ORIENTATION TO DEEP STRUCTURE WHEN TRYING EXAMPLES: A KEY TO SUCCESSFUL PROBLEM SOLVING

KAYE STACEY and NICK SCOTT
University of Melbourne

There are many characteristics of a good mathematical problem solver: having wide mathematical knowledge, understanding content thoroughly, being confident and persistent, having experience of solving problems, being able to use heuristic strategies and having metacognitive knowledge. Factors such as these work together to create good problem solvers. Sometimes, for example, extensive knowledge and experience makes problem solving simple; at other times good use of heuristic strategies, metacognitive expertise, confidence and persistence can compensate for lack of knowledge.

In this chapter, we analyse sample protocols of graduate and undergraduate mathematics students working on a problem in number theory. The problem lends itself to the use of numerical examples, and so the major heuristic strategy employed is to try examples. The chapter highlights how problem solvers used the strategy “try some examples” in very different ways. Some used numerical examples to gain insight; others only collected evidence. Most tried too many examples but another tried too few and consequently lost contact with the problem. Such issues of metacognitive control will also be illustrated. Examples were often used to discover the facts of “what” was true—what we will call the surface features of the solution—but only some participants used examples to uncover reasons for these facts—what we will call the deep structure of the problem. The problem solvers oriented to deep structure knew to look beyond patterns in the “data” obtained by trying examples and use their mathematical techniques and skills to expose mathematical structure. Some observations of the role of teaching in the development of orientation to problem solving conclude the chapter.

Try Some Examples: A Heuristic for Seeking Surface Features or Deep Structure

“Trying some examples” is probably the simplest of all problem solving heuristics. Mason, Burton and Stacey (1982) place it into the somewhat wider category of “specialising”, which they identify as one of four basic processes of thinking mathematically. In their analysis, specialising and generalising are dual processes that drive mathematical thinking, along with the second dual pair of conjecturing and convincing. They advocate specialising as a simple technique that everyone can use to get started on a question or to make progress when otherwise unable to proceed.

Commentaries by experienced problem solvers such as Polya (1957) and Mason et al (1982) point out several roles for the heuristic of trying examples—to become familiar with a situation, to find out what might be the case and to gather clues about why it might be so. Mason et al, for example, comment:

> "*Your aim in [trying examples] is two-fold: to get an idea of what the answer to the question might be, and at the same time to develop a sense of why your answer might be correct. Put another way, by doing examples you make the question meaningful to yourself and you may also begin to see an underlying pattern in all the special cases which will be the clue to resolving the question completely.*" (p. 2)

Elsewhere, they describe how specialising can be carried out randomly, systematically and artfully. In this chapter, we will show how some participants used examples in all of these ways, whilst others limited their use to finding out what rather than why.

Despite the agreement of experts on the importance of specialising, there are very few research studies of how students implement it. In one of the few studies of students' understanding of the use of examples in problem solving, Stacey (1992) noted that groups of 14 year old students had difficulty generalising from specific numerical examples, but that there was good understanding of the role of numerical examples to disprove conjectures.

Studies of expertise in problem solving have frequently identified how experts focus on deep structure, whereas novices focus on surface features. Now classic experiments, summarised by Schoenfeld (1985), have shown that expert chess players perceive positions of pieces on a chessboard in terms of broad arrangements (deep structure), rather than as separate placements of pieces (surface structure). Similarly, physics problems are sorted by experts according to the physical principles involved in their solution (deep structure) rather than according to the objects in the problem statements (surface structure). Similar differences have been observed in successful and less successful students' memories of mathematics problems. The successful students were said to have recalled deep structure whereas the less successful students were said to have recalled surface features.

Our use of the terms deep and surface is slightly different. We will describe the facts (or partial facts) of what is true about a mathematical situation as the surface features and the reasons (or partial reasons) why they are true as deep structure. In terms of the processes underlying mathematical thinking described by Mason et al (1982), searching for surface features ends at conjecturing whereas deep structure is concerned with convincing.

Describing Problem Solving Protocols with Episode Graphs.

In the field of mathematics education, Schoenfeld (1985, 1992) has produced the most influential insights into control behaviour during mathematical problem solving episodes. He observes:

> "*With good control, problem solvers can make the most of their resources and solve rather difficult problems with some efficiency. Lacking it, they can squander their resources and fail to solve problems easily within their grasp.*" (1985, p. 44)

In his 1985 book *Mathematical Problem Solving*, Schoenfeld details a method for graphically representing the various changes in behaviour that result from control decisions during a problem solving episode. He used these "episode graphs" of the problem solving sessions to characterise the differences in control behaviour exhibited by novice and expert problem solvers. These graphs will be used in this chapter to summarise some aspects of the problem solving protocols that we have collected.

