Mathematical Problem Solving in Groups:
Are Two Heads Better Than One?

Kaye Stacey
Institute of Education
University of Melbourne

Author's address:

Dr Kaye Stacey
School of Science and Mathematics Education
Institute of Education
University of Melbourne
Parkville 3052
Victoria, Australia

Problem Solving in Groups:
Are Two Heads Better Than One?[*]

It seems natural to expect that students solving problems in groups would get better results than students solving the same problems individually. However, in a written test of problem solving, group performance (pairs and triples) was not better than the individual performance. To investigate why this happened, protocols of students solving problems in groups were carefully examined. In much of the group discussion, students constantly passed over right ideas in favour of simpler, but erroneous, ideas. Some groups checked effectively, but for others the checking was either ineffective or non-existent. Groups had plenty of ideas but the choices that they made between the ideas were often not considered at all carefully. Teachers using group work for problem solving need to be aware of these aspects of group behavior. Group problem solving experiences where simple answers are strongly favored will not enhance students' ability to solve problems.

---

[*] This research was supported by a grant from the Australian Research Council. The opinions expressed here are those of the author and not necessarily those of the Australian Research Council. I would like to thank Mollie MacGregor for her assistance in collecting and analysing these video-tapes.

INTRODUCTION

This paper reports some results on problem solving in groups which have implications for classroom teaching of problem solving, for understanding the fundamental processes of mathematical problem solving and for assessing the place of group work in teaching Mathematics.

Co-operative group work is an increasingly popular strategy for teaching mathematics and is very frequently associated with the teaching of problem solving. In Australia, where this study was undertaken, group work is recommended by writers and authorities such as Lovitt and Clarke (1988), Victorian Curriculum and Assessment Board (1989), Stacey and Groves (1985), Victorian Ministry of Education (1988), Richards and Trotter (1989). The reasons given for the use of group work in problem solving include the opportunity for pooling of ideas, the natural need that arises to explain and express ideas clearly, and the reduction in anxiety for tackling something hard. The association of group work with problem solving is also strongly made elsewhere in the world. Good, Grouws and Mason (1990), for example, report that "problem solving" was the most frequent answer given by teachers in a sample of 1509 U.S. elementary school teachers to the question "For what mathematical topics, if any, is it especially important to use small groups?"

The observations which prompted this paper were initially made by accident. As part of a larger project, we were interested to monitor the progress of whole classes in mathematical problem solving. For the reasons listed above, it was expected that group work would lead to better answers to non-routine mathematical problem solving questions. Such a result had been reported for other types of problem solving, for example by Mathews, Lane, Reber, Buco, Chaney and Erffmeyer (1982) who compared group and individual solutions to the NASA Moon Survival Problem and Egerbladh and Sjodin (1986) who used word puzzles.

Surprisingly, data which is presented below, from a large sample of Year 9 students, showed that group performance was not better than individual performance on a written test of mathematical problem solving. A similar finding has been reported by Treilibs (1979), who found that group performance of able British A-level students was not superior to individual performance in mathematical modelling. He speculated that the effect was due to the inexperience of students in modelling and in group work, as well as the large number of diversions of the argument that were typical of the modelling observed. He concluded that teachers need to take positive action if they are to create co-operative working groups and that they should avoid a total commitment to group work in modelling. Egerbladh and Sjodin (1986) who used two types of word puzzles reported that group productivity (whether the groups or individuals scored more highly) interacted significantly with the accessibility of the task, that is whether the task was accessible to a large or small extent to all members of the group. They commented that if a correct proposal

> "is only understood by this member, the educational interaction should be inhibited. If a given answer is immediately understood by all members, the same interaction should be stimulated." (p19)

Rather than being particularly suitable for group work as the pedagogical literature suggests, the relative inaccessibility of mathematical problems may make this a difficult are for groups to work productively.

In order to explore further the processes involved in group work for mathematical problem solving, students were video-taped whilst working on a range of problems. This paper reports an analysis of the protocols for three structurally similar problems, where it is possible to trace a small number of ideas through the group discussion.

The reasons for average group performance not being superior to average individual performance are important pointers to understanding both students' problem solving behaviour and some of the dangers of exclusively using group work for mathematical problem solving.

