Conceptual Gain and Successful Problem-solving in Primary School Mathematics

PAMELA DAVENPORT
The Anna Freud Centre

CHRISTINE HOWE
Centre for Research into Interactive Learning, Department of Psychology,
University of Strathclyde, Scotland, UK

SUMMARY This study investigated the effects of children solving addition and subtraction problems collaboratively in comparison with solving problems in the traditional manner of the classroom. Seventy-seven children were divided into experimental and control groups, the experimental children being assigned to groups of four with note taken of the ability and gender mix. Following a pre-test–intervention–post-test design, the experimental children worked together in their groups using problem-solving guidelines to solve a number of problems, thereafter ‘teaching’ their problem to a fellow pupil. Each child worked on six problems over a 3-week period, three of the problems in their groups subsequently teaching them to another, the other three problems being taught to them by another child. Over the same time period, the control group solved the same problems working individually at their desks.

The pre- and post-tests were analysed for number of problems correct or ‘score’, problem-solving strategy and execution of procedures, with pre-test scores being subtracted from post-test scores to give measures of change. The results indicated a main effect of ability on strategy change and a two-way interaction between gender and condition. They also indicated a main effect of condition for execution of procedures. Dialogue analyses indicated that more below average children improved their strategy understanding by listening to peers. The results themselves revealed variations in the way that children of different ability levels and gender can benefit from collaborative group work and thus have some interesting implications for the organisation of collaborative groups in the classroom.

Introduction

Peer Collaboration in the Mathematics Classroom

It is widely recognised that when children begin school, they come with a formidable knowledge of relevance to mathematics, this knowledge consisting of what Ginsberg & Allardice (1984) term System 1 and 2 skills. System 1 skills are preformalised and expressed mainly in action, while System 2 skills are cultur-
ally transmitted from adults, TV and books and expressed primarily in everyday language. The precise mix of System 1 and 2 skills varies from child to child, but no matter what form it takes there is a mismatch with the System 3 knowledge of written symbolisms, algorithms and explicitly stated principles that is encountered at school. This mismatch is in fact a clear instance of what Vygotsky (1978) was referring to in his discussion of academic or ‘scientific’ concepts. Scientific concepts are, according to Vygotsky, explicit and interconnected, and they differ in important ways from the ‘everyday’ concepts that children learn before they attend school.

Even worse though, the mismatch which exists for mathematics is something of which teachers are largely unaware, for the existence of prior knowledge via System 1 and 2 skills is not widely recognised. For instance, Smith (1994) found that no infant teachers in a primary school sample expected knowledge of addition and subtraction and less than 25% \( (n = 15) \) anticipated the number concept or number recognition. Because prior knowledge is not on their agenda, teachers have little chance of taking it into account, with predictable consequences. At best, System 3 knowledge will coexist with other forms, providing alternative routes to problem solving. At worst, it will be assimilated to the other forms making for a confused set of principles of service to no one. Although not talking in System terms, Siegler’s (1989) research into problem-solving strategies can be regarded as providing evidence for both coexistence and confusion. Moreover, since Siegler worked with children who were beyond the earliest years of primary schooling, his data document a sustained and recurring problem.

Teachers must then take prior knowledge into account, but how should they do this? The most obvious answer is to embed school mathematics in meaningful contexts, a point also emphasised by the National Curriculum for England and Wales (Department for Education, 1995). These guidelines promote a contextualised learning approach with reference to number skills and concepts, anticipating that ‘pupils should use number … in practical tasks, in real-life problems and to investigate within mathematics itself’. However, while this is desirable, it is unlikely to be sufficient. Real-life contexts would naturally imply real-life language, and the latter is not equivalent to the maths register. Forman (1992) makes this very point stating that while mathematics must be embedded in meaningful contexts, teachers should not allow these contexts to dictate the register. The difficulties that may arise if real-life contexts do dictate have been well illustrated by Walkerdine (1982). Walkerdine describes how children encouraged to say ‘No-oo’ to mathematics errors in pantomime style continued gleefully with ‘No-oo’ when inappropriate.

It may be suggested that the way forward is for teachers to present real-life problems in the formal register, and to explain carefully what the terms mean. However, on the Vygotskian (1978) account, this is unlikely to be a complete solution, for it renders children relatively passive in the explanatory process. For children to ‘internalise’ the register–problem linkages, it seems desirable that they should practise making the linkages by explaining for themselves.
Moreover, while doing this by explaining back to a teacher would have an element of phoniness, since teachers are assumed to be experts, no such assumptions will be made about classmates, suggesting that explanations to peers may be of value.

Although nothing is said about peer-based explanation within the National Curriculum, the recommendations of Mathematics 5–14 (Scottish Office Education Department, 1993), often seen as the National Curriculum's Scottish equivalent, do re-echo the above line of reasoning. There, the peer collaboration approach to problem-solving and enquiry skills is advocated, pupils to 'be challenged to think about what they are doing, to question and to explain'. Forman (1992) has also urged along these lines, indeed making the additional point that children are themselves aware that simple knowledge of the solution to a problem does not teach you anything. They realise how crucial it is to give an explicit explanation of how to solve the problem. Forman further found that children working in classrooms where collaborative discussions were actively encouraged, perceived themselves as intellectual resources, offering and receiving help from each other, sharing frustrations and being exposed to a variety of ideas.

