Data Handling: An Introduction to Higher Order Processes

KEYWORDS:
Teaching;
Investigations
informal inference

Jane M Watson
University of Tasmania, Australia.
Rosemary A Callingham
Dept. of Education and the Arts, Tasmania, Australia.

Summary
An activity is described which allows students with a range of abilities to become involved in data analysis and informal inference while working in self-selected small group environments.

◆INTRODUCTION◆

Now that the mathematics curriculum documents of most countries require data handling to be part of students' mathematical experiences, it may be safe to assume that most students are collecting data and drawing graphical representations. There is a tendency however, particularly in the primary years, to stop here and leave the drawing of inferences for the high school years. Such a postponement is not in line with most curriculum statements. For example, one aspect of the National Council of Teachers of Mathematics (NCTM, 1989) *Standards* for Grades 5-8 is the ability to "make inferences and convincing arguments that are based on data analysis". In Australia this objective is covered in Band B for the middle school years in *A National Statement on Mathematics* (Australian Education Council, 1991) with "...draw informal inferences from data...

◆ACTIVITY◆

The activity described here had three major objectives:

- to allow Grade 6 children to explore their own resources for asking questions and drawing inferences about data sets,
- to expose them to some simple techniques commonly used to analyse data, and
- to give them the opportunity to prepare a report summarising the conclusions drawn.

It was decided to allow students to work in self-selected groups in order to encourage the sharing of ideas in an area with which they had little familiarity. Three 45-minute lessons were allocated for the activity. This time frame did not allow for students to collect their own data. Fictitious data were used in order to provide a range and number of measurements which were able to be processed in the time allocated and to provide interesting outcomes.

The materials for the activity consisted of identical sets of 16 data cards, one set for each student, of which the one shown in Figure 1 is an example.

Although fictitious, the data were devised using statistical norms to represent realistic values and categories of information. The ages ranged between 8 and 18; other activities included football, netball, swimming, TV, and reading. Eyes were blue, green and brown. The weights ranged from 26 kg to 74 kg and fast food meals per week from 1 to 12. A complete data set is available from the authors.

Name: Jennifer Rado
Age: 9
Favourite activity: Board games
Eye colour: Green
Weight (kg): 33
Fast food meals per week: 4

Figure 1 Example of a data card

Two classes of Grade 6 students participated in the activity. One had completed a data collection and representation activity in a local cemetery several months earlier but otherwise the students had no previous experience with data representation and analysis. All students were enthusiastic about the activity from its inception and most looked forward to working in groups which was not the normal mode of operation in the classrooms.

In the first lesson the introduction discussed data collection as a method of answering questions and the children were asked to speculate about how they would answer the question, "Do Grade 6 children prefer salt-and-vinegar potato chips?" After some discussion, the idea of sample arose and this was developed to distinguish it from the idea of population. Students were then introduced to the data cards as an example of a sample of children for whom information had been collected. With no further instructions on specific methods of investigation to follow, each child was given a set of cards and asked to form a group with some others to see what interesting questions they could ask, what they could learn and how they could convince others of their conclusions. At the end of the period the groups joined together to share what they had learned.

The second lesson, two days later, began with a review of what had been done before and a brief introduction to various more formal ways of representing information. These included constructing a table to summarise one variable, drawing a Venn diagram to classify two or more characteristics, drawing a scattergram to show the association of two variables, and creating a two-by-two table to study the relationship of two categorical variables. The examples given to illustrate these methods deliberately did not include any information that was on the data cards. Students then returned to their groups to further explore the data cards and make representations which would convince others of the relationships they believed they had found. Again there was a sharing time at the end of the lesson.

The third lesson was devoted to finalising the preparation of reports and posters to share with the rest of the class. The posters were then to be displayed around the classrooms. Throughout the lessons the teachers interacted with the students, answering questions and asking leading questions where it was deemed appropriate. At no stage, however, did the teachers impose a methodology on students.

◆INITIAL ATTEMPTS◆

During the first class period, the most common initial strategy was to list the names from the cards according to the other variables. For some children it was difficult to move away from the identities of the individuals in the samples. For example, two girls working together divided the tasks. One neatly listed all names by age, then by eye colour and then by weight, while the other listed names by number of fast food meals and favourite activity. A pair of boys, working at a large art table, sorted and re-sorted the cards on the table according to age groups, alphabetical order, fast food meals, weight, and boys and girls (by age). Other groups listed the information systematically, including a group who organised information about age, eye colour and first letter of last name into simple tables.