PRELIMINARY STUDY: WRITTEN TESTING

A 45-minute written test of problem solving (developed by the present author) was given to all Year 9 students (average age 14 years) at two secondary schools. Two of the four classes at each school were randomly allocated to take the test under the individual work format and the other two classes at each school used the group work format (mostly pairs with some triples). The teachers had not observed any differences in ability between the classes. The class teachers supervised the test, allocating students to the groups in whatever way they normally would, usually friendship groups.

The test contained six items, each of which was presented in a "real-world" context and was non-routine and unfamiliar to the students at least in the sense that it did not look like a typical text-book exercise. One of the items, for example, required students to estimate the number of microwave ovens required to make one cubic metre of popcorn in a school day, given various data. Another item was similar in mathematical structure to the three problems which are used in the protocol analysis discussed below in this paper but it involved a quadratic (rather than linear) pattern. The questions asked did have right answers; several also required interpretation of a mathematical result. The marking scheme used was developed by a teacher as a typical scheme for a test, alloting marks for complete or partial correct answers.

At both schools, the group performance was slightly worse than the individual performance. Analysis of the written scripts from one school made us suspect that the testing in groups may not have been taken as seriously as the individual work. At the other, a Catholic girl's school, all classes worked very seriously on the task. The results for this school, reported in Table 1, showed that students working individually performed slightly better, although a t-test showed that the difference was not statistically significant. (Note that the finding that the overall mark did not improve does not imply that particular individuals were not advantaged by the group work format.)

The written scripts were examined to see if there was an obvious explanation for the lack of advantage for groups. It was thought that because groups would need to reach agreement, they may have spent more time discussing each question and thus attempted less questions. This was not the case, as the average percentage of questions answered (counting the 18 answer-giving parts of the total test separately) was less for the individuals (see Table 1). It seems then that the group work may have increased confidence and perhaps the number of ideas considered, but this did not translate into more correct answers.

Table 1

Results on Written Problem Solving Test

	Number of students	Mean score	Standard deviation	Questions answered
Groups	37	9.3	5.05	73%
Individuals	43	9.5	4.45	63%

COLLECTION OF PROTOCOLS

To explore why the group problem solving had no advantage over individual work, video-taped protocols of seven single-sex groups of three students were examined. The students were selected by their teachers as groups of above-average ability students who would work well together. The girls were from a private girls' school and the boys were from a co-educational private school. As had been recommended by Derek Foxman of the Assessment of Performance Unit (personal communication) the observer provided the problems and encouraged students to discuss but otherwise took no part in the discussion, feigning lack of interest. In order to promote co-operative working, only one pencil was provided for each group. The students were a group of girls (here labelled G9) and a group of boys (B9) from Year 9 (age 14 years), two groups of girls (G8a and G8b) and two groups of boys (B8a and B8b) from Year 8 (age 13 years) and one group of girls (G7) from Year 7 (age 12 years).

There were three problems which were to be analysed, interspersed amongst a larger batch of problems that the groups attempted during a 45-minute session. These three problems (*Ladders* (see Figure 1), *Sequence* (see Figure 2) and *Christmas Trees* (see Stacey, 1989)) have the same mathematical structure, all being simplified versions of one of the problems on the written test, although presented in different contexts. Previous work (Stacey, 1989) had established that there are a small number of very prevalent errors made by students with problems of this type. The proposal, rejection and acceptance of a small number of ideas could be therefore by traced in the protocols of attempts at 42 questions (attempts by seven groups at three problems, each asking two questions). In fact all groups answered the first question in *Sequence* by successive addition of 6 without any discussion, so this question is not included in the analysis below and two groups only had time for two of the three problems. This left protocols for a total of 31 questions for analysis.

