STUDENTS' CERTAINTY AND CHECKING BEHAVIOUR DURING MATHEMATICAL PROBLEM SOLVING

CECILIA DEL BEATO and KAYE STACEY
Institute of Education, University of Melbourne

This study investigates the certainty and uncertainty that students feel as they work on a mathematical problem and how this relates to the checking that they carry out. It is hypothesised that the over-confidence in decisions that characterises reasoning in many fields of human endeavour is also exhibited in mathematical work and that it may partly explain why students generally are reluctant to check their work. Students who feel certain that their work is correct would see little reason to check it. In the problem used in this study, students became uncertain when they moved from a particular case where they could count to a much larger case where a general rule was required. They also became uncertain when the arithmetic became harder - the size of this effect had not been expected. Students with wrong methods that gave easy arithmetic were, in the end, almost as certain that their answers were correct as students with the correct method. Students often did not know how to use extra information to check their answers. About half of the students who were correct became less certain after being given supporting information.

How confident students feel in the correctness of their answers to mathematical questions is a subject that has been little investigated. This paper proposes that it is of interest because how certain one feels in one's decisions and solutions should affect whether one checks or not. A frequent finding of research into the methods and strategies used by students to solve mathematical problems is that they neglect to check (Galbraith, 1986; Stacey, 1989). Several reasons have been given to explain this phenomenon. One is that "reflection" is a neglected phase of problem solving receiving little instructional time. Students do not seem to feel a need to check their work: they respond in an immature way, leaving to the teacher the responsibility of determining whether something is right or wrong. There is also considerable evidence that children do not know strategies for checking their work. Stacey and Groves (1985) noted that many students interpret a verbal instruction to check as only an instruction to repeat. Bell, Costello and Kuchemann (1983) and Galbraith (1986) have also noted that there is widespread misunderstanding about the roles of examples and counter-examples in checking. To check, one needs an adequate understanding of the content involved and also an appreciation of the logical structure of proving things to be true and false. Lee and Wheeler (1987) , for example, point out that students can only use substitution of numbers as a strategy for checking algebra if they have a reasonably clear understanding of the relationship between arithmetic and algebra.

That certainty may also influence checking behaviour was suggested to us by Fischbein's (1987) analysis of the role of intuition in mathematical thinking. Fischbein reviews research from a variety of sources to support his hypothesis that certainty in a reasoning process is produced by reliance on immediate, self-evident, intrinsically certain intuitions. These are necessary for reasoning to be efficient and productive. He sees certitude as a fundamental need of the human mind; without it a reasoning process could not continue. In fact, a substantial body of research into the subjective evaluation of confidence has consistently found that people are overconfident in the accuracy of their own knowledge, decisions , interpretations and solutions (Lichtenstein, Fischoff and Phillips, 1982). This natural tendency to overconfidence may contribute to students' lack of checking, particularly of their reasoning. Although Fischbein's analysis mostly concerns stable conceptual structures and clusters of beliefs, he also recognises "anticipatory" intuitions which are specific to the problem solving process.

Galbraith also noted that when students were asked to select the better of two explanations, many students made their choice of the grounds of simplicity. This observation has also been made by Stacey (to appear) who observed that when groups of students were solving problems together, they often chose a simple incorrect solution, even when a correct solution had been proposed by a group member. The crucial factor in whether the group solution was correct seemed not to be having the ideas, so much as choosing between them. Misplaced confidence in a simple idea and lack of adequate checking strategies were common faults.

AIMS

With the background outlined above in mind, data were collected to explore the following questions:

i) What factors cause certainty and uncertainty in problem solving amongst students?

ii) How does the certainty of students who answer a question correctly compare to the certainty of students who answer it incorrectly?

iii) Is there a relationship between certainty and checking? Are students who do a problem in only one way more certain of their results than students who search for other approaches?

iv) What is the effect on a student's certainty of information supporting and information contradicting an answer?