Other groups were more idiosyncratic in their sorting. A group of four girls looked for duplicates of information, for example age groups (five pairs), eye colour (blue, brown, green), activities (lists of names), fast foods (recording numbers for those with multiples) and weight (only one duplicate, 32 kg). They then looked for what they called "mixtures", which were relatively random combinations of small group associations. For favourite activity and eye colour they recorded ages. This approach varied however

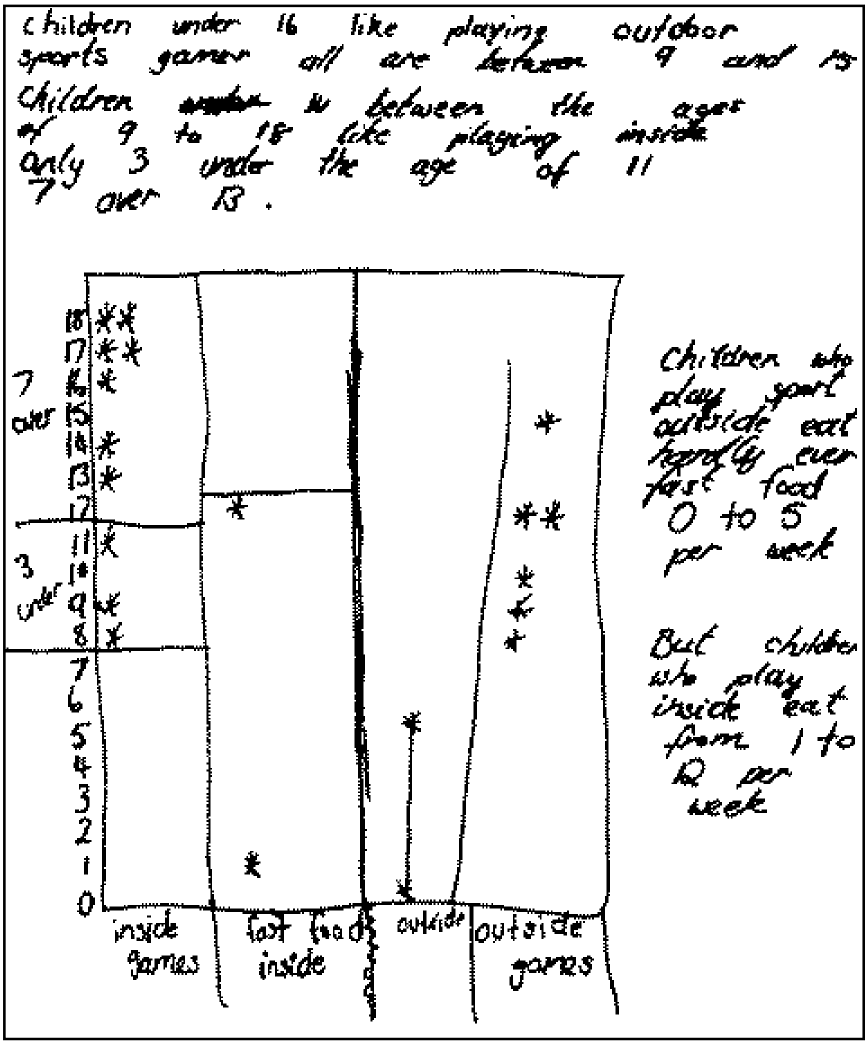


Figure 2 Posing Questions and Conclusions

when repeated data values in categories were used up and, instead, the lowest and highest values of other variables were recorded. There appeared to be no effort to look overall for a characteristic to study. Rather the group was interested to find two or three children who shared two or three features.

A group of three girls was the first to consider all of the data cards in terms of two categories, "inside" and "outside" activities. After putting the cards into the two categories, they looked at weight, fast food meals and age. One girl attempted to show her results graphically and this is shown in Figure 2. Her vertical scale of whole numbers represents age in the first and fourth columns and number of fast food meals in the middle two columns. For age there is a star for every child while for fast foods only the range of values is represented. Although there are some difficulties in the conclusions, the representation shows considerable ingenuity for someone who had not produced any graphs or charts during the current school year (which was 75% complete).