Previous Research into Explanation and Problem-Solving

What then does the evidence suggest? Certainly it appears that peer collaboration with an emphasis on explanation can lead to an enhanced awareness of the cognitive processes and strategies which can be used to solve number problems. Research evidence that the explanations which children give to each other can be beneficial was reported by Webb (1992) in a review of 12 studies. Likewise, Peterson et al. (1984) found that higher ability children were more likely to recognise, to report attending to and to understand both lessons and the specific strategies that these used to connect new information with prior knowledge. However, can conditions be created where lower ability children benefit?

Work by Bargh & Schul (1980) suggests more all-round benefits if group work is followed by structured pairwise teaching. In particular, Bargh and Schul found that given the responsibility of teaching to another instead of simply learning for themselves, children of all levels can produce more highly organised cognitive structures. A similar point is made Thomson (1993): given responsibility for tutoring, a group children with learning difficulties were as effective at ensuring success as their mainstream peers. This said, the work was not located in the mathematics classroom and it remains to be seen how combinations of group and pair work would operate in that specific context. Moreover, even if the group–pair combination is successful in inducing strategic progress with all levels of ability, the question must be posed as to whether this can be translated into problem-solving success. Rapprochements between Piagetian and Vygotskian theorising such as Rogoff (1990) and Wood (1986) propose that the conditions conducive to conceptual or strategic advance are not equivalent to
the conditions conducive to procedural advance, suggesting that the answer might be negative.

A further factor for consideration with reference to working in groups is that of gender. Few studies have paid attention to this, but working from those which have, there are several issues to be resolved. Webb (1984), from a study examining sex differences in interaction and achievement in secondary level pupils working in groups of four, reported that girls asked for help more often than boys. Nevertheless, they were twice as likely to be ignored in comparison with boys who tended to receive more explanations per se. As a response to this difficulty, Peterson et al. (1991) reported a tendency towards equalisation of interaction within groups which were balanced, e.g. had the same number of girls and boys, in comparison to groups where one sex was in the majority. Further to this, McCaskin et al. (1994) suggest that balanced groups are superior to same sex groups when it comes to levels of helpful activities. Still on gender differences, when considering academic achievement Webb (1984) found boys’ levels of achievement to be higher than girls’ in groups where the sex distribution was unbalanced, while McCaskin et al. found no indication of gender differences in achievement. Consequently, while gender differences undoubtedly exist (Howe, 1997), the challenge is to address some of the inconsistencies which have arisen in the results from previous work carried out in this area.

All in all then, there are reasons for thinking that contextualised mathematics will be more effective if introduced during group-based collaboration between pupils, particularly if supplemented by pair-wise follow-up. However, conclusive evidence has yet to be obtained, and little is known about how effectiveness interacts with ability and gender. The prime concern of the present study was to provide such evidence by examining the effectiveness of a teaching programme which involved (a) presenting contextualised material, (b) dividing children into groups to solve problems collaboratively, and (c) assigning children to the role of teacher within pairs to guide each other through problems previously solved in their groups. Three questions were of interest:

1. Did the experimental group of children, those involved in the collaborative problem-solving intervention, benefit more than a control group who solved the same problems individually?
2. Did the benefits depend on the children’s ability/gender?
3. Is there a relation between benefit and the generation of good explanations during interaction?

Method

Design

The study followed a pre-test–intervention–post-test format, with both experimental and control children completing the pre- and post-tests. Following the pre-tests, the experimental children were assigned firstly into groups of four then
into pairs to participate in the problem solving intervention. The control children completed the same problems individually while sitting at their own tables.

**Sample**

The children were Primary 6 pupils attending two schools in the East End of Glasgow. Children from Primary 6 classes, approximately 10 years of age, were used principally for practical reasons. Children at this stage have reading skills sufficient to deal with the written problems and group instructions they would be dealing with and have experience of problem solving. In addition, they are not under the heavier curricular pressures associated with Primary 7, the final year of primary school. The two schools involved were equivalent in socio-economic background, both being located in a council estate with the majority of the population described as working class. School A had two Primary 6 classes running in parallel, one assigned as the experimental class of 27 children, the other as the control class with 25 children. School B had one class assigned as experimental with 25 children. Thus overall, there were 52 children in total in the experimental group and 25 in the control.

**Materials**

*The addition and subtraction problems.* The materials for the pre- and post-test comprised small booklets consisting of single- and multiple-answer story format problems. The multiple-answer problems were an amalgamation of two or more single-answer problems giving longer and more complex problems (Appendix 1). The problems were all addition and subtraction, but they varied in topic area and structure. The decision to create only addition and subtraction problems was made for several reasons. Addition and subtraction are central to mathematics teaching and are probably perceived as such by children—even very young children can perform simple addition and subtraction problems when the contexts are real and meaningful (Fuson & Hall, 1983; Hughes, 1986; Smith, 1994), yet these same children have difficulties when faced with equivalent problems in the formal maths register. Analyses of the way children solve addition and subtraction problems can also indicate the existence of competing and/or confused problem-solving strategies (Siegler, 1989). The topic areas covered by the problems were money, time, weight, distance (kilometres) and distance (metres). Working from the text books currently in use by the Primary 6 classes in both schools, these topics were identified as the ones covered within problem-solving for that academic year. The problems were created to conform in detail with the curriculum for Primary 6.

In addition to mathematical operations and choice of topic areas, attention was also paid to the structure of the problems. Carpenter & Moser (1982) have identified six different structures to story format problems. These are defined as follows:
(a) joining;
(b) separating;
(c) part–part–whole;
(d) comparison;
(e) equalising–add on; and
(f) equalising–take away.

Carpenter and Moser have evidence that the structures invoke different problem-solving strategies in children. Given the interest in the different strategies which children might employ when problem-solving, these structures were incorporated within the problems (see Appendix 2).