METHOD

One problem (see Figure 1) was given to 227 Year 8 students (average age 13 years) at two girls' schools during class time. After completing the problem, the students answered a questionnaire. About 15 students (exhibiting different responses) were interviewed shortly after the problem solving to further elucidate the reasons for the responses. The questionnaire asked students to rate their certainty in their answers to the three parts of the problem on scales from 0 to 10. The zero, five and ten positions on the scale were annotated with descriptive comments, such as "completely sure your answer is wrong" at zero. This measurement of certainty was adapted from calibration studies reviewed in Lichenstein, Fischhoff and Phillips (1982). Then students were asked to indicate, separately for the 10x10 and 50x50 blankets, whether they had done the question in only one way or in more than one way and whether they had obtained one or more answers. This was the criterion selected for judging whether students had checked their work. Finally students were asked to re-assess their certainty on each part of the problem after being presented with one of three pieces of additional information. They were randomly given the correct answer of one of the following cases: 20x20 blanket (version 1), 50x50 blanket (version 2), or both 3x3 and 6x6 blankets (version 3). Version 2 gives them direct information about their answer to the 50x50 blanket. Version 3 is intended to draw students' attention to the incorrectness of a method which involves finding the answer for a larger blanket as a whole number multiple of the answer for a smaller blanket. This will be referred to as the "multiples of previous answer" method. A student using this method assumed the amount of sewing was directly proportional to the size of the blanket, so may, for example, have multiplied by the answer for the 10x10 blanket by 5 to get the answer for the 50x50 blanket. The purpose of giving the three different versions was to compare the relative effect of direct and indirect information on re-assessing certainty.

Imagine you are going to make a patchwork doll's blanket by sewing together some tiny squares of material measuring 1cm by 1cm. You want to know how much sewing you will have to do to make the blanket.

To make a square blanket measuring 3cm by 3cm, you need 9 squares of material and it takes 12 cm of sewing. (Diagram given of 3x3 blanket with sewing between the squares, but not along the outside edge of the blanket, clearly shown)

- *How much sewing is needed to make a square blanket measuring 5cm x 5cm?*
- *How much sewing is needed to make a square blanket measuring 10cm x 10cm?*
- *How much sewing is needed to make a square blanket measuring 50cm x 50cm?*

Figure 1: Problem solved by students in study.

RESULTS AND DISCUSSION

About half (55%) of all students and two thirds of the students who were correct checked their work in some way. Of those who did check, 31% were correct whereas only 18% of those who did not check were correct. This difference is statistically significant (chi-squared=5.3, d.f.=1, $p<0.02$). A t-test showed that students who checked their work and found the same answer in two different ways had significantly greater certainty than students who did not check ($p<0.05$). These in turn had higher certainty than students who checked but found two different answers. These results are as we had expected.

In order to relate changes in certainty to mathematical behaviour, responses were classified according to the solution method used. Some students followed one general rule for all three answers. Some began by drawing the 5x5 blanket and counting the number of centimetres of sewing but used a general rule for the 10x10 and 50x50 blankets. Others used the generalisation only for the third blanket. The mean certainty on each part of the problem is shown in Figure 2 for students consistently using the four most popular methods. Data for the whole group of students are given by Stacey and del Beato (to appear). The most common methods were:

i) the perimeter rule (51 students merely gave the perimeter of all the blankets);

ii) the correct rule, in effect $2n(n-1)$ (46 students, some of whom may have counted to obtain correct results for the smaller blankets);

iii) the "multiples of previous answer" method (14 students) explained above;

iv) the formula n^2+n to find the amount of sewing for the n x n blanket (20 students). Algebraic notation is implied here, although no students used it.

The perimeter and the n^2+n rule were probably popular because they fitted the given information that a 3x3 blanket requires 12 cm of sewing. In all cases shown in Figure 2, and all those not shown here but given by Stacey and del Beato (to appear), students were most certain about their answers for the 5 x 5 blanket and then certainty dropped for the 10x10 blanket and again for the 50x50 blanket. The higher certainty for the 5x5 blanket was associated with use of drawing and counting to obtain the answer. The only two groups that did not count (the perimeter and the n^2+n rule) had the lowest certainties for this part. These groups of students seemed to settle immediately on a solution that happens to fit the given data without exploring the situation in any depth, but they did not do this because they were very certain of the answer. In Fischbein's terms, they did not have an anticipatory intuition of which they were very sure. Instead, they seemed merely to accept the uncertainty and not do anything about it - even the few of them who drew the 5x5 blanket did not even count.

Uncertainty on moving to a generalisation and when calculating.

The responses of some students in the interviews indicated that the two important factors contributing to the drop of certainty as the blanket size increased were:

i) the uncertainty students felt when moving to a generalisation; and

ii) the uncertainty they felt about the results of calculations involving large numbers.

One student, for example, said "as the numbers increased I was less certain that the numbers were correct" and another said she became less certain because "I couldn't draw it so I had to work it out in my head."

The effect of the first factor (making a generalisation) was evident in the much larger mean drop in certainty when students moved from counting to use of a general rule (1.47 averaged over 27 clear instances) compared to the drop when students used the same rule on two consecutive parts of the problem (0.78 averaged over 18 instances). (Students who were correct were excluded from this analysis because it was sometimes unclear whether they had counted or used a rule.)