Schoenfeld makes several claims about episode graphs, especially that the type of episode graph is linked to the success in problem solving and also to the solver's general expertise in problem solving. One qualification of the last claim is that the problem should be challenging to the solver. An expert working on a routine problem, for example, would not exhibit a great deal of control because this is not needed in a straightforward task. Despite the impact of Schoenfeld's work on thinking about control behaviour—and the attractiveness of the graphical representation he developed—there have been very few articles published which have used the graphing technique, analysed it from a methodological viewpoint, or further explored the links between control and expertise. In *Mathematical Problem Solving* Schoenfeld describes only six protocols in detail. This chapter therefore provides an opportunity to extend the very small data-base on which the conclusions about the link between episode graphs, expertise, control and problem solving behaviour have been built and tested. Several methodological points relating to the parsing technique have been investigated in relation to the protocols that are described in this chapter. These are not presented here but are discussed thoroughly by Scott (1996).

OVERVIEW OF METHODOLOGY

For this study, we obtained detailed descriptions of nine solutions from participants "thinking aloud" as they worked on a challenging mathematical problem. The protocols documented cognitive actions and metacognitive control actions, to the extent that these can be observed or confidently inferred. The participants were ten first year undergraduate students specialising in mathematics (about 19 years old) and four postgraduate students enrolled in research degrees in mathematics. The post-graduate students and one exceptional undergraduate who had represented Australia in an International Mathematical Olympiad were classified as experts and worked alone. The others were classified as novices and solved the problem in pairs, as in Schoenfeld's work.

Participants worked on a problem, Staircase Numbers (see Figure 1) from Stacey and Groves (1985). This was chosen because it is challenging yet amenable to a range of solution approaches and accessible to problem solvers of varying levels of mathematical ability and training. A full solution to this problem can be found in Mason et al (1982). The concrete presentation of the staircase numbers was selected (rather than the abstract presentation of the same problem as "which numbers are sums of consecutive numbers?") to increase accessibility and increase the possibility of visual and intuitive methods. One of the features of this problem is that it is easy to try numerical examples to gain information about many different aspects of the problem. The degree of success achieved by each solution attempt is measured crudely by recording the number of aspects of the solution to the problem that were stated. Eight aspects of the solution, described in Figure 2, were chosen for these measures. This is a very rich problem and much more can be found out about it than these eight solution outcomes, but only one participant found more than these in the half hour available for each recording session.

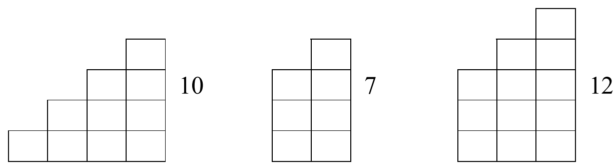
Schoenfeld’s (1985) methodology for collecting and parsing protocols was followed as closely as circumstances permitted. The problem solving sessions were videotaped and then transcribed. Transcripts were parsed and episode graphs were constructed.

Staircase numbers

A staircase number is a number that can be expressed as the sum of consecutive integers. For example, 10, 7 and 12 are staircase numbers because

$10 = 1 + 2 + 3 + 4$
$7 = 3 + 4$
$12 = 3 + 4 + 5$

We can imagine the ‘staircases’:



In the case of 10, the heights of the stairs are 1, 2, 3 and 4.

The number 4 is not a staircase number because the only way of writing it as a sum of consecutive integers is the trivial and uninteresting way with one stair, of height 4.

Which numbers are staircase numbers and which are not?

Find a recipe or rule that will give the heights of the stairs for any staircase number.

(Adapted from ‘Strategies for Problem Solving’, K. Stacey and S. Groves, 1985)

Figure 1. Staircase Numbers problem sheet used in this study.

Outcome	Description of Outcome
1	Identifying that all odd numbers (after 1) are staircase numbers
2	Informally demonstrating how odd numbers can be written as the sum of two consecutive numbers (e.g. half of 17 is 8.5 and so 17 = 8 + 9)
3	Stating outcome 2 in a general way, possibly algebraically [e.g. $n = (n-1)/2 + (n+1)/2$]
4	Identifying that powers of two are not staircase numbers
5	Identifying that staircase numbers with an odd number of stairs are multiples of that number, or vice versa.
6	Using outcome 5 in some way to provide some reasons for outcome 4 e.g. by observing that powers of two have no odd factors other than one and so cannot be staircase numbers with an odd number of stairs.
7	Devising a rule for expressing even numbers (other than powers of two) as staircase numbers.
8	Describing how to write for any given number, other than a power of two, as a staircase number.

Figure 2. Description of the main outcomes in the solutions to Staircase Numbers problem.

SUMMARY OF SOLUTIONS AND PROTOCOLS

In total nine useable protocols were collected. Table 1 shows that two of the experts found all eight of the outcomes specified in Figure 2 whereas the other two experts and the five novice pairs found from two to four of them. After examining the protocols, there seemed to be four different types. Three of the protocols, those by experts LI, LR and SF, display individual features. However, substantial commonality was found in the other six protocols (one by expert DW and all those of novices) and so in the discussion below, the protocol of NW & JM will be discussed as representative of them all.