Ladders , *Sequence* and *Christmas Trees* all have similar mathematical structures. Let us denote the number of matches needed for a ladder with r rungs as $M(r)$, and the number of lights on a Christmas Tree of size s as $X(s)$. In fact $M(r) = 3r + 2$ and $X(s) = 4s - 1$. Denoting the nth term of the number sequence as $S(n)$, we have $S(n) = 6n-2$. The first relationship that students notice and try to use is the constant difference property, that $M(r+1) = M(r)+3$, $S(n+1) = S(n)+6$ and $X(s+1) = X(s)+4$. This can be used for the first question asked in each problem, but is impractical for the second questions, which require students to find and use (implicitly or explicitly) a relationship of the form $y = mx + c$ (with $c \neq 0$). Previous work (Stacey, 1989) has established the common wrong solutions for these problems: in particular, over 40% of a sample of Year 7 and 8 students used the "whole-object" method to find $M(1000)$, $S(100)$ and $X(100)$. This "whole-object" method uses false relationships of direct proportion such as $M(20) = 4 \times M(5) = 68$, $S(100) = 10 \times S(10) = 580$ and $X(100) = 5 \times X(20) = 395$. It is called the "whole-object" method because students apparently assume that a ladder with mn rungs is actually made up of m complete copies of a ladder with n rungs.

Figure 1: Ladders
to be inserted about here

Sequence

Fill in the blanks in this number sequence which continues on and on in the same pattern

4, 10, 16, 22, 28, , , , , , ,

What is the 10th term of the sequence?
What is the 100th term of the sequence?

Figure 2. Sequence

ANALYSIS OF PROTOCOLS

Table 2 shows the correct answers and the answers given by the seven groups of students interviewed and the number of episodes of checking which could be identified in the protocols. The Year 7 and Year 8 girls groups exhibited the most checking and also discussed the problems for the longest times. The two year 9 groups (G9 and B9) and the Year 7 group (G7) answered all their questions correctly. B8a would have answered all questions correctly but made two "slips", one arithmetic and one logical, just as the final answers were being recorded and the group was preparing to move on to the next problem. The remaining ten wrong answers in Table 2 are all due to various implementations of the seductively simple whole-object method.

The decision to use or not to use this simple wrong method was therefore critical to the success of the groups, so the instances when these decisions were made were examined. As shown in Table 3, G9 and B9 made no whole-object conjectures, whilst groups G7 and B8a were tempted but successfully resisted using it. Groups G8b, G8a and B8b fell into the trap, to varying degrees. Table 3 also shows the number of problems where at least one group member has proposed a correct general method sufficiently clearly that, in the opinion of the author, were it to be adopted it could have been readily used to solve the second question for the problem, i.e. to find M(1000), S(100) or X(100). For example, in the transcript of Group G8a below, Beth's first proposal offers a clear method for finding M(20) which is readily generalisable to M(1000). However, G8a did not adopt it.

Table 3 shows that all groups except B8b proposed ideas that could have been used to solve the problems correctly. Conversely all groups except the two year 9 groups proposed the wrong whole -object method at some stage. It seems not to be getting the ideas that is difficult, but choosing which ideas to implement. The different behavior exhibited by the groups is discussed below with detailed reference to the protocols.

Table 2

Answers obtained by the groups

		Ladders		*Sequence*	*Christmas Trees*		
	Number of Year	M(20)	M(1000)	S(100)	X(20)	X(100)	Checking Episodes
G9	9	62	3002	598	79	399	3
B9	9	62	3002	598	79	399	3
G8a	8	68	3400	580	79	395	5
G8b	8	62	3020	598	*	*	5
B8a	8	62	3004 ##	418#	79	399	3
B8b	8	70	3500	580	110	550	1?
G7	7	62	3002	598	*	*	10+
Correct answers		62	3002	598	79	399	

* Insufficient time to tackle this problem.
#arithmetic error ##logical error

Table 3

The number of problems where use of correct and whole object methods are suggested or accepted for final answers.

	Number of problems attempted	WO method suggested	WO method accepted	Correct method suggested	Correct method accepted
G9	3	0	0	3	3
B9	3	0	0	3	3
G8a	3	3	3	2	0
G8b	2	1	1	2	1
B8a	3	1	0	3	3
B8b	3	3	3	0	0
G7	2	2	0	2	2

SUMMARY OF THE PROTOCOLS

Groups using specific known cases for checking

Three groups, G9, B9 and B8a checked their ideas by comparing predictions against specific known cases. This checking was always effective. The methods which were checked in this way were not subsequently used by the group in any question. For example, in finding M(20), B8a had tested both the whole-object method and their theory that $M(r) = 3r + 2$ (expressed verbally only) by drawing a 20-rung ladder and counting the rungs. This checking was effective: they did not use the whole-object method for M(20) and did not mention it in any later question.

