different answers to the same question. The number of answers per question was not related to general mathematics ability, as measured by the Y1 (r=-0.06). There was no correlation between the strategy used and number of correct solutions either before or after the intervention. The only significant correlation was between strategy and total number of solutions at delayed post-test (r=0.39, p<0.05). Hence, any relation between the strategy children in the autonomous group used on the computer and performance outcomes is weak.

General Mathematics Ability

The Y1 correlated significantly with the number of correct solutions generated by the subjects at pre-test (r=0.40, p<0.01) but not at post-test or delayed post-test (r=0.253 and r=0.235). The correlation with total number of solutions was also significant at pre-test (r=0.330, p <0.025) but not after the intervention (r = 0.165 & r = 0.172). It would seem that general mathematics ability was related to children’s initial ability to produce multiple solutions on these tasks but that intervention weakened this relation.

Discussion

The first goal of the experiment was to investigate children’s initial performance at producing multiple answers to coin problems. Given the research on the nature of children’s beliefs about mathematics, we proposed that children would initially perform poorly on this task. This hypothesis was supported by our experiments. We found that at pre-test, the 6-7 year old children produced an average of 2.66 correct answers across three questions, *i.e.* less than 1 correct answer per problem. This poor performance was also true for the 8 - 9 year old children, they produced an average of 1.22 correct answers per question at pre-test. Even if we include errors, the children’s scores do not improve substantially. The younger children produce an average of 2.2 answers per question and the older children, 2.7 answers per question. Hence, it would appear that primary school children do not easily produce multiple answers to these problems upon demand.

COPPERS appears to support the development of these skills. In Experiment One, subjects improved upon their pre-test scores by nearly 400% to produce an average of 3.5 correct novel solutions per question. In Experiment Two, subjects produced an average 2.95 correct answers per question at post-test. This increase in performance can be compared to the control group who produced 1.07 correct answers per question at pre-test and 1.03 correct answers at post-test. Improvement in performance seems to be relatively robust. In Experiment Two, there was no change in performance from post-test to delayed post-test. There was a drop in performance in Experiment One although scores remained significantly higher than pre-test. However, delayed testing took place in the last week of the summer term - less than ideal circumstances. We can claim, with some justification, that COPPERS is effective at teaching children to produce multiple solutions in this domain.

Types of solution

In addition to increasing the number of solutions, we were also interested in examining whether the types of answers changed after the intervention. It is difficult to perform much pre-test to post-test comparison. The number of correct solutions was very low at pre-test, and our main aim was to test the effectiveness of the learning environment, not to probe children's strategies which would have required more detailed protocol analysis.

We found that both the average number of coins per question and the average number of different types of coins per question increased from pre-test to post-test. The vast majority of decompositions at pre-test were very routine. For example, the most common correct answer to '2 x 1p + 2 x 20p' was '10p + 10p + 10p + 10p + 2p', 11 out of the 18 children who gave a correct answer to this problem, generated this solution. The second most common was '20p + 20p+ 2p' given by 9 of these children. Given the limited number of correct answers at pre-test, these two solutions accounted for the majority of correct answers. In total, only 15 different decompositions were identified for this question at pre-test from a possible total of 271.

At post-test there is a completely different picture. There was much greater variety of solutions, both within and between individuals, although, there were still some preferred responses ('2 x 20p + 3 x 2p' = '1p + 5p + 10p + 10p + 20p', and '1p + 5p + 10p + 10p + 10p + 10p'). These two decompositions were given by 11 and 12 children respectively out of a possible 39, and together accounted for only 15% of the solutions. From the 151 total answers generated for question one, 46 different decompositions were identified. In addition, solutions tended to be much less routine, '1p + 2p + 2p + 2p +2p + 2p + 5p + 10p + 10p + 10p' or '1p + 1p + 1p + 1p + 1p + 1p + 1p + 1p + 1p + 2p + 5p + 10p + 20p'. The performance of the children at post-test suggests a much more flexible and inventive approach to decomposing numbers. Rather than learning a few common approaches to these problems from the computer, the range of solutions given suggests that children were generating their own decomposition strategies.

Multiple Solutions

In the first experiment, subjects who were required to give four rather than one solution per question did not produce more correct solutions at post-test. However, for the lower performing children giving multiple answers did lead to significantly better performance. We suggest that one explanation is that the number of solutions requested by the computer was not sufficiently stretching for the high performing subjects. The computer asks for four correct answers per question - higher performing subjects had produced an average of four answers initially. Hence, the second experiment required some children to give eight answers. All the experimental groups had significantly better performance at post-test. However, the eight answer group had significantly more correct answers than the four answer and control groups at post-test, and the control group at delayed post-test.

Improvements in generating correct solutions could be due to the two different skills of calculation and multiple decomposition. If we examine the total number of solutions irrespective of accuracy, we simply measure the number of decompositions. For this measure, the prediction that all experimental groups would produce more answers than the control group was not supported. However, the eight answer group did show significant improvement in performance from pre-test to post-test. Thus, the significant improvement for total correct solutions observed for all the experimental groups had different causes. The four and autonomous groups improved primarily because the accuracy of their calculation increased, whereas the eight group, in addition to becoming more accurate, also increased the total number of decompositions.