Table 1: *Solution outcomes for participants.*

Participant(s)	**Category**	**Solution outcome**							
		1	2	3	4	5	6	7	8
LI	Expert	×	×	×	×	×	×	×	×
LR	Expert	×	×	×		?			
SF	Expert	×	×	×	×	×	×	×	×
DW	Expert	?	×	×	?				
NW & JM	Novices	×	×	×	×				
SC & DH	Novices	×				?			?
SL & WS	Novices	×	?	?	?	?			
AC & BC	Novices	×			×				
JL & MC	Novices	×	×	×	×	?			

KEY × outcome achieved ? some aspects of outcome achieved

Note: Participants classified as experts worked alone, novices worked in pairs.

Participant LI

LI is a very talented undergraduate mathematics student. His protocol shows a solution that seeks the deep structure of the problem, operationalised by using an algebraic formulation. A summary is given in Appendix 1. His high success on the problem came from the use of targeted examples and strategically important global assessments and his confidence in the various algebraic formulations that he devised, even to the point of ignoring small errors. Although there was much more to be discovered in this rich problem, in sixteen minutes, LI produced a solution covering all outcomes expected, as shown in Figure 2. He used numerical examples in several ways:

- to reduce unnecessary generality from algebraic expressions,
- to search for deep structure by looking for the general in the particular,
- to identify algebraic errors.

Possibly because LI did not try a wide range of examples, he did not realise there was a limitation to the formula that he had developed (it sometimes gave negative stairs). Had he tried a wider range of numerical examples, he may have been lead to investigate the substantially deeper results linking odd factors with staircases with even numbers of stairs (See Mason et al, 1982).

LI's episode graph is shown in Figure 3. In constructing an episode graph, behaviours identified from the participants' speech and writing are divided into episodes and the episodes are categorised into six types: read, analyse, explore, plan, implement and verify. Points where control activities occur (such as assessing the amount of progress that has been made) are highlighted using the triangles that can be seen in the graph. The episodes are then arranged on a timeline that indicates episode duration, transition between episodes, and managerial activity. Further details can be found in Schoenfeld (1985).

(Insert Figure 3)

Figure 3. Episode graph for LI

The episode graph of LI (see Figure 3) was similar to the episode graphs that Schoenfeld associated with experts. This reflects the fact that LI constantly monitored his work—assessing the results of calculations and procedures to see how they fit with expectations and continually refining his global solution plan in light of these assessments. Two aspects of the episode graph show evidence of this control. Firstly, the triangles indicate individual instances of control. Secondly, the graph moves up and down showing that LI's work consisted of various categories of episodes and that the solution moved frequently between episode types. Schoenfeld claims that the transitions between episodes of different types are the result of metacognitive control. This graph shows the characteristics that Schoenfeld (1985) identifies in the episode graphs of other expert problem solvers.

Participant LR

A doctoral topology student, LR had an exceptional record of study in mathematics. However, in his solution (see Appendix 2), LR covered only four outcomes from Figure 2 (see Table 1). He justified an algorithm for writing all odd numbers as staircase numbers and made preliminary assertions about staircase numbers with three columns being divisible by three. Much of his time was spent investigating subgoals; the criteria for staircase numbers to have three, four or five stairs. Like LI, he sought deep structure in the problem, evidenced by an algebraic formulation. However, LR used very few examples, so lost touch with the problem and hence made very little progress towards a solution. LR used examples to:

- gain initial familiarity with the problem (briefly),
- identify algebraic errors.

His algebraic representations were primary and numeric work was done in the context of the algebra. For example, 21 was demonstrated to be the sum of m and $(m - 1)$ where $N = 21$ and $m = (N+1)/2$. LR did not seek to illustrate the algebra with examples, although in two instances numeric examples pointed him to errors in reasoning and manipulation ($N = 12$ and $N = 14$). Had he looked at more examples systematically, he may have been able to move away from his one approach, which he had assessed early on as "*not really getting anywhere*".

(Insert Figure 4)

Figure 4. Episode graph for LR

The episode graph (see Figure 4) was similar to the episode graphs that Schoenfeld associated with experts. There were many instances of control, but it was local, assessing progress towards subgoals rather than the overall goals of the problem.

Participant SF

Expert participant SF was a Masters student in operations research, having been an excellent student throughout her career. Her protocol (see Appendix 3) shows that she also looked at the deep structure in the problem, but she did this differently to LI and LR, by collecting structural evidence from examples. She used a wide range of examples and she chose them systematically and with care. Like participant LI, SF made strategically important global assessments, which probably contributed to her successful solution. All eight outcomes from Figure 2 were covered by her investigation.

Despite her success and her "expert" status, her episode graph (see Figure 5) was similar to the graphs that Schoenfeld associated with novice problem solvers. There is little control behaviour, as evidenced by few triangles indicating solution assessments and few transitions from one category of behaviour to another. Schoenfeld asserted that the graphs of protocols done by novices generally appeared like that of SF.