The pre- and post-test booklets each comprised 20 problems chosen from a pool of 250 addition and subtraction problems. The problems differed between pre- and post-test, but in both cases they covered the five topic areas and six structures. Piloting revealed that, while the problems discriminated between ability levels, they were of equivalent difficulty between pre- and post-test and across topic and structure. A selection of problems from the same pool were also used for the intervention section of the study.

Collaborative task: group and pairs instruction sheets. The problem-solving instructions given to the group, for which one child was responsible (Appendix 3), structured the activity by indicating when to read, think, talk, write, etc. The group was also given their own copy of the problem for that session on a white sheet of paper and a ‘group decision’ problem printed on a yellow sheet of paper. Briefly, the instructions had the children firstly working on their own to solve the problem, thereafter sharing answers and calculations. Work by Howe et al. (1990) has shown that having children commit themselves individually in writing is a powerful stimulus to participation in subsequent group discussion, hence the individual stage prior to the collaboration. If there was agreement at this stage, the instructions prompted the children to transfer the calculation and answer to the group decision sheet. They were then instructed to check the answer card and see if they were correct. If incorrect as a group, the children then had to try and identify through discussion, where they had gone wrong and attempt to remedy this. If, however, when initially exchanging solutions and pathways, there was disagreement, the instructions told the group to talk to each other and choose what they thought the correct answer ought to be.

Following the group phase, the children were placed into a pair with a child from another group who would have solved a different problem from them. Each child, once again with problem-solving instructions specifically for the pairs task (Appendix 4) took a turn at being the teacher, guiding their partner through the problem which they themselves had solved in their group the previous day. Briefly, the instructions directed them to have their partner read the problem and then ask what calculations had to be done and what numbers had to be used, correcting any mistakes at this point and checking if the ‘pupil’
understood how to solve the problem. The ‘pupil’ was then to work out the solution.

Procedure

Pre-test. All the children in the study completed the pre-test, which was in two stages, the first being to the whole class and the second to the children taken individually. For the whole class stage, the two researchers administering the test started by introducing themselves, and giving out the booklets. The children were asked to try and work out the answers to as many of the problems as possible on their own and to write down the way they did this. They were also told that if they calculated the answer mentally, this was perfectly okay but they should also write down how this was carried out. In addition the class was told that if they were really stuck and could not work out an answer, it was okay to miss it out and go on to the next problem. It was reiterated that they were working on their own and should ask no one for help, not even the teacher.

As soon as some children had finished, individual interviewing commenced. The children were taken two at a time to a large quiet room where the two researchers were waiting, one at either end of the room at a table and chairs with audio recording equipment on the table. One researcher worked with each child. The researchers initially chatted to the children to put them at their ease and then guided the conversation round to the interview by asking how they thought they had done and what they thought about the problems. They were told that they were to be asked some questions about two of the problems. They were told that the questions were to find out what type of calculation they had done and why they had done this. They were invited to talk through:

(a) how they had carried out the calculation or how they had done it in their head;
(b) anything they had used to help them, e.g. fingers to help them count; and
(c) how sure they were that they had the correct answer.

Questioning was carried out following a written interview schedule (Appendix 5). While all the interviews were audio recorded, brief notes of responses were also taken by the researchers in case of tape failure.

Problem-solving intervention. The children in the two experimental classes were placed into groups of four based on teacher rating of problem-solving abilities—below average, average or above average. Research by Kenealy et al. (1991), indicates teacher ratings of ability to be a reliable method of categorisation. Although grouping was constrained by the need to place all the children in the experimental classes in some group or the other, an attempt was made to achieve combinations of both same and mixed ability. Moreover, within the mixed ability groups, some were two-level and some three. Gender was not
Table I. Gender and ability mix of experimental groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Ability and gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H(m), H(m), H(m), H(m)</td>
</tr>
<tr>
<td>2</td>
<td>A(m), A(m), A(f), A(f)</td>
</tr>
<tr>
<td>3</td>
<td>B(m), B(f), B(f), B(f)</td>
</tr>
<tr>
<td>4</td>
<td>H(f), H(f), A(f), A(m)</td>
</tr>
<tr>
<td>5</td>
<td>A(f), A(m), B(f), B(m)</td>
</tr>
<tr>
<td>6</td>
<td>H(m), A(f), A(f), A(f)</td>
</tr>
<tr>
<td>7</td>
<td>A(f), A(f), A(m), B(m)</td>
</tr>
<tr>
<td>8</td>
<td>H(f), H(f), H(f), H(f)</td>
</tr>
<tr>
<td>9</td>
<td>A(f), A(f), A(f), A(m)</td>
</tr>
<tr>
<td>10</td>
<td>B(f), B(f), B(f), B(m)</td>
</tr>
<tr>
<td>11</td>
<td>H(f), H(f), A(f), A(m)</td>
</tr>
<tr>
<td>12</td>
<td>H(m), A(f), A(m), B(m)</td>
</tr>
<tr>
<td>13</td>
<td>H(f), H(f), A(m), B(m)</td>
</tr>
</tbody>
</table>

H, above average; A, average; B, below average; M, male; F, female.

considered in grouping but it was noted. The final range of ability and gender combinations is shown in Table I.