We suggest that the reason only eight answers (as opposed to four) proved effective is related to the zone of proximal development, (Vygotsky, 1978). This is the region of activity in which learners can perform successfully given the aid of supporting context, in this case that of the computer. Taking this view, it is necessary to set problems on the computer that would be out of reach for children without assistance. However, to diagnose the dimensions of the zone of proximal development is a difficult task. Nevertheless, we should be able identify its lower boundary by analysing the child’s unaided performance. With this information, it should then be possible to set problems that are out of reach for the unsupported child and which therefore fall within their zone of proximal development.

Summary Table

The representations used by COPPERS were examined in Experiment One. We examined the use of a summary table to provide information about the current and, where relevant, previous answers to a question. The hypothesis that the table would improve the children’s performance was supported. The children who encountered the summary table produced significantly more solutions at post-test than children who did not see the summary table.

Lampert (1986a) proposed that such tables would be useful for allowing subjects to compare their previous answers, especially those that had been in error, and to provide a record of work for their teacher to analyse. In addition, the table reminds students of previous answers so that they are not repeated, either because of memory lapses or because of misunderstandings of commutativity of addition. However, these functions were proposed in the context of multiple solutions; the current study found that the table led to better performance regardless of the number of answers the children gave on the computer.

The table may affect generation of solutions in many ways and a number of different approaches were observed: children tried to use as few columns as possible or as many as possible; they might try to get high numbers in particular columns; make patterns across the columns, *etc*. For example, one subject noted his answer read like a palindrome across the table, “its the same backwards as forwards” and tried to create another palindrome on his next go. Given recent emphasis on mathematics as ‘the science of patterns’ we feel it is encouraging that children were beginning to seek and generate patterns in their answers.

Why should the table and the place value notation but not the place value by itself have led to these hypothesised strategies? Research over a number of years has found that learning and problem solving may be enhanced by the use of appropriate external graphical representations (*e.g.* Larkin & Simon, 1987). Tables tend to make information explicit, emphasise empty cells and hence direct attention to unexplored alternatives, highlight patterns and regularity and represent variability (*e.g.* Cox & Brna, 1995). In this case, it also emphasises order. Hence, the table makes different information salient. In addition, the table serves as a symbolic representation of the multiplication and addition procedures involved in finding solutions to the problems. Numbers in the columns must be multiplied by the numbers in the column heading and then added together to get the total amount of money. The operators used to interpret a table therefore require children to practise multiplication and addition, skills that COPPERS is attempting to teach.

The table and place value representations simultaneously provide information on the same problem in different ways. Recently, a number of researchers have argued that by employing such multiple linked representations, learners will come to see and apply complex ideas in novel ways and that the learning that results from multiple representations will not be shallow and procedural, but flexible and insightful (Kaput, 1989). It is claimed that multiple representations allow different ideas and strategies to be supported (*e.g.* Tabachneck, Koedinger, & Nathan, 1994) and that translating between different representations leads to better understanding (Kaput, 1989). Hence, although there are many explanations of the improved performance of the table group, the current experiment does not allow us to isolate which one(s) it may be. However, these results do indicate the importance of (multiple) external representations for learning in this domain.

Strategies and Learning

The first experiment found that the way an initial value was calculated affected subsequent performance. If children calculated the answer to the whole problem (or part of the problem) and then chose which buttons to press to reach this total, they practised addition and multiplication, and decomposing a total. However, if they chose to simply press the coins presented in the question, they limited their practice at addition and multiplication. Unsurprisingly, children who used this strategy did not show the same amount of improvement as children who commonly used other strategies.

The autonomous condition in the second experiment looked at the strategies learners used when given choice about how many answers to give per question. Children given this choice provided an average of 1.66 answer per question. This is perhaps surprising given that the children knew that they would receive more points if they gave more answers per question. A first hypothesis that they were not motivated by the points system seems unlikely, given their comments. They were concerned to know how well they were doing (and how their friends were doing!). Many children seemed to be caught between the lure of the points and that of new questions making explicit bargains with themselves (*e.g.* “I’ll have a new one, but I’ll answer three on the next one”). Again, it would seem that children’s beliefs about the nature of mathematics affected the way they used the system, in a potentially deleterious way.

It is difficult to tell how strategy affected performance. There was little relation between how many questions subjects answered on the computer and post-tests measures. The only significant result was that children who gave more answers on the computer performed better at delayed post-test. However, although the range of answers per question was large, the majority of children gave very few answers per question, 79% gave on average less than two answers per questions. The one subject who chose to give 16 answers to one question on the computer, showed the greatest improvement within the autonomous group.

Aptitude and Multiple Solutions

No simple pattern was found of relations between general mathematical aptitude and the production of multiple solutions. It was not the case that children who scored better on the general mathematics tests were better at this task. For example, in Experiment One there was no correlation between the mathematical test scores and any measure of post-test performance and only one significant correlation at pre-test (the total solutions regardless of accuracy). In Experiment Two, there were significant correlations at pre-test with accuracy and correct solutions but none at post-test. Nor is there a consistent affect of how ability mediates interaction with the computer. In Experiment One, we found that lower performing children tend to pick a worse interaction strategy but in Experiment Two there was no relation between mathematical ability and strategy in the autonomous group.

It is tempting to speculate that one of the reasons for these diverse results is that children with better mathematics skills have already stabilised their notions about the nature of mathematics (*e.g.* one correct answer per question as quickly as possible). Thus, the mathematical skills and knowledge that they could use on these problems are under-utilised.

Conclusions

We have evaluated a computer based learning environment that attempts to embody the types of instructional method being called for in primary mathematics. In particular, we concentrated upon examining children's skills at producing multiple answers for a single problem. Initially children were found to produce a very low number of solutions, but, with a limited amount of teaching, they show impressive and sustainable improvements. Two aspects of the computer system were found to be strongly positively related to learning outcomes. The first was the requirement to direct children to produce (with support) more solutions than they would naturally give. The second was the benefit of providing a second tabular representation of users' answers. Hence, when structuring children's investigation into multiple solutions (skills and beliefs argued to be fundamental to mathematical knowledge) we propose that the provision of this type of support will facilitate learning.

References

Baroody, A. J. (1987). Children's mathematical thinking: A developmental framework for pre-school, primary and special education teachers. New York: Teachers College Press.

Cox, R., & Brna, P. (1995). Supporting the use of external representations in problem solving: The need for flexible learning environments. Journal of Artificial Intelligence in Education, 6, 239-302.

Basic Number Screening Test (1976). W.E.C. Gilham & K.A. Hesse, SevenOaks: Hodder and Stoughton.

Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. Grouws (Eds.), Handbook of teaching and learning mathematics. New York: Macmillan.

Hughes, M. (1986). Children and number. Oxford: Blackwell.

Kaput, J. J. (1989). Linking representations in the symbol systems of algebra. In S. Wagner & C. Kieran (Eds.), Research issues in the learning and teaching of algebra. Hillsdale. NJ: Lawrence Erlbaum Associates.

Kaput, J., & Maxwell-West, M. (1994). Missing-value proportional reasoning problems: Factors affecting informal reasoning patterns. In G. Harel & J. Confrey (Eds.), The development of multiplicative reasoning in the learning of mathematics (pp. 237-292). Albany, New York: State University of New York Press.

Lampert, M. (1986a). Knowing, doing, and teaching multiplication. Cognition and Instruction, 3, 305-342.

Lampert, M. (1986b). Teaching multiplication. Journal of Mathematical Behaviour, 5, 241-280.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11, 65-99.

Nunes, T., & Bryant, P. (1996). Children doing mathematics. Oxford: Blackwell.

Schoenfeld, A.H (1986). On having and using geometric knowledge. In J. Hiebert (Eds.), Conceptual and procedural knowledge: The case of mathematics (pp. 225-265). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem Solving, metacognition and sense making in mathematics. In D. Grouws (Eds.), Handbook of teaching and learning mathematics (pp. 334-270). New York: Macmillan.

Steinberg, E. (1989). Cognition and learner control. A literature review. Journal of Computer Based Instruction, 16, 117-121.

Tabachneck, H. J. M., Koedinger, K. R., & Nathan, M. J. (1994). Towards a theoretical account of strategy use and sense making in mathematical problem solving. In A. Ram & K. Eiselt (Ed.), Proceedings of the 16th Annual Conference of the Cognitive Science Society, Georgia Institute of Technology, Atlanta, Georgia: LEA.

Underwood, J., & Underwood, G. (1987). Data organisation and retrieval by children. British Journal of Educational Psychology, 57, 313-329.

Vygotsky, L.S. (1978) Mind in society: The development of higher psychological processes. (M. Cole, V. John-Steiner, S. Scribner, E. Souberman, Eds. and Trans.). Cambridge, MA: Harvard University Press.

Y1 (1979). D.Young. SevenOaks: Hodder and Stoughton.

Author Note

The initial implementation of this system was completed at the University of Sussex with the invaluable guidance and support of Ben Du Boulay and supported by the EPSRC (formally SERC). Further development and the experimental work was conducted at the ESRC Centre for Research in Development, Instruction and Training. The authors would particularly like to thank Peter Bibby, Sue Cavendish, Jean Underwood and Helen Woodiwiss for their time and help. Finally, we would like to acknowledge the other members of the centre for providing a stimulating and supportive atmosphere in which to conduct research.

Correspondence concerning this article should be addressed to Shaaron Ainsworth, ESRC Centre for Research in Development, Instruction & Training, Dept. of Psychology, University Park, Nottingham, NG7 2RD.