The effect of the second factor (more difficult calculations) was examined by comparing the mean drop in certainty from the 10x10 to the 50x50 blankets for students whose rules required only simple calculations (e.g. the perimeter rule which only required 50 x 4) with the mean drop in certainty for students whose rules required calculations involving a product of two digit numbers (e.g. 50x50). Students using hard calculations dropped their certainty by 0.96 (averaged over 25 clear instances). A t-test showed that this drop was significantly greater at the 5% level than the mean drop of students using easy calculations (0.21 averaged over 58 instances). The size of this effect had not been expected. This may partly explain why the natural checking of students concentrates so much on repeating arithmetic calculations, where teachers see a need to concentrate on checking the reasoning.

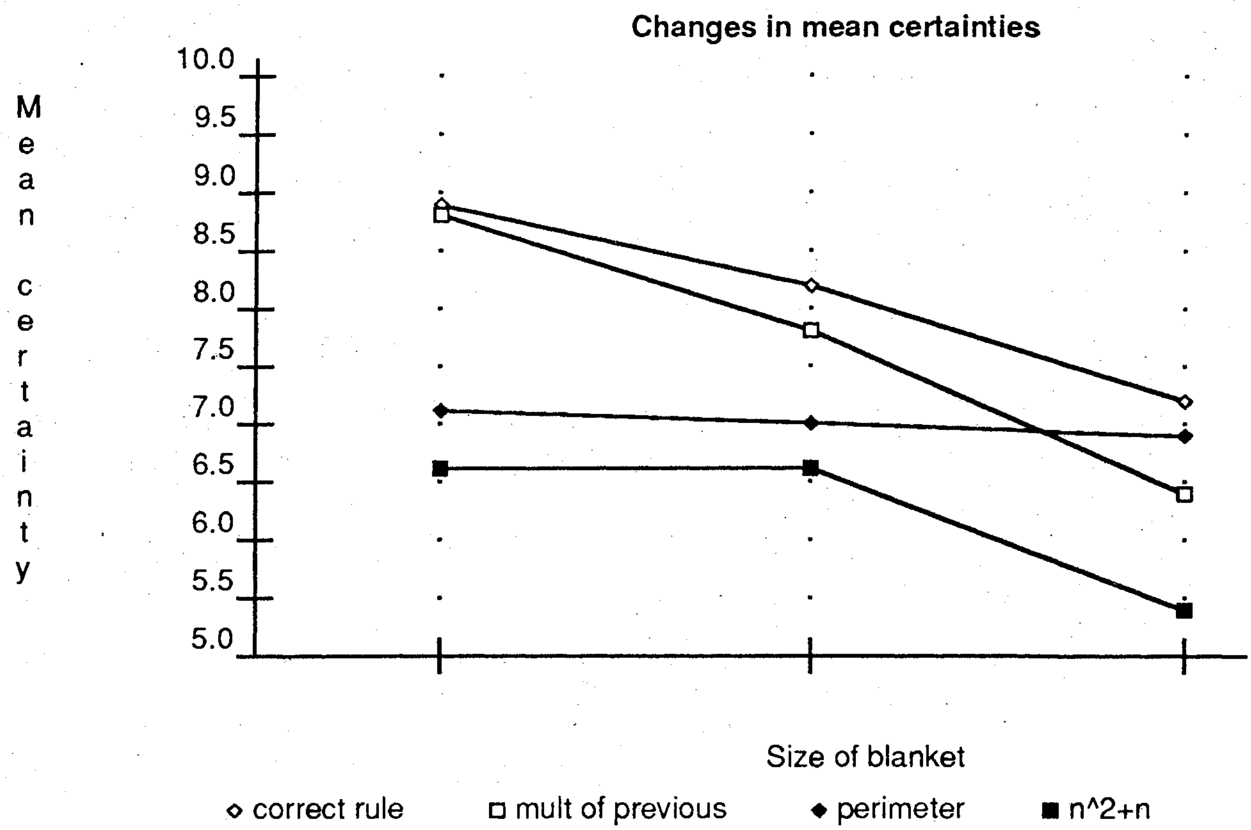

Figure 2: Changes in mean certainty of users of different rules.

Certainties of students with a simple or complex rule.