The first time the groups met to work together, they were given a practice problem and instructions to work through this with one of the researchers. Incorporated within this session was a short lesson given to all groups. The lesson was based around reading the problem, and included exploring with the children what constituted good stopping places when there was a long problem to read, e.g. full stops, question marks and the end of paragraphs. Attention was also paid to identification of clue words which might indicate what type of calculation was required, e.g. 'more', depending on the context, might indicate that an addition was appropriate. This short lesson became a 'reminder' lesson and was reiterated each time the groups met to carry out problem-solving. Following this, the groups were then given the problem-solving instructions and their problem for the day to work through. Throughout the entire procedure, an adult was on hand should the group need assistance. On completion of the group problem-solving session, the children were put into pairs and worked through the pairs sessions following the problem-solving guidelines as described in the Materials section. As with the groups, an adult sat with the pairs on hand to give assistance if required. This assistance usually took the form of help with the task requirements, e.g. making sure that the children followed the instructions to the letter. This type of help was withdrawn as the children became accustomed to the instructions themselves. Very rarely, help was requested with the actual problem, e.g. explaining why a specific type of calculation was carried out. This type of help was only provided if the children were absolutely stuck and could progress no further through the task without adult intervention.
The groups/pairs cycle was carried out three times over a 3-week period. Two days per week were devoted to the group sessions and two to the pairs, 4 days in all. All sessions were videotaped. While the experimental classes were engaged in this activity, the control classes also solved the same problems. The researchers played the role of 'teacher' standing in front of the whole class, initially delivering the same lesson as the groups had received at the beginning of the task. The children were then requested to solve the problems on their own sitting at their tables, writing down any calculations they did. The researcher was on hand if any help was required with reading the problem. The children were encouraged to try and work it out for themselves and to omit it should they be unable to work out an answer. This was in common with the group task when children were working on the problems on their own—they were required not to ask for help at that stage and to miss it out should they be unable to work out and answer, as they would have the opportunity to work together and help each other get an answer later.

Post-test. Immediately following the intervention, all the children completed a post-test following the same procedure as the pre-test, that is, the whole class followed by individual interview. Apart from the differences in problem content, the only difference from the pre-test was that the children who had worked in groups/pairs were asked how they felt about the experience, what they had learned and why. All the interviews for both pre- and post-test were subsequently transcribed.

Coding

Measures were taken of the children's pre- and post-test performance and of their dialogue during the group sessions.

Pre- and post-test coding. Three measures were applied to the pre- and post-test data. First, note was made of the total number of problems out of 20 (the 'score') which each child successfully solved. In addition, each child's pre-test score was subtracted from their post-test score to give a figure representing the score change across the two tests. Second, the children's responses to the interview questions regarding how they went about finding the solution to two of the problems were summarised and referred to as the 'strategies' that the children used. The strategies were arranged hierarchically in order of successful implementation, the least effective rated as '0', the most successful as '4' as shown in Table II.

The children were awarded a strategy rating for the four potential solutions within the two problems they were asked about, resulting in values between 0 and 16 for both pre- and post-test. The pre-test strategy values were subtracted from the post-test providing measures of strategy change. Third, working again from the pre- and post-test interviews, the 'execution' performance was defined


TABLE II. Hierarchical coding scheme of problem solving strategies

<table>
<thead>
<tr>
<th>Rating</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Calculation not attempted</td>
</tr>
<tr>
<td>1</td>
<td>Can’t explain how solution reached</td>
</tr>
<tr>
<td>2</td>
<td>Incorrect reference to question</td>
</tr>
<tr>
<td>3</td>
<td>Re-reads question</td>
</tr>
<tr>
<td>4</td>
<td>Correct reference to question</td>
</tr>
</tbody>
</table>

as the appropriateness of the calculation and solution. Children could do one of three things, not attempt a solution, make an incorrect attempt or make a correct attempt, these being coded hierarchically as 0, 1 and 2. This resulted in a value for both pre- and post-test execution within the range 0–8, representative of two two-answer problems from each test. The pre-test execution values were also subtracted from the post-test, resulting in measures of execution change.

**Inter-rater reliabilities.** The pre- and post-test interview responses from 41 subjects were independently coded with reference to strategies and execution by two judges who were blind to condition. Inter-judge agreement was 88% for strategy change (Kappa = 0.81, \( p < 0.0001 \)) and 82% for execution change (Kappa = 0.65, \( p < 0.0001 \)). Reliability checks were not carried out on score change since it was an objective measure.

**Dialogue coding.** When the children were working together in groups, they could be involved in up to four broadly defined categories of activity as follows:

(a) listening;
(b) giving appropriate explanations;
(c) giving inappropriate explanations; and
(d) no discussion.

Each group session was rated from the video recording, with each child’s contribution noted as per the categories above. This was a straightforward process as throughout the session, the children were remarkably consistent in their contributions.

**Results**

Firstly, the two experimental classes were compared for their pre-test scores, strategies and execution. There were no significant differences between the
classes on any of the measures so they were combined and treated as one group. The experimental class’s pre-test performance was also compared with the control group’s, again showing no significant differences between groups. This was taken as justification for using change in scores for analyses as any differences would then be due to the intervention.

**Score Change, Strategy Change, Execution Change: results**

ANOVA’s were used to look at the effects of condition (experimental vs control), gender (male vs female) and ability (below vs average vs above average as classified by teachers) on change in score, strategy and execution. There was an approaching significant main effect of condition (experimental/control), upon score change, \((F = 3.61, \text{df} = 1.67, p < 0.06)\), Table III showing the means for each condition and clearly indicating that the experimental classes were getting better overall scores than the control.

Looking at the children’s strategy change, while there were no significant main effects for gender or condition, we found a main effect of ability \((F = 5.49, \text{df} = 2.60, p < 0.007)\). Breaking this down by ability levels as shown in Table IV, the children in the below-average category progressed the most between pre- and post-test. Interestingly, the children in the above-average category appeared to be regressing, using less successful strategies.

Although there was no main effect for gender or condition, there was a 2-way interaction for strategy change by gender and condition \((F = 6.45,\)
df = 1.60, p < 0.01). Looking at the mean strategy change score in Table V, while all the children in the experimental condition showed improved levels of strategic understanding when working in groups, the girls improved twice as much as the boys. Girls working on their own deteriorate in levels of understanding in comparison with the children in the experimental group.

Turning to execution change, there was a main effect of condition \( (F = 4.38, \text{df} = 1.68, p < 0.04) \), with no other significant main effects or interactions. Looking at the means shown in Table VI above, both the control and experimental groups seem to be deteriorating in the ability to execute the correct calculation, but the experimental to a lesser degree.

Moving away from the data collected from children while they were working as individuals during the pre- and post-tests, we also examined the children’s dialogue when working collaboratively in groups. As explained above, this was collapsed into four categories, listening, giving appropriate explanations, giving inappropriate explanations and no discussion. The frequency of usage of each category was calculated showing first for positive score, strategy and execution change and then negative. Table VII shows the percentages of pupils who had either positive or negative score change, and the types of strategies they used. The children were remarkably consistent in general using one strategy, e.g. listening to the group discussion. Where a child used more than one strategy it was the predominant strategy which was coded. As can be seen in Table VII, positive scores change is associated with giving appropriate explanations and negative with no discussion.

For positive and negative strategy change shown in Table VIII, positive changes in strategy use are associated with listening and negative with giving inappropriate explanations or with no discussion occurring in the group.
TABLE VII. Use of dialogue categories by pupils who show improvement and deterioration in pre-to post-test scores

<table>
<thead>
<tr>
<th>Score change</th>
<th>Dialogue used by those with a positive pre- to post-test score change (%)</th>
<th>Dialogue used by those with a negative pre- to post-test score change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listen</td>
<td>24.13</td>
<td>8.1</td>
</tr>
<tr>
<td>Appropriate explanations</td>
<td>62.06</td>
<td>21.62</td>
</tr>
<tr>
<td>Inappropriate explanations</td>
<td>0</td>
<td>24.32</td>
</tr>
<tr>
<td>No discussion</td>
<td>13.79</td>
<td>45.94</td>
</tr>
</tbody>
</table>

TABLE VIII. Use of dialogue categories by pupils who show improvement and deterioration in pre-to post-test problem solving strategy use

<table>
<thead>
<tr>
<th>Strategy change</th>
<th>Dialogue used by those with a positive pre- to post-test score change (%)</th>
<th>Dialogue used by those with a negative pre- to post-test score change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listen</td>
<td>43.33</td>
<td>25.00</td>
</tr>
<tr>
<td>Appropriate explanations</td>
<td>26.66</td>
<td>10.00</td>
</tr>
<tr>
<td>Inappropriate explanations</td>
<td>6.66</td>
<td>30.00</td>
</tr>
<tr>
<td>No discussion</td>
<td>23.33</td>
<td>30.00</td>
</tr>
</tbody>
</table>

For execution change, Table IX shows that positive changes in execution are associated with giving appropriate explanations to others, while negative changes are not particularly associated with any one particular type of dialogue behaviour.

Looking at those who specifically showed improvement in strategy use and breaking this down by levels of ability, it can be seen from the summary of data

TABLE IX. Use of dialogue categories by pupils who show improvement and deterioration in pre-to post-test execution

<table>
<thead>
<tr>
<th>Execution change</th>
<th>Dialogue used by those with a positive pre- to post-test score change (%)</th>
<th>Dialogue used by those with a negative pre- to post-test score change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listen</td>
<td>12.82</td>
<td>20.68</td>
</tr>
<tr>
<td>Appropriate explanations</td>
<td>46.15</td>
<td>27.58</td>
</tr>
<tr>
<td>Inappropriate explanations</td>
<td>12.82</td>
<td>17.24</td>
</tr>
<tr>
<td>No discussion</td>
<td>28.20</td>
<td>34.48</td>
</tr>
</tbody>
</table>
shown in Table X that those in the below-average category spent most of their time listening in comparison with other ability levels who had more of a spread of activities. This is of interest, as it was the below-average children who made most improvement.

Turning to when the children were working in pairs, dialogue was analysed looking at the impact of:

(a) acting as a ‘pupil’ as opposed to a ‘teacher’;
(b) the relative ability levels of the pupils and teachers; and
(c) whether or not there was improvement pre- to post-test in score, strategy and execution.

With reference to pre- to post-test score change there was one significant result regarding those acting as teachers. Of those who had a deterioration in pre- to post-test score and were acting as teachers, the below-average children asked for more help of the researcher when explaining to their pupil how to do the sum in comparison with the above average children \((F = 3.96, \ df = 2.54, \ p < 0.02, \ \text{follow-up Scheffé test with significance level 0.05})\). There were no other significant results concerning the pairs data.