The two year 9 groups did not propose the whole-object method and solved all questions correctly. They checked their ideas both with general arguments and against specific known cases. Group G9 worked algebraically, in each case seeking a formula with constant reference to specific numerical examples. Group B9 used the geometric structure in the problems to develop general methods, again with reference to specific examples. Only one wrong method was suggested: finding M(1000) by multiplying the number of rungs (1000) by 3 because " there are three matches per rung". This idea was disposed of by checking it against the smaller ladders already drawn. The idea did not re-surface.

Persistent checking

Groups G7 and G8b exhibited reasonably effective checking. In their discussion, the G8b girls proposed seven methods for solving *Ladders*, following a pattern of conjecture, and refinement or rejection. Once again it was the case that when ideas were checked by comparing predictions against specific cases, they did not re-surface later. Checking which pointed to general reasons was less effective. In the extract from the G8b protocol below, note how Ros uses a specific case to dismiss Tan's second wrong method (which does not re-surface later), but only a general reason is advanced by Sue against the whole object method, which is subsequently implemented. This group drew the 20 rung ladder to find M(20), found M(100) using a correct linear relationship but then used the whole-object method to find M(1000) without further discussion. This phenomenon of students slipping into error changing from a correct method to an incorrect method as the numbers get larger has been noted in a variety of content areas (eg Hart, 1981; Stacey, 1989).

Sue: (seeing the 1000 rung question in *Ladders*)
"there must be some easy way you can work it out"
Tan: (proposing whole-object method)
"work out how many for 100 and multiply by 10"
Sue: (objecting to whole-object method)
"*no, we've got to add 3 for each rung, you have to add three"*
Tan: (proposing a second wrong method)
"1000 threes?
Sue: "*yeah, 1000 threes"*
Ros: (objecting using specific example, noting that M(5) is not equal to 5 threes)

"no, because for 5 rungs you have 17 matches"

Sue: (considering M(1) and perhaps suggesting simple direct proportion again)

"yes, you have 5 matches for every one rung "

Ros: (proposing and dismissing a variation of Tan's wrong method)

"so there's going to be 5000 matches; that can't be right"

Tan: "*how many in 5 do you need ? 17* "

Sue: (picking up on Tan's original suggestion, despite her own previous general objection)

"*find out how many we need for 100"*

(some dialogue omitted here)

Sue: (after all have worked out M(100) as 302 correctly)

"so its 302 matches "

Ros: (implementing whole-object method again)

"*for 1000 rungs, 3020 matches."*

It was persistence that enables the G7 girls to answer all questions correctly, despite suggestions to use the whole-object method for both M(20) and S(100). The whole-object method arose several times in their lengthy protocol for *Sequence*, which they finally solved correctly by comparing the answers obtained by three different methods, one of which was the whole-object method and two of which were correct. They had rejected the whole-object method when someone in effect proposed S(100) = 10 x S(10) yet they accepted it for S(50) x 2 = S(100). As shown in the extract below, Kris, like so many other students reported by Hart (1981), reveals that she sees doubling as unlike other multiplications. Like the G8b girls, the G7 girls had six somewhat different ideas for doing *Ladders*, dropped two (one right and one wrong) without discussion and tested two before use. One of these methods was the correct

$M(r) = (2r + 2) + r$ which was tested on several specific cases and reasoning from the structure of the ladder. Note the sophisticated checking suggested by Jess in the extract below which saves them from the whole object method in *Ladders.*

Kris: (suggesting the whole-object method for M(20), knowing that M(4) = 14 and M(5) = 17)

"Wouldn't it be easier to do . . . 5 rungs is 17 so 5 by 4 is 20 so if you times 17 by 4 you get how many twenties?"

Jess: (suggesting a way of checking the whole-object rule)

"try 14 by 5 and see if you get the same answer"

Kris: (agreeing, calculating 14 x 5 and 17 x 4, then seeing that they are not equal)

"oh, good idea" . . . what's that? oh dear"

Jess: *"you didn't get the same answer"*

Kris: (proposing doubling as a special case of the whole-object method)

"Should it? Yeah, it should of. . . . Oh well.

How many is 10? If 5 is 17 then ten would be double 17 wouldn't it? Wouldn't it be double what 5 is?"

Discussion then moved on to another proposal. These girls had plenty of ideas and worked persistently to choose between them.

Ineffective and non-existent checking

Two groups, B8b and G8a, exhibited little effective checking. Queries went unanswered, or were answered merely by repetition of the proposed calculation. Group B8b adopted the whole-object method for M(20) without comment and used it without discussion in every question thereafter. In the total protocol for all three problems, the only one instance of checking behaviour was Alan's mild query, shown in the extract below from their second use of the whole-object method. Alan received no reply and the erroneous method was subsequently used in every question.

Con: (after having noted correctly that as the number of rungs increases by one, the number of matches increases by three and using this fact to find that M(4) = 14)

"if its 14 for 4 . . . 50 . . , 70 . . 14 . . . goes into 20 five times so if you've got to work it out you times that by 5 gives you 70"

Ben: (repeating Con's calculation of M(20) using whole-object method)

"14 x *5 = 70 "*

Con: (beginning to find M(1000))

"then you can take it as a multiple of that and times it by 250. Times that by 50 then 70 x *50 = 3500"*

Alan: (expressing doubts)

"one rung is three . . ."

Con: (ignoring the query and continuing to calculate)

"20 goes into 1000 5 times so you do 5 times the answer of that 50 by 70"

End of dialogue.

The girls in group G8a also failed to check ideas properly although, as will be illustrated by the extracts of the transcripts below, their disquiet with the whole-object method grows as they continue to use it. Group members proposed five ideas in trying to find M(20), three of which could have led to the correct answer. They were all dropped without any discussion at all except for the one objection raised by Cath but not sustained (see * below in the transcript). Instead of discussing proposals, new ideas were put forward. The whole-object method was implemented for M(20) without discussion and then used without further consideration to produce M(1000). Two different correct methods were actually used for X(20) but the whole-object method was used again for X(100) and then again for S(100). Each time it was used, the uneasiness increased. As the transcripts below show, Cath suggests the whole-object method each time. Beth suggests correct relationships in *Ladders* and *Sequence* that could easily be implemented to give correct answers for M(1000) and S(100) but mysteriously she drops them each time, although with increasing reluctance as shown in Encounters 1, 2 and 3! In Encounter 1, the whole-object method is adopted without discussion, although other ideas are proposed. In Encounter 2, a query is answered with a repetition of the proposed calculation. Finally, in Encounter 3, just when the checking was clearly displaying their errors, the girls fail to notice what it means.

Encounter 1: *Ladders*

Cath: (proposing whole-object method for finding M(20) from M(5), known to be 17)

"5 rungs has 17 and so 4 times 5 . . . what's 4 by 17?"

Beth: (proposing a correct method and the correct answer)

"no, the number of rungs is . . .there is an extra one for the number down , so if there's 20 rungs there's 21 down each side so that's 42 plus 20 for the rungs equals 62"

Ann: (ignoring Beth's correct and elegantly stated solution and proposing another idea)

"17 from 5 then there's 15 more rungs"

Cath: (proposing another idea)

"let's draw it . . . if there's 17 from 5 "

Beth: (proposing another idea)

"*how many matches in a rung?"*

Cath*: (offering the only objection made to any idea)

"*no , that's for the whole ladder"*

Beth: (adopting Cath's original whole-object approach)

"*4 by 17 equals 68"*

Ann: "*yep*"

Cath: "*yeah*"

Encounter 2: *Christmas Trees*

In the second problem, two slightly different yet correct methods with convincing explanations were given by Beth and Cath for finding the number of lights on a Christmas tree with 20 levels. Then Cath proposed the whole-object object method to find the number of lights of a tree with 100 levels. Beth and then Ann queried this, but their queries were immediately dismissed when Cath repeated her proposed calculation (79 x 5) with no other explanation.