As the students manipulated their cards, they began to pose questions for themselves. The group of boys who had organised their cards on the large art table asked, "Do boys weigh more and eat more fast food than girls?" To answer this they found totals to show that the boys ate 45 fast food meals while the girls ate 15. One boy in a group of two sorted the cards first into age groups but randomly placed the groups around the table with no ordering. At the end of this session, the boy looked at the cards for the four boys who ate more fast foods than the others and came to the idea that watching TV might be related to weight and fast foods, but he could not organise his ideas in the light of the other 12 data cards. This boy's partner had trouble sorting in ages because so few were the same but when asked if he could "consider values close together" immediately started to group in l0's.

It was interesting to note that students did not at first necessarily choose characteristics which would be expected to be of the most interest statistically. Interest in eye colour and first letters of the names persisted for some time. It will take time, exposure and instruction to help children become aware of variables in social settings. The sharing session after the first lesson was valuable in that the children showed greatest interest in comments about fast foods related to activity, gender or ages, which are the variables of statistical interest.

By the end of the first lesson most students were attempting to draw some conclusions or state conjectures. These included the following: "There were more last names starting with M and W, and two Smiths." "More boys watch TV." "One boy ate and weighed more than two of the girls." "The ones who eat no fast food do the outside sports." "Older people weigh more." "People with blue eyes watch more TV." The conjectures illustrate the variety of focus possible: from individuals to categories to associations. This is to be expected in novices and is a rich source of discussion of the situations where statistics can and cannot be used effectively. In the beginning it is important not to be too critical of unsophisticated suggestions

◆METHOD OF JUSTIFICATION ◆

The outcomes of the second lesson were of interest because of the exposure to some of the conventional ways of displaying data to reinforce arguments. It was interesting to note the natural tendencies of students in the production of work. Although all the students sat at tables in groups, some produced individual graphs or tables and others produced a joint product. Only one student was interested in the "sorting" (Venn) diagram and required considerable support to produce this. She was not able to repeat the process with different attributes.

A variety of representations was used by other students. Some children took up one suggestion quickly and would not relinquish this for something more complex. One girl modelled very carefully the table created by the teacher in order to record the number of fast food meals in the left hand column and the number of children on the right. This she recopied three times during the second lesson after which she wrote a one sentence summary, "My graph shows that most people eat only up to 5 fast food meals per week."

Scattergrams were used by a number of groups. Some of the scattergrams followed the model given very closely while others used categories on one axis. A conventional scattergram is shown in Figure 3 displaying the association of number of fast food meals per week and weight. Others showed weight and age relationships. The scattergram shown in Figure 4 is effective in showing the differences in fast food intake for the various activities on the cards. Several children adopted this approach. Side-by-side bar charts were also commonly used, although in many cases adjacent bars did not display the same information. The conclusions drawn from these were less convincing. In one case a bar chart which alternated columns showing fast food meals per week and weight was accompanied by the label, "THIS PROVES PEOPLE THAT EAT MORE WEIGH MORE!"