SF used algebra only as a notation to signify quantities in her written work. However, her use of examples was highly systematic and thoughtful. She used examples:

- to look for patterns and suggest conjectures,
- as carriers of general structure,
- as a source of further directions for investigation.

(Insert Figure 5)

Figure 5. Episode graph for SF

Participants NW & JM

The other protocols, one from an expert and five from novice pairs, exhibited many commonalities, and so only one example is given. NW and JM were a pair of novice problem solvers, who asked to work as a team and worked well in this way. They were both above average students in mathematics at school and neither found the mathematics they were currently studying too difficult. Their protocol displays concentration only on the surface features of the problem. They used examples to collect data about what happens, but not about why it happens. In one instance, they used a numerical example to explain a rule also stated algebraically. Table 1 shows that they discovered the main solution outcomes for the problem, but they did not know why these results were true. Their global assessments did not impact on their solution process. The episode graph (Figure 6) was similar to the episode graphs that Schoenfeld associated with novices.

(Insert Figure 6)

Figure 6. Episode graph for NW & JM

ANALYSIS OF THE PROTOCOLS

Table 2 summarises differences between the four protocols. Analysis of the protocols showed that the major differences contributing to success on the problem seemed to lie in the following dimensions, which will be illustrated below:

- whether the participant looked at the deep structure of the problem or concentrated on surface features,
- how the participant used examples and
- the nature of some of the assessments of solution progress that were made.

Table 2. Characteristics of Selected Protocols

Participant(s)	LI	LR	SF	NW & JM
Category	Expert	Expert	Expert	Novice*
Outcomes reached	8	3	8	4#
Structure	Deep	Deep	Deep	Surface
Formulation	Algebraic	Algebraic	Numeric	Numeric
Examples chosen	Few, targeted	Too few	Wide, targeted	Too many
Episode graph	Expert-type	Expert-type	Novice-type	Novice-type
Assessment of progress	Strategic and global	Frequent but local	Strategic and global	Low

* The NW & JM protocol is typical of six protocols obtained, including one from an expert.
# The NW & JM protocol was the most successful of the six that it represents.

Deep Structure and Surface Features

As indicated in the descriptions above, a major division was found in how participants approached and thought about the problem. LI, LR and SF all considered the problem from what we call a "deep structure" perspective. Their work acted on the problem at a structural level. This was either in terms of an algebraic formulation of the problem (in the case of LI and LR), or through the reconciliation of evidence from examples perceived in structural terms (participant SF). These protocols explained both *why* the solution (or part solution) worked as well as *how* it worked. These solutions involved formulation of structural components of the problem.

The other six protocols focussed on "surface features" of the problem. These solutions did not go beyond summarising information noted from the generation of numeric examples, usually systematically. There was no real attempt to formulate the problem to make it amenable to analysis. Consequently, they were limited in what they were able to discover (see Table 1). For example, three of the six protocols included a rule for writing odd numbers as staircases of the form

$$N = \frac{N-1}{2} + \frac{N+1}{2}.$$

However, this generalisation was a summary of observations rather than the result of reasoning. As another example, SC & DH looked at dividing numbers by various factors as a basis for determining stairs, but they could never take this beyond a trial and error process, and it was never fully tested.

Deep Structure and Success

Table 1 shows that only two solutions, both exhibiting deep structure, succeeded in achieving more than half the solution outcomes for this problem. Consideration of deep structure however was not enough to ensure success. LI and SF covered all the solution outcomes described in Figure 2, although there was certainly more to be found in this rich problem. The deep structure solution of LR, on the other hand, was less successful than the novice, surface feature pair of NW & JM.

To see what else might be required, compare the solutions of LI and SF to that of LR. Although LR worked within a framework that helped to elicit structural elements of staircase numbers, he did not note some simple properties like powers of 2 not being staircase numbers (outcome 4). He had a representation for describing and manipulating staircase number descriptions but he had no examples within which he could ground his thinking. In contrast, both LI and SF used examples to "point them in the right direction" so that they could focus their investigations on one or two salient aspects. Some of LR's work, on the other hand, became so abstracted from the "real" situation that he began to lose touch and make algebraic slips. It was quite a while before LR was able to follow an example through his algebraic work far enough to detect a simple mistake. The combination of deep structural investigation *along with* targeted examples seemed to be required for progress.

The Episode Graphs

The three expert deep structure solutions show two different types of episode graph. LI and LR have similar graphs, both of the type Schoenfeld associates with experts, although one was successful and the other not. This shows that success is not implied by an expert-type episode graph. On the other hand, the highly successful solution of SF gave a graph of the form that Schoenfeld would associate with non-expert problem solving behaviour.

In contrast to the variation in the deep structure episode graphs, all the surface feature protocols resulted in episode graphs that looked like Schoenfeld's typical novice episode graph. (Some methodological points relating to this claim can be found in Scott, 1996). However, it was noticeable that quality of thinking was not reflected in the episode graphs. For example, the pair JL & MC looked at uniqueness of staircase representations and discussed ideas about averaging stairs around an average stair height. They also worked on a way of algebraically representing staircase numbers—but with little success. The fact that their episode graph was similar to the graphs of the others and did not highlight the clear differences in the quality of their investigation is an important point to consider when using episode graphs.

Trying Examples

NW & JM used the examples they generated to identify that powers of 2 were not staircase numbers. They acknowledged that they had not proven this assertion, but did not attempt to do so. The pair found a rule and a reason expressed diagrammatically for odd numbers (outcome 2), but otherwise they looked for patterns in the examples they generated, without ever considering what properties of staircase numbers led to these patterns. Similarly, AC & BC exhibited little in the way of analytic behaviour, spending the majority of their time checking examples on a calculator. When stuck, they often went back to more example generating simply because it was something they knew they could do. AC & BC justified these actions with comments like *"at least we're still collecting data"*.

All six sets of participants in the surface feature group showed this behaviour of "try a few *more* examples". In contrast, when participants providing deep structure protocols were stuck and chose to turn to examples, their behaviour was different. The move back to generating more examples was not handled in the same way. For instance, when LI stopped to try some examples when he did not know how to proceed, his choice of examples for consideration was very particular. He knew exactly the type of examples he needed to generate in order to confirm or refute his current ideas about the problem. He only worked with these examples for as long as they provided him with new information. Then he went back to analysis to accommodate this information into his previous work. This made the exploration focussed and useful. In a similar fashion, SF's solution depended on her systematic generation and interpretation of examples.

Two contrasting uses of examples are seen in the protocols of LR and of NW & JM, whose achievement of outcomes (Table 1) was comparable. LR's solution suffered from too few examples, whilst NW & JM suffer from too extensive a use of examples. LR spent the majority of his session working with an algebraic formulation of the problem intended to model the structure of staircase numbers. He generated few examples and spent little time considering which of his examples were staircase numbers and which weren't. His local and global solution assessments were well placed and accurate when he made them, but unfortunately, the lack of assessment of the examples he worked with meant that he never discovered that powers of two were not staircase numbers. On the other hand, NW & JM were very good at spotting patterns and consolidating information collected during their explorations. They "discovered" important points relating to odd staircase numbers and powers of 2 but left them unproven. In the end LR knew less, but knew it better; while NW & JM "knew" a little more, but in an unsubstantiated way.

It is hard to tell from the protocols whether concentration on surface features was due to reluctance to consider deep structure or inability. Some of the surface feature protocols show that participants did try to go beyond simple pattern spotting from examples. For instance, SC and DH attempted to generalise their finding that the stairs of a 2-stair staircase number balanced around the mean value (e.g. stairs of 8 and 9 balance around 8.5) to staircases of other sizes as illustrated in Figure 7. Unfortunately they were unable to refine the process of choosing the best factor and this left their resulting procedure at the "trial and error" level.

(Insert Figure 7)

Figure 7. SC & DH balance stairs around a mean stair height

In summary, while "try a few examples" was a strategy that all participants knew and employed at some point in their work on the problem, there were significant differences in the way in which the strategy was implemented and how the information collected was used. Where examples were specifically designed to illuminate insights or to confirm or refute assertions, they could be effectively used as a way of advancing in the problem (see participants LI or SF). For the six sets of participants in the surface feature group, examples were generally not used in this way. They generally used the strategy "try some more examples", just collecting more data in the hope of spotting some more results.

CONCLUSIONS

Necessary Resources for Solving the "Staircase Numbers" Problem

The protocols in this study provide an interesting set of contrasting examples of problem solving. As summarised in Table 2, some but not all of the deep structure solutions were successful as were some but not all of the solutions corresponding to expert-type episode graphs. Moreover, both experts and novices produced surface structure solutions, which seemed to result in similar episode graphs.

This was a hard problem for many of the participants because their initial behaviour in generating examples was generally not conducive to progression towards considering the underlying structure of staircase numbers. Only two participants made any real attempt at algebraically formulating general expressions for staircase numbers that they could then investigate. Two other participants made some use of the visual representations of staircases provided with the problem to think about "balancing" stairs around some central "mean" stair height. This work did not generally move beyond "drawing pictures" however and did not translate into additional solution outcomes being demonstrated.

While "try some examples" is a common strategy advocated in the problem solving literature and in teaching, it was used poorly by many participants in this study. It certainly seems that to solve the "Staircase Numbers" problem, the use of examples was essential. The two complete solutions provided by participants in this study made use of examples that were specifically designed to illuminate insights or confirm or refute assertions. The other participant in the deep structure group—one of the strongest mathematics students in the study—suffered from *not trying enough examples*. He was unable to discover some simple facts about staircase numbers.

For the six sets of participants in the surface feature group, the difficulty generally lay in *trying too many examples*. In some respects, these participants seemed to equate trying examples with solving the problem. They only used examples to "discover" facts about the problem, or when they were stuck and did not know what else to do. Their solution attempts demonstrated very little "targeting" of examples to suit specific questions. Most of their time was spent lost in a sea of examples with nothing but surface features obvious in the information to guide participants. In general, only a combination of deep structural investigation *along with* targeted examples seemed to assist solution progress.

There was substantial commonality amongst the degree of success and the episode graphs of the surface feature solutions (as would be predicted from Schoenfeld's work) but there was variation in quality of thinking amongst them that neither measure revealed. It seemed that the episode graphs could be used effectively for the identification of *changes* in behaviour brought about by local assessments of the knowledge state and solution progress. In this respect Schoenfeld's parsing approach had real value. It can also be concluded that graphs that indicate control behaviour through repeated variations in episode categorisations are indeed indicative of control-related problem solving practice. However good control may well exist in cases where such features are not prominent in an episode graph, as in the case of SF.

Implications for Teaching

The younger participants in this study (all the novice pairs and LI) attended secondary school when the teaching of problem solving first achieved wide popularity in our education system. Our personal observations and anecdotal evidence indicates that in many classes, strategies like "try some examples" have been taught to students in a superficial way. Stacey (1991) provides some documentation of this. The use of examples to identify surface features of solutions and as an aid to guessing patterns has been foremost, with little attention given to the use of examples to gain deeper insight into a problem. Discussions with some of the participants in this study indicate that the message that gets across in teaching is that trying examples is an adequate method of problem solving. As one said, *"you can solve almost any problem if you work out enough examples"*. It may not be coincidental that, apart from LI, the undergraduate participants showed little interest in uncovering the deep structure in the problem. They were good at collecting information, but their inability to reconcile it in meaningful ways let them down. Along with lessons about how, when and where to use the "try some examples" strategy, it is important to build students' skills in formulating problems structurally so that patterns observed are more than surface level indicators of a possible answer.

Our assessment of some of the ways in which curricula valuing problem solving have been implemented in many Australian schools is that students have been exposed to a strongly empirical approach to mathematics. Gathering data and observing and describing patterns have become common activities in classrooms, whilst proving results or questioning why results hold (or indeed if they always hold) have a low profile. General reasoning is often not expected by the teacher. Moreover, the problems chosen for students to work on are often puzzles, where solutions can be found by trial and error but are too difficult for students to prove (Stacey, 1995). Accompanying this has been a marked decrease in proof in areas of the curriculum where it was traditionally found. For example, a common textbook series justifies the theorems about angles in circles only by measurement of a few examples. It may have been expected that an emphasis on open problem solving may have provided many opportunities for students to make conjectures and to prove them, even if not formally. However, these possibilities seem not to have become reality.

In the light of these curriculum trends, it is perhaps not surprising that the novice students did not use examples to search for deep structure in the problems. However, it does set a clear goal for changing teaching so that students learn to use "try some examples" as the powerful strategy that it is.

REFERENCES

Mason, J., Burton, L. & Stacey, K.: 1982, *Thinking Mathematically,* Addison-Wesley, London.

Polya, G.: 1957, *How To Solve It.* (2nd Edition), Princeton University Press.

Schoenfeld, A. H.: 1985, *Mathematical Problem Solving,* Academic Press, Orlando, Florida.

Schoenfeld, A. H.: 1992, 'Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics', in D. A. Grouws (ed.), *Handbook of Research on Mathematics Teaching and Learning*, Macmillan, New York, pp. 334-370.

Scott, N. H.:1996, *The Utility of Schoenfeld's Protocol Parsing Technique for Detecting and Illustrating Control-Based Behaviour in Non-Routine Problem Solving Episodes.* Unpublished M.Ed. thesis, University of Melbourne.

Stacey, K.: 1995, 'The challenges of keeping open problem-solving open in school mathematics' *Zentralblatt fur Didaktik der Mathematik* **95**(2), 62-67.

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

Stacey, K.:1991, 'What Does Teaching About Problem Solving Teach?', in F. Furinghetti (ed.), *Proceedings of the Fifteenth International Conference on the Psychology of Mathematics Education,* Assisi, Italy, pp. 278-285.

Stacey, K.: 1992, ‘Mathematical Problem Solving in Groups: Are Two Heads Better Than One?’, *Journal of Mathematical Behavior,* **11**, 261-275.

APPENDIX 1

Protocol of LI (Expert with a Successful Deep Structure Solution)

After reading the problem, LI began an algebraic formulation of the problem based on the sum of a general arithmetic progression, to investigate the effect of varying the size of the starting stair and the total number of stairs. After 30 seconds of silent contemplation at the 1.05 minute mark, he made some important observations about the conditions of the problem (e.g. there are a minimum of two stairs and the smallest staircase number is three). This work removed some of the generality from his algebraic formulation.

After a silent transition at 2:43 minutes, LI decided to "*try some more examples*". LI appeared to be in deep thought during this transition; perhaps he knew that he was having difficulty with the algebra of the problem and that trying some examples would help him to get a better feel for the variables that he was working with. LI systematically checked all the numbers from 1 to 7 during which he noted that "*odd numbers can obviously be done just by taking the integer part of the half*" (outcome 2). With the resulting information LI quickly moved back to analysis. He seemed loath to try any more examples at this stage (as would have been reasonable—he hadn't encountered many numbers that were not staircase numbers at that point). This action reinforces the view that he was confident of his original formulation of the problem, and preferred to focus on the structure and relationships between the variables, rather than look for surface level patterns in the data generated during this episode.

From 3.34 to 6.22 minutes, LI focussed his analysis on even numbers, finally considering what happens if the sum is four times a square. This unproductive approach was abandoned and not considered again. During this work, he further reduced unnecessary generality from his algebraic formulation.

At 6:45 minutes, LI began systematically checking all the numbers to 19 to see which were staircase numbers, summarising these results in a table.

At 9:15 minutes, LI said "*OK, let's go back to the start again*" and he began implementing an approach based on parity that would have been familiar to him from his Olympiad training. Going back to the start meant *right back* before his arithmetic progression work. In the subsequent implementation episode, LI covered solution outcomes 4 and 6 from Figure 2.

LI made several algebraic slips during this episode (e.g. writing $n+m$ instead of $n-m$) which might have prevented him from deriving his result. However, he still reached the correct conclusions. Because LI did not verify his solution at this point or at any later stage, or use it to generate some staircase number examples, he did not see that his algebra was wrong until he used it sometime later on a sample number, 23.

At 12 minutes, he then moved on to generating a procedure for producing the stairs for any staircase number, with a straightforward implementation of his previous results. This work covered solution outcomes 5, 7 and 8 from Figure 2. The problem solving session finished at 16.25 minutes.

APPENDIX 2

Protocol of LR (Expert with an Unsuccessful Deep Structure Solution)

LR read the question, listed the first few staircase numbers and noted that they included the set of "triangular" numbers. Then he formulated the problem algebraically as the sum of $n + 1$ stairs with largest stair m:

$$S = m + (m - 1) + (m - 2) + \ldots + (m - n)$$

He summed this expression, which culminated in an algebraic description of the set $\sum$ of staircase numbers:

$$\sum = \left\{(n+1)\left(m - \tfrac{1}{2}n\right) : m, n \text{ natural numbers, } n < m\right\}$$

At 5.58 minutes he asked: *"Well, what's there to know [about $\sum$]? It's got, ... it's a countable set ..."* and the interviewer suggested he move onto the second part of the question, trying, perhaps to write 23 as a staircase number.

LR did this by solving the equation $m + (m - 1) = 23$ after deciding that 23 must have at least two stairs. He went on to write the equation above as a more general expression for 2-stair representations:

$$m = \frac{N+1}{2} \text{ and } m + (m-1) = 23$$

and concluded that this expression always has a solution if *N* is odd. This work covered solution outcomes 1, 2 and 3 from Figure 2. He then turned to expressing even numbers as staircase numbers with three stairs. From the equations:

$$m + (m - 1) + (m - 2) = N, \therefore 3(m - 1) = N$$

he concluded: "*So 3 outside (m - 1) is equal to an even number so there's no solution*." He then checked the problem sheet and asked (rhetorically): "*Well, hang on, there's a solution here for 12, [points to the question sheet] so what have I done wrong?*"

He soon realised that the $(m - 1)$ term in his expression above must be even, that *N* must be a multiple of 3 and that if a number is not divisible by 3 then it cannot have a

3-stair representation.

Then LR then worked on the case of staircase numbers with 4 stairs, concluding that "*N on 2 plus 3 must be even*", and that "*that's not such a nice looking formula... We're not really getting anywhere*". Working occasionally with $N = 14$ as an example, he concluded that in a 4-stair representation, "*N is congruent to 1 mod 4*", an incorrect result deriving from the algebraic error $2(28 + 1) = 4\xi + 1$. He did not pick up this mistake until 23.56 minutes, when he looked again at the case N = 14. Later he also worked similarly on staircases with 5 stairs eventually concluding that "*I need 2N plus 15 to be divisible by 10*". His algebraic approach is illustrated in Figure 8.

(Insert Figure 8)

Figure 8. Extract from participant LR's transcript working on the 4 stairs case demonstrating his algebraic approach.

APPENDIX 3

Protocol of SF (Expert with a Successful Deep Structure Solution)

SF's transcript was difficult to interpret because periods of silence and incoherent sentences made behaviour and underlying thinking difficult to interpret. Also because of this, there was more interaction from the interviewer than in other protocols.