When interviewed after the post-test, the children in the experimental classes \((n = 45)\) were also asked a number of questions about working in groups and pairs. Initially, they were asked if they enjoyed the procedure more than, the same or not as much as number in the classroom. Seventy-five per cent enjoyed number in groups/pairs more than in the classroom, 17% enjoyed it the same as number in the classroom, with 6% preferring number in the classroom. They were then asked if they preferred to work in groups or pairs or had no particular preference. Fifty-three per cent preferred to work in groups, 31% preferred to work in pairs, with 16% expressing no particular preference. As a follow-up to those questions, those who expressed a preference were asked why, their reasons detailed below.
Reasons Why Children Enjoyed Group/Pair Work

- You get to work together: 54.76%
- It's more fun: 16.66%
- Friends don't get cross if you are wrong: 4.76%
- You know how others are doing: 4.76%
- Because we need help: 4.76%
- You use your brain more: 4.76%
- You use words which are understandable: 2.38%
- Feel more confident: 2.38%
- You find out different things: 2.38%
- You get to be the teacher: 2.38%

Discussion

This teaching intervention was designed to examine three main issues, namely (a) the effect of working collaboratively to solve problems as opposed to working individually in the traditional manner, (b) the extent to which the benefits of group collaboration were dependent on gender and ability, and (c) the effect, if any, of the generation of good explanations given within groups. To reiterate the results as shown in Table III, there was an approaching significant main effect of condition upon score change, the experimental children's performance improving in comparison with the control children's whose performance was deteriorating. So it is possible to state that, in this case, the trend is for improvements in performance when working collaboratively in groups. The children's score change correlated significantly with their execution change (\( \rho = 0.38, p < 0.001 \)), confirming that those who knew which calculation to do and who could execute it correctly were getting better scores, the converse also being true. While both experimental and control groups were deteriorating in their pre- to post-test ability to execute the correct calculation, the control group was significantly poorer.

Turning to the errors of execution which children made while carrying out calculations, they can be described as commonplace. They tended to be of two types, the first being the wrong choice of calculation to carry out, e.g. an addition instead of a subtraction, and the second demonstrating the children's lack of understanding of the mechanics of calculations, ranging from the basic error of trying to subtract a bigger number from a smaller one (with written calculations), to errors relating to rules for exchanging and carrying forward. Often such errors were spotted by the children when asked to talk through a calculation during the pre- and post-test interviews. Remembering that overall, both the experimental and control groups deteriorated in their ability to execute correct calculations, but demonstrated relative improvements in levels of understanding, it could be suggested that the mechanics of mathematics might well be perceived and treated as separate from internal representations of the problem,
thus supporting research evidence for ongoing parallel development of informal and formal mathematical systems as suggested by Siegler (1989). Following this line of argument, this would also suggest that our intervention failed to marry the two systems together.

Of relevance to this argument were the children's use of problem-solving strategies. Couched in terms of domain-general and domain-specific knowledge, Siegler discusses how both types of knowledge work together for the individual to produce a variety of strategies which can be applied to the same type of problem. In this case, domain-general knowledge can be seen as the mechanics of mathematics, domain-specific the internal representation of the problem. He argues that children perform consistently better when using back-up strategies to solve a problem e.g. using fingers when adding/subtracting, than when using retrieval (procedural information). Using our approach, children were encouraged to discuss and explore the conceptual component of problem-solving and thus increase their understanding at this level. They were then faced with the task of recording their decision, all the children choosing to do this in the traditional way of carrying out the calculation as a written sum, thus invoking the retrieval strategy in order to do this. So, while our approach facilitates a better conceptual understanding of the problem and in turn application of more appropriate problem solving strategies (domain-specific knowledge), it does nothing to foster better understanding about the procedural aspect of problem-solving (domain-general knowledge).

Alternatively, there could be a problem with the execution measure as the above rationale does not explain why children's understanding improves and their scores improve but their ability to execute the correct calculation deteriorates over both experimental and control conditions. The children were encouraged to work out the solutions to problems in any way they preferred, whether it be by traditional written sum, using fingers, or mentally. We then asked the children at their interviews following both pre- and post-tests, how they worked out the answer. This may in fact have been biased against those who worked out the answer in their own idiosyncratic way which they might not be able to describe accurately. This would allow for children to be seeming to deteriorating in execution but still improving in score and understanding.

Staying with strategic change, Siegler (1989) also suggests that children can and do use several different strategies when problem-solving. Examining the children's use of strategies when problem-solving on the pre- and post-tests, the first thing was that these particular children used a limited variety, usually employing single strategies, occasionally two at the same time. These included correct and incorrect reference to clue words, correct and incorrect reference to the question in general, use of a specific strategy first, e.g. 'I did an adding because I always try adding and if it doesn't look right I'll try a taking away', re-reading the question and being unable to explain why.

This issue of strategy choice was also of interest when the children were working in groups of different ability and gender. Having identified strategies in use when the children were working individually, we examined at the descriptive
level, the strategies in use when the children were in their groups. On examination of pre- to post-test strategy change by ability (Table IV), those children rated as below average made the most improvement followed by those rated average. This would appear to contradict much of the earlier research reported by Webb (1992) and Peterson et al. (1984), who found a tendency for improvements in average and above-average children. Given that listening was associated with improvement in strategy use and that listening was the main strategy used by below-average children, this would suggest that placing below-average children in groups where they can listen to the reasoning of their more able peers is more beneficial than placing them in groups where they don’t have access to such information.

In contrast to the studies mentioned above, our results show the above-average children using less successful strategies pre- to post-test. One explanation for this can be found in the work of Greenfield et al. (1972), who see the departure from using previously reliable strategies as indicative of changes in internal forms of organisation. As experience accrues, children have more skills to draw from when faced with problem-solving situations, thus a rule which was perceived as adequate beforehand might be called into question as the child perceives different facets of the problem and has to sort out which components to consider for success and what to reject to avoid failure. This involves attempting to perceive what is salient within a particular problem, thus accounting for departures from previously reliable strategies. The implication is, of course, that given time to rearrange their internal representations of the problem-solving situation, these children would once again use successful and reliable strategies.

Support for this line of reasoning is further found in the work of Howe et al. (1992). After working collaboratively in groups on science problems, the children completed two post-tests, one immediate, that is within 24 hours of the collaborative task, and one a few weeks later. Performance on the immediate post-test was indistinguishable from performance on the pre-test while performance on the delayed later post-test was markedly better. In another study (Tolmie et al., 1993), again there were two post-tests, one now 4 weeks after the group tasks, the other 11 weeks. Change from the pre-test to 4-week post-test was positive, but it was outstripped by change from the 4-week to 11-week post-test. Such results led Howe and colleagues to suggest that the bulk of group-based learning takes place after the group tasks. Accepting this and also given Vygotskian perspectives on the internalisation of register–problem linkages as explained earlier, it would appear unwise to conclude from the present results that group-based learning is ineffective in mathematics for above-average children. Further research is required which takes full account of the need for internal reorganisation.

Interestingly, with regard to improvements in strategic understanding, there was a gender effect where the girls in the experimental groups outperformed all the other children. Given the size of the sample for this study, we cannot draw any firm conclusions about this result, but it is still worthy of comment. It
cannot be said that this improvement was due to there being an unequal distribution of below-average girls who were benefiting from listening, as there were eight below-average girls to seven below-average boys. Instead, it might be suggested that such improvement be attributed to the positive effects from girls acting as facilitators when in the group situation. Jenkins & Cheshire (1990) describe girls’ conversational styles as facilitating competent and effective contributions from other group members, encouraging other group members to participate, in addition to being good listeners. In this context, given that girls may be more involved in verbalising ideas, raising questions and listening to responses, their higher levels of performance could be attributed to these activities, which in turn would have the effect of imposing meaning on the material for that individual, thus enhancing their internal representation of the problem-solving situation. Support for this conjecture comes from the teachers of the classes involved in the study, who commented in general on the efficacy of groups which had girls in them, especially with this age group.

As far as working in pairs was concerned, based on the dialogue analyses, there appeared not to be much of consequence happening here. The one significant result found was with reference to those children acting as teachers and who had a deterioration in pre- to post-test scores. The children which fell into this category and were in addition of below-average ability, asked the researcher for more help than any other group when trying to explain to their pupil what to do. In all events, this would seem a perfectly understandable and obvious outcome as they were the ones who had the least understanding of the process or the least confidence in explaining to another child how to do it. In general, it may be hypothesised that the children were using the pairs sessions as a time to practice the skills which they were introduced to in the group sessions; however, this would require further research to substantiate.

Finally, the children were also asked if they enjoyed what they had been doing and their preference if any for the groups and pair. Out of 45 children interviewed at this stage, 75% said they enjoyed it more that the traditional classroom approach to number, 53% preferring to work in groups, 31% in pairs, the remainder having no particular preferences. As working collaboratively in a situation where individuals are required to share methods can be perceived as a threatening situation, and in fact one which is very often actively discouraged in the classroom, it was heartening to see so many children enjoying what they were doing, and perhaps enlightening with reference to how children prefer to learn.

Summary

In summary, this study indicates that working in groups can facilitate cognitive gain in the mathematical problem-solving context, with girls in groups showing more relative improvement. Children rated as below average in ability appear to benefit the most from this approach as it affords them the opportunity to listen to their peers discussing problem-solving in a language they can understand.
more readily. However, this did not translate to improvements in execution, suggesting that there is the need to find some way of tackling the ongoing problem of linking the conceptual and procedural components of mathematical problem-solving.

Acknowledgements

We would like to express our thanks to the University of Strathclyde Research and Consultancy Unit for funding to carry out this study and to Jenny Low for help with data collection. Our thanks are also extended to the children and staff of St Martha's and St Helen's Primary Schools, Glasgow, for their enthusiastic participation in the project.

NOTES

[1] Address for correspondence: Dr Pamela Davenport (née Smith), The Anna Freud Centre, 21 Maresfield Gardens, Hampstead, London NW3, UK.

REFERENCES


Appendix 1: Examples of multiple-answer problems

Example 1
Fiona had to take 2 trains to get from Glasgow to Elgin. The first train from Glasgow to Aberdeen took 2 hours 52 minutes and the train from Aberdeen to Elgin took 1 hour 39 minutes. How much time did Fiona spend on trains altogether?

She then had to get a bus to her friend’s house. Going by her new watch, she had to wait for the bus outside Elgin Railway Station for 7 minutes 23 seconds. The bus journey lasted 15 minutes 39 seconds. How long did it take Fiona to get from Elgin Railway Station to her friend’s house? How long did Fiona spend travelling altogether?

Example 2
Sharon and Rose are playing in the playground. They are counting how many giant steps it takes to get from one end of the playground to the other. They make sure that their giant steps are
exactly the same size. Sharon has taken 26 giant steps but has not reached the end of the playground yet. Rose is at the end of the playground and took 43 giant steps to get there. How many more steps does Sharon have to take to get to the end of the playground?

Jenny joins in the game. Her steps are slightly smaller than Sharon’s and Rose’s. To get from one end of the playground to the other, Jenny takes 51 giant steps. How many more steps than Sharon and Rose does Jenny take?

Appendix 2: Examples of Carpenter & Moser’s (1982) six structures presented as addition problems for time

Joining
The class were running a race. Susan took 1 minute 27 seconds to finish. Her friend Janice took 19 seconds longer to finish. How long did Janice take to run the race?

Separating
Peter took part in a cross-country race. He ran for 25 minutes 39 seconds and he spent 3 minutes 50 seconds walking during the race. What was his finishing time?

Part Part–Whole
Bobby got a new watch with a ‘second-hand’ for his birthday and decided to time how long it took him to get to school. He waited for the bus for 7 minutes 23 seconds. The journey lasted 15 minutes 39 seconds. How long did it take him to get to school altogether?

Comparison
On Monday it took Linda 7 minutes and 35 seconds to walk home from school. On Tuesday she took 2 minutes and 46 seconds longer. How long did it take her to walk home on Tuesday?

Equalising Add-on
Mum set the timer on the microwave for 3 minutes 25 seconds to cook a pizza. This was less than the cooking time printed on the packet so it was not properly cooked. Mum set the timer for another 1 minute 50 seconds which made the total time the pizza was cooked the same as it said on the packet. How long did it say on the packet to cook the pizza?

Equalising Take-away
The children had already spent 20 minutes 40 seconds doing different activities in the gym. The gym teacher said time was running out and the class still had one more activity to do. They usually spent 25 minutes in the gym. How much time could they spend at their last activity so that they could leave when they were meant to leave?

Appendix 3: Group problem-solving instructions

You are going to work together in a group to solve some number problems. Sometimes you will work on your own and sometimes you will talk to each other and work together.

(1) Do you all have the sheets with Problem 6 in front of you?
STOP READING AND CHECK
(2) Get one person in the group to carefully read the problem out loud to everyone else. 
   STOP READING AND LISTEN
(3) Now work on your own and find the answers to the sums. 
   Let each other know when you are finished. 
   Re-read the problem first of all. 
   STOP READING AND WORK
(4) Has everyone finished or done all that they can do? 
   STOP READING AND CHECK
(5) Now take it in turns to show the rest of the group your answers, starting with whoever 
   is in charge of the yellow sheet. 
   STOP READING AND TALK
(6) Did you get the same answers? 
   If NO, go to number 7 (GO TO NUMBER 7 NOW) 
   If YES, write the sums and the answers on the yellow sheet. 
   STOP READING AND WRITE 
   Now go to number 9.
(7) If you did not all get the same answers, talk to each other and decide what the sums 
   and the answers should be. Tell each other why you chose to do this type of sum. 
   STOP READING AND TALK
(8) Write the sums and the answers on the yellow sheet. 
   STOP READING AND WRITE
(9) Check the answer card to see if you are correct. 
   STOP READING AND CHECK
(10) If you are correct, let the teacher know you are finished. 
(11) If you are wrong, try and work out what kind of sums you should be doing again. 
   STOP READING AND WORK
(12) When you have done this, go back to number 8.

Appendix 4: Pairs instructions

You are the teacher. 
First of all, re-read the problem to remind yourself what it was all about. 
STOP AND READ 
Turn the problem over and tell your partner what the problem is about. 
STOP AND SPEAK 
Give the problem to your partner and ask him or her to read it. Tell them to let you know when 
they have finished reading it. 
STOP AND WAIT 
Now you are going to ask your partner some questions. Turn to the next page. 
ASK YOUR PARTNER

(1) What are you asked to find out? 
IS YOUR PARTNER CORRECT? 
IF YES, READ OUT QUESTION 2. 
IF NO, TELL THEM WHAT THEY HAVE TO FIND OUT.

(2) What kind of sum would you do to find that out? 
IS YOUR PARTNER CORRECT? 
IF YES, READ OUT QUESTION 3. 
IF NO, EXPLAIN WHAT THEY HAVE TO DO.
(3) Why would you do that type of sum?
IS YOUR PARTNER CORRECT?
IF YES, GO TO QUESTION 5.
IF NO, EXPLAIN WHY THEY WOULD DO THAT SUM THEN GO TO QUESTION 4.

(4) Do you understand why we do that sum?
IF THEY ARE SURE, GO ON TO QUESTION 5.
IF NOT, EXPLAIN AGAIN THEN GO TO QUESTION 5.

(5) What numbers will you use?
IS YOUR PARTNER CORRECT?
IF YES, GO ON TO QUESTION 6.
IF NO, TELL THEM WHICH NUMBERS TO USE.

(6) How will you write the sum down?
IS YOUR PARTNER CORRECT?
IF YES, TELL THEM TO WRITE THE SUM DOWN.
IF NO, EXPLAIN WHAT THEY SHOULD BE DOING AND LET THEM WRITE IT DOWN.

Check and see if they are correct.
If they are correct, tell them well done.
If they are wrong, help them until they get it right.
When you have finished, let the teacher know.

Appendix 5
NUMERACY PROJECT—INTERVIEW SCHEDULE

NAME ________________________________

Introduction
1. Let’s choose one problem. I’ll read it through and you can listen. Then I’m going to ask you how you did it.
It doesn’t matter if you didn’t get it right. What’s important is how you did it.

Read problem
2. What kind of sum did you do?
Option: Let’s look at the sum and see if you can remember.
Let’s look at the answer and see if you can remember.

3. Why did you do adding/take away? Were there any clues in the story?
If strategy is to add/subtract first, explore why.

4. I see you’ve done a sum here. Can you talk through what you did?
There’s only an answer here. Did you do it in your head?
Can you talk through what you did?

5. Probe for strategy
How did you add/take away?
Did you use your fingers, ruler, know the answer?
Which side did you start with?
Which number went on top? Why? (if subtracting)
6. **Problem specific queries**

Can you remember the rule for ...?
Was the sum easy/hard ...?

7. **Confidence**

I want to find out how sure you are you’ve got the right answer. Look at these faces.

   (1) isn’t sure at all
   (2) could be right, could be wrong
   (3) pretty sure it’s the right answer
   (4) knows it’s the right answer.

Point to a face that shows how you feel about your answer.