Encounter 3: *Sequence*

For *Sequence*, stronger objections are made but the two checks attempted are not carried through properly!

Cath: (proposing the whole object method to obtain S(100) from S(10))
"*So you just times 58 to 580*"

Beth: (querying the whole object method)
"*How do you get that? Does it work?*"

Cath: (offering the first real reply to an objection in the whole session)
"*But that's our principle for all the rest of them. Well if that's going to be 10 times that number . . .because you're going to keep going . .*"

Beth: (proposing a general check on the whole object method)
"*that's like saying this [S(100)] is the fifth number times 20. See if you get the same thing*"

Beth then writes 28 x 20 and calculates it correctly as 560. Meanwhile, Ann and Cath together try to check whether similar relationships among the first ten terms of the sequence. They check whether S(10) = S(2) x 5, for example. The checks show the relationships do not hold, but this confuses Ann and Cath, who then reject the tests. Making an almost Freudian slip, Beth also fails to notice the result of her test:

Beth: (reading the correctly calculated and clearly written answer of 560 as 580!) *"580 . . .here, it's the same. Yeah it's 580. Because that's the fifth number by multiplying that by 20 to get 100 . . .it's the same number."*

Cath: *"oh, I see what you mean"*

End of dialogue - the whole-object method apparently passes the test. Whereas the girls in G7 were able to use the sophisticated checking of the rule which Jess had proposed, the G8a girls, although a year older, were only confused when Beth made the same suggestion.

DISCUSSION OF RESULTS

Solving these problems in groups was characterised by plenty of ideas and hypotheses, so that getting ideas seemed to be no problem for any group. But the choices that were made between the ideas were often not considered at all carefully. In the discussion of every group other than that of B8b, at least one correct method for solving each problem (although not necessarily both of the questions asked about the problem) was proposed. All these groups of students therefore had access (although not equal access) to correct methods of solution.

Conversely, at some stage every group except G9 mentioned a definite, wrong method for finding answers. The whole-object method, for example, was proposed in discussion at some stage by each of the five Year 7 and 8 groups. Three of these groups used it and two did not. The important difference between the groups was not having the right idea, but selecting it carefully from all the ideas proposed and following it through. In many instances, it was the attractive simplicity of the whole-object method that caused it to be selected: it is both simple to understand and simple to implement. Queries were often dispelled merely by repeating the calculation to be performed. This points to the crucial role that checking behaviour played in these problems.

The groups which fell most heavily into the trap of the whole-object method exhibited very little effective checking behaviour. Ideas were proposed and dropped, then revived without discussion later. The checking that was most effective in discarding wrong ideas compared predictions with specific known cases. Where general arguments were used to support or dismiss an idea, the wrong ideas tended to resurface later. For example, when G8a, G8b and G7 supported or objected to their methods with general reasoning, in each case they used the wrong method later in the same problem.

Correct methods were checked or defended with both specific known cases and general arguments. Groups that did not accept correct methods after they had been proposed did not actively reject them: instead they tended to be passed over. This is in accordance with the observation of Egerbladh and Sjodin (1986) who noted the effect of inaccessibility of ideas on group interaction. In the general pattern of dialogue, there are many instances of proposals being floated into the conversation, yet receiving no response. Ideas, both right and wrong, were often dropped or used without any discussion. Groups G8a, G7 and B8b were particularly prone to this.

CONCLUSION

In mathematical problem solving by junior secondary students, there is no evidence here that two or even three heads are better than one. The behavior of the groups documented above provides a clear explanation for the failure of groups to score better than individuals on the problem solving test. Although group work did lead to enormous pooling of ideas, groups had difficulty selecting correct approaches. East accessibility of the ideas seems to be a key factor. In the group situation, the benefits of using a simple idea are obvious. Everyone "understands" the method which is being used and it produces answers quickly. The whole-object method, for example, is generally put forward and also justified only as a calculation, not as a general principle. Students with doubts are generally not given time to formulate and articulate them. It is likely that students who only see the simple idea are more confident than students who glimpse the complexity. This confidence may make them more likely to prevail in a discussion. This hypothesis is being investigated by my colleague, Cecilia del Beato.

Teachers using group work in the classroom need to be alert to the attraction of simple ideas. To overcome them, teachers must strongly reinforce group behaviour where ideas are considered carefully and students are given time to articulate their ideas. Teachers must also be certain to provide feedback on the correct answers.

As has been demonstrated by Egerbladh and Sjodin (1986), the effects of group work on problem solving will depend on the nature of the problem solving task. Much of the research that recommends problem solving in groups has been concerned with problems of a quite different type to the mathematical problem solving being discussed here. The tasks used in this study of mathematical problem solving, and Treilib's modelling problems (1979) represent a type of problem solving which is perhaps not well tackled by groups. For example, although students engaged in a good deal of "brainstorming" (which is well done by groups), careful, reflective thinking is important in these and other mathematical problems. There is also no strong opportunity in these mathematical problems for diversity amongst students' background, talents and abilities to contribute to a solution.

These observations support the call by Lester (1989, p 123) who warns against too great an emphasis on small group work in problem solving. The use of groups in teaching mathematical problem solving can still meet some of the aims (e.g. to reduce anxiety and for the pooling of ideas) which were outlined in the Introduction to this paper, but care needs to be taken so that superficial and impulsive thinking are not reinforced. It also is clear from the protocols that the use of group work was not successful in providing a natural forum where clear explanations are required. Students did not demand explanations from their peers.

Finally, the study also has disturbing implications from the point of view of teaching about mathematical proof. Dismissing or accepting ideas on the basis of general argument was an unsuccessful strategy for almost all of the groups. One of the key ideas of mathematical proof is that numerical examples alone do not contribute to proof. It is essential to understand the mathematical structure in a general way, yet the evidence here shows the enormous difficulty that students have in dealing with general arguments. On the other hand, for those groups who knew to try a specific case, the import of a numerical counter-example was surprisingly well understood.

REFERENCES

Egerbladh, Thor & Sjodin, Sture (1986) Joint Effects of Group Composition, Group Norm, Type of Problem and Group vs Individual Responding. *Scandinavian Journal of Educational Research,* 30(1) March 1986 pp 17 - 23.

Good, Thomas L., Grouws, Douglas A. & Mason. DeWayne A. (1990) Teachers' Beliefs About Small Group Instruction In Elementary School Mathematics. *Journal for Research in Mathematics Education*, 21 (1), pp 2 - 15.

Hart, K (Ed.) (1981) *Children's Understanding of Mathematics : 11 - 16.* London: John Murray.

Lester, Frank, K. (1989) Reflections about Mathematical Problem Solving Research. In Charles, Randall I. and Silver Edward A. (Eds.) *The Teaching and Assessing of Mathematical Problem Solving.* pp 115 - 124. Reston Va:Lawrence Erlbaum and NCTM.

Lovitt, C. & Clarke, D. (1988) *Mathematics Curriculum and Teaching Program.* Canberra: Curriculum Development Centre.

Mathews, Robert C., Lane, Irving M., Reber,Robert A., Buco, Steven M., Chaney, Courtland M.& Erffmeyer, Robert C. (1982) Towards Designing Optimal Problem-Solving Procedures: Comparisons of Male and Female Interacting Groups. *Group and Organization Studies*, 7 (4), pp 497 - 507.

Richards, M. & Trotter, J. (1989) Problem Solving Task Centres. In B. Doig (Ed.) *Everyone Counts*. (p 431) Melbourne: Mathematical Association of Victoria. .

Stacey, K. (1989) Finding and Using Patterns in Linear Generalising Problems. *Educational Studies in Mathematics*, 20 (2), pp 147 - 164.

Stacey, K. & Groves, S. (1985) *Strategies for Problem Solving,* Melbourne: Latitude Publications.

Treilibs, Vern (1979) *Formulation Processes in Mathematical Modelling.* Thesis for the degree of Master of Philosophy, University of Nottingham.

Victorian Curriculum and Assessment Board (1989) *Mathematics course development support material.* Melbourne: publisher as author.

Victorian Ministry of Education (1988) *The Mathematics Framework P - 10.* Melbourne: Victorian Ministry of Education (Schools Branch)