SF began by writing down several algebraic ways of describing the structure of a staircase number as and, like LR, asked if more was required. The interviewer suggested that SF try to determine whether 47 was a staircase number. In response to this, SF moved into what was to become a very productive exploration episode that lasted until 23:18 minutes.

SF started this episode by systematically recording what happens when i and n vary: "*to see what happens*". This quote gives an indication of the nature of this episode. SF wasn't sure of the type of answer she should expect and so was effectively "searching the problem space" without much direction. SF set up an ingenious table (see Figure 9) for looking for patterns by recording the staircase number against x, the "starting stair" and n, the number of stairs.

(Insert Figure 9)

Figure 9. Extract from participant SF's transcript

Her first attempts at filling out these tables for n = 2 and 3 steps included some arithmetical errors which she later rectified. These initial errors were propagated throughout the n = 4, 5 and 6 rows. She did not fix these. After looking for patterns in the table in Figure 9 with little success, she then extended her table to include staircase numbers with starting stairs of 5, 6 and 7. She then quickly noted that two stairs could represent any odd number (from the second row of the table) and that even staircase numbers would need at least three stairs. During this phase, SF was taking note of observable patterns without a clear idea of why they were occurring. SF's table was probably responsible for her ease in picking up patterns in the problem. It seems likely that SF must have had an idea about the important structural elements of the problem before she chose to establish her table as she did. After 9.00 minutes, SF had covered solution outcomes 1 and 2 as well as part of 5.

However, there were long periods which were incoherent to the interviewer (but not to SF). The interviewer therefore began to intervene, asking about a rule for generating odd numbers as 2-stair staircase numbers (given at 9:44 minutes) and SF began to describe "balancing" numbers around a middle stair. For example, $18 \div 3 = 6$ so that three stairs balanced around 6 will give 18: $18 = 5 + 6 + 7$. The difficulty of understanding SF's "think aloud" protocol when she was lost in thought is illustrated by the following utterance, her first attempt at describing which numbers aren't staircase numbers:

> "*So basically that means that ... any number by an even number ... may not ... and not by other numbers ... are not going to be able to be written as staircase numbers. Because we can't sort of add and subtract ...*" .

Because of the incoherence, at 14:22 minutes the interviewer interjected and prompted with a request to try writing 28 as a staircase number. SF used part trial and error and part "balancing" to come up with $28 = 1 + 2 + 3 + 4 + 5 + 6 + 7$. After a 35 second pause from 16.17–16.52 minutes SF noted that she could have done this example by "considering the pairs" either side of 4—the result of dividing 28 by the odd factor 7.

SF continued working with the balancing process, largely in silence to the end of the session on the example numbers 26 and 44, both suggested by the interviewer. Finally, SF wrote the description in Figure 10 of how to express a number with an odd factor as a staircase number.

(Insert Figure 10)

Figure 10. Extract from participant SF's transcript

This iterative procedure for generating staircases relied on identifying an odd factor; writing it as a 2-stair staircase number; and then adding pairs of stairs either side of this starting pair until the staircase had the correct value. For example, for 44 take the odd factor 11. Write it as $5 + 6$ first and then as $4 + 7$, $3 + 8$ and $2 + 9$. As a result

$44 = 2+3+4+5+6+7+8+9$. She did not resolve the question of what happens when this procedure leads to negative numbers, although she knew it was a difficulty.

The transcript from SF does not provide explicit justification of her assertions.

However, the interviewer accepted that SF knew the solution *and knew why*. SF's conviction was in her x/n table, and in her procedure that entailed balancing based on an odd factor.

APPENDIX 4

Protocol of NW &JM (Novice Pair with a Surface Features Solution)

This novice team was able to identify (but not prove) that powers of 2 were not staircase numbers and write a procedure for finding a staircase representation for odd numbers (of the form $N = \frac{N-1}{2} + \frac{N+1}{2}$). They were able to use this procedure successfully on a number of examples. They had little success devising a similar algorithm for even staircase numbers. They covered most of the elements of solution outcomes 1, 2, 3 and 4 from Figure 2. They were able to collect and summarise useful information relating to the problem but could not integrate this information through any form of analysis.

NW & JM began by clarifying what they meant by the "heights" of stairs and considered a number of examples which lead by 5:40 minutes to the conclusion that powers of two were not staircase numbers. Until 19:00, they looked at ways of systematically generating staircase numbers and discussed the uniqueness of the representations sparked by the example of 18 = 5 + 6 + 7 and 18 = 3 + 4 + 5 + 6. JM argued from a diagram that all odd numbers could be written as 2-stair staircase numbers and NW agreed, using algebraic symbols and a numeric example to write down the rule (although not to prove it) as shown in Figure 11. By 22:05 minutes they were discussing possible patterns amongst their representations of the even numbers from 6 to 22, but had made no progress except to continue the list of data points to around 30 when the session stopped at 30:00 minutes.

(Insert Figure 11)

Figure 11. Written work by NW & JM.