A comparison of students with the correct rule and the perimeter rule shows the practical implications of the drops in certainty due to generalising and harder calculations. The two factors combine, so that on the challenging part of the problem, the certainties of students who are correct and those who have grabbed the simplest wrong rule apparently without any investigation, are comparable. The perimeter group began with an immediate generalisation, based on little evidence. They did not subsequently have to make the transition to a generalisation and their calculations were at all stages very simple. Thus, although the perimeter group were significantly less certain of their 5x5 blanket answers ($p<0.01$) than the correct group, there was no significant difference between their certainties for the 50x50 blanket and those of the correct group.

Effect on certainty of information supporting or contradicting answers.

The number of students involved in this part of the study was 120 of whom 28 had correct methods and 92 had incorrect methods. Students had been asked to reassess their certainty in their answer for the 50x50 blanket after receiving either Version 1 (correct answer for 20x20 blanket), Version 2 (correct answer for 50x50 blanket) or version 3 (correct answer for 3x3 and 6x6 blankets). The numbers of students in each category are unequal because versions were distributed at random, when students finished solving the problem. For students who had correct methods, Version 2 caused an average increase in certainty of 2.27, whilst the other two versions caused slight decreases (-0.50 for version 1 and -0.14 for version 3) . For students who had incorrect methods, Version 2 caused the greatest average decrease in certainty (-3.46), whilst the other two versions caused smaller decreases (-2.41 for version 1 and -1.65 for version 3). Although Version 2 (the correct information) caused the greatest changes in certainty, only half of the incorrect students who received it dropped their certainty rating to zero. Responses from the interviews suggest that for some students there is a reluctance to completely reject an answer even in the light of contradictory evidence. This, a prediction of Fischbein's theory, has implications for checking. If students are unwilling to admit that an answer is wrong, they are unlikely to check.

The decreases in mean certainty for correct rule students had not been expected. This effect could not be attributed to a few outliers. Of the 17 students in this category, 8 had reduced their certainty. Clearly they had not seen the added information as supporting their (correct) ideas. We postulate that this was because they did not appreciate just how fast a quadratic function such as 2n(n-1) can grow and felt that their (correct) answer of 4900 was much larger than the answers given (760 for the 20x20 blanket). This phenomenon was also observed during the interviews. Some students used the additional information provided in Versions 1 and 3 only as a gross check on the size of their answer, rather than as a test for their own rules. Thus if the answer for the 20x20 blanket was between their own answers for the 10x10 and 50x50 blanket, they were often satisfied that they were correct. This is consistent with other findings that students do not fully appreciate how to check their work.

CONCLUSION

The use of certainty ratings has proved to be a promising research tool for understanding students' thinking during problem solving. When presented with evidence that an answer is incorrect, students are often reluctant to reject their answer completely. Students who jumped quickly to generalisations based on little evidence were just as certain about the correctness of their answers to the final part of the problem as students who were correct. Thus, in a group discussion, both a simple wrong rule and the correct rule may be propounded with equal conviction, leading a group to choose the wrong answer over the right answer. Both making a generalisation and doing arithmetical calculations caused students to lose certainty in their work. Students who have made wrong assumptions at the start of their work which happen to lead to simple processing will be no more likely to check their work at the end than students who are correct. Their simple (wrong) rule has "proved itself" in the ease of calculation of the answers it produces.

When given information which would enable an independent check of a generalisation, many students used it only as a guide to size. They possibly did not see the general rule they had themselves invented and used, as applying in a precise way to the given data. This is one example of the many aspects of checking which students might be taught about.

REFERENCES

Bell, A.W., Costello, J. & Kuchemann, D. (1983). *Research on learning and teaching.* Windsor: NFER-Nelson.

Fischbein, E. (1987). *Intuition in science and mathematics.* Dordrecht: Reidel.

Galbraith, P.L. (1986). The use of mathematical strategies: factors and features affecting performance. *Educational Studies in Mathematics*, 17, 413 - 441.

Lee, L., & Wheeler, D. (1987). *Algebraic thinking in high school students: Their conceptions of generalisation and justification.* Research report. Montreal: Concordia University.

Lichenstein, S., Fischhoff, B., & Phillips, L.D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahnemann, Slovic, P. & Tversky, A. *Judgement under uncertainty.* Cambridge: Cambridge University Press.

Stacey, K. (to appear). Problem solving in groups: are two heads better than one? *Journal of Mathematical Behavior.*

Stacey, K. (1989). Finding and using patterns in linear generalising problems. *Educational Studies in Mathematics*, 20,147 - 164.

Stacey, K. & del Beato, C. (to appear). Sources of certainty and uncertainty in mathematical problem solving. In *Proceedings of the Fifteenth Conference of the International Group for the Psychology of Mathematics Education*, Durham, NH, August, 1992.

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