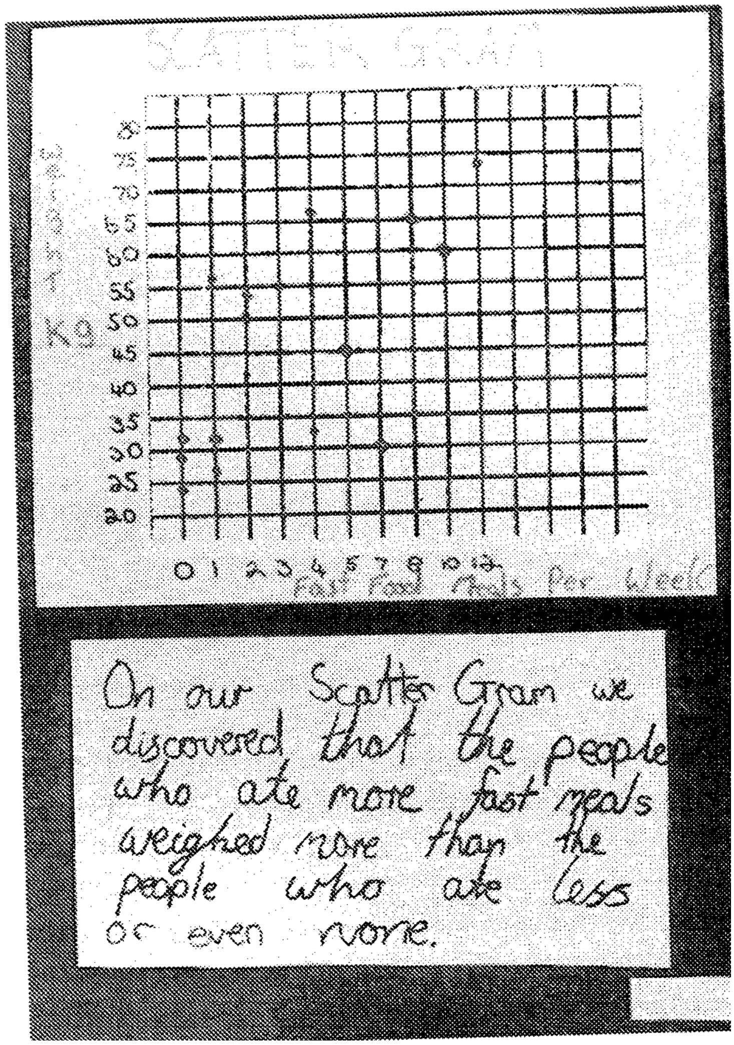


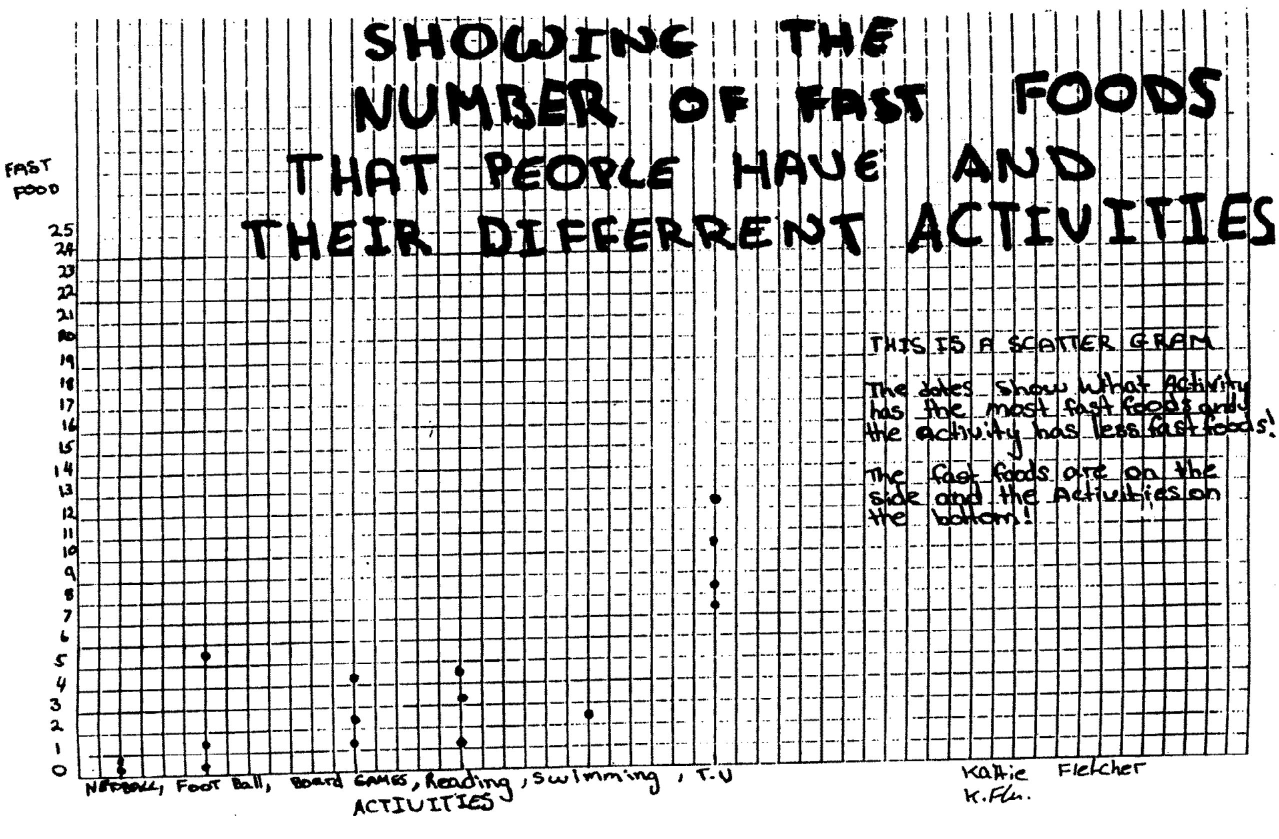
Figure 3 Conventional Scattergram

The use of two-by-two tables was taken up only by those who wanted to compare boys and girls. The poster shown in Figure 5 is typical of what was produced. The order of the presentation of the argument may appear unusual, with the conclusion in the middle. An earlier draft had the argument in a more logical order: “The same amount of girls play sport as boys... A lot more boys eat fast food than girls... The weight difference is probably because of fast foods.” It is likely that the group failed to realise the importance of retaining this order while pasting together the final product.

The most spectacular instance of choosing the data to make the point desired is shown in the bar chart in Figure 6. In order to represent fast food intake for the activity groups, one boy chose to use the datum for the person in each group with the largest number of fast food meals per week. This display provided the opportunity for a class discussion on how people may choose the data they use depending on the point they wish to make.

During the first session a group of four girls thought of using the arithmetic mean to summarise the data. They came up with this as a way of comparing the number of fast food meals eaten per week by the boys and girls. Their conclusion based on averages was, “in this graph we show that boys average a heavier weight than

Figure 4 Alternative Scattergram



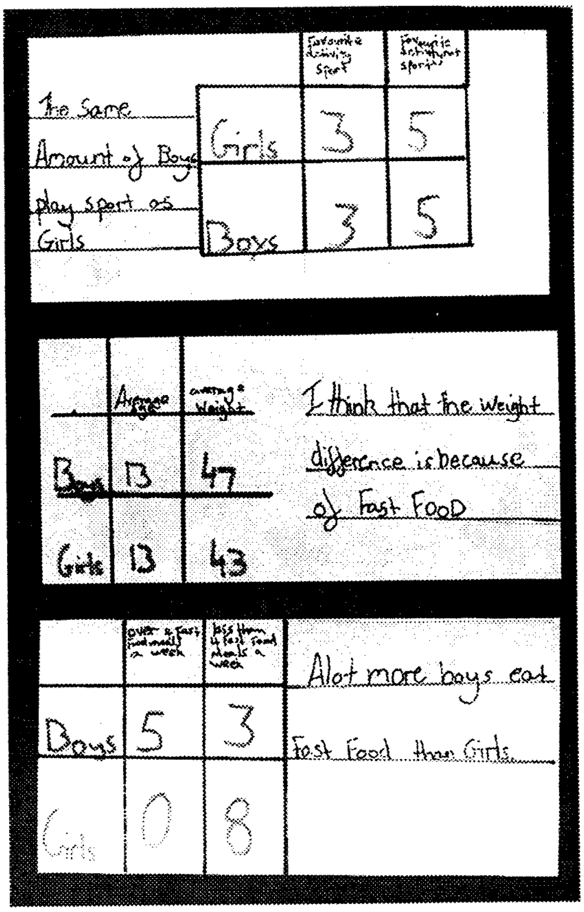


Figure 5 Typical Tables

girls and that boys have an average of 4 fast food meals more than girls per week." They did not however make a causal claim for the relationship.
Several other groups, including the boys who presented Figure 5, took up the idea of using the mean after the first sharing session.

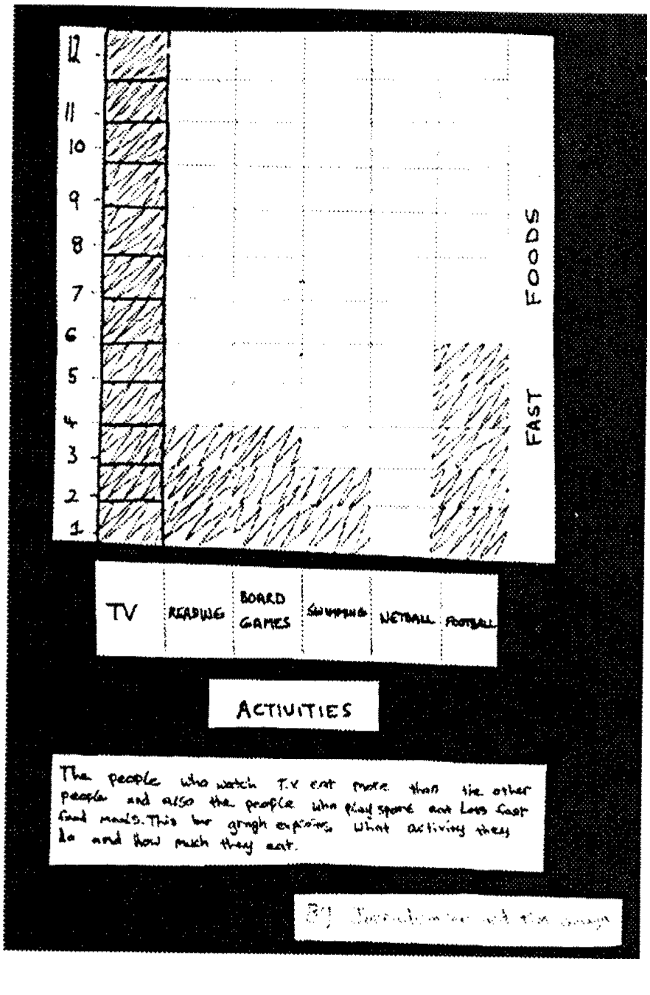
◆DISCUSSION ◆

The benefits of an activity such as this compared to a single worksheet activity are enormous. The ingenuity shown in the representations devised by the students themselves indicates an aspect of problem solving which is not often encouraged in the classroom in relation to chance and data. Although full marks might not be given to a graph such as shown in Figure 2 if produced by a senior student who had completed a course in statistics, it shows considerable ability to represent an association for one with no training.

The fact that some students did not progress past categorising the information with respect to one variable, while others explored cause-effect relationships, would indicate that the activity is appropriate for classes of students of varying abilities. All students expressed satisfaction with the activity and the products of their labours. The lower ability students were not frustrated by having to cope with associations which were too complex for them. The student who presented a simple one-way table was obviously proud of what she produced and understood the categorisation she described. The acceptance and valuing of this student's work does not mean that other students were not encouraged to consider more complex relationships than those with which they may have begun. Some students responded positively to suggestions while others chose to stay with their original decisions in terms of variables studied. The openness of the activity and the freedom allowed in interpreting the task generated a sense of success in all students and a feeling of satisfaction in the teachers involved.

The group work involved in the data cards activity was purposely imposed in an informal fashion. Groups were formed by the children and could be classified as friendship groups. All were single sex. In the end they varied in size from one to five, with only one group at each extreme. One girl worked by

Figure 6 Bar-Chart Representation



herself very effectively and surprised her classroom teacher with the quality of her work compared with normal mathematics activities. The group of five boys who worked together produced the lowest quality output but at least there was some output associated with the task, which was not always the case for these five boys in normal classroom activities.

While statisticians might criticise some of the posters presenting the results of the students' investigations, it must be remembered that in all cases this represented a first attempt at data analysis of this type. Whereas some may argue that the ability to draw sophisticated inferences does not develop until students are in later grades, there is ample evidence from these students that it is possible at this level to appreciate an inferential task and for most students to gain valuable experience in taking part in one. Laying the groundwork for inference and creating the expectation in students that there is some purpose in "analysing" data, will improve student learning and make the teaching at higher levels much more rewarding. It was clear that many of these Grade 6 students appreciated the idea of association between variables. There was also an excellent discussion of bias as a result of the poster in Figure 6. The idea that the data were representative but did not tell the whole story was discussed as was the claim of "proof' in one poster. It is only by experiencing and discussing these occurrences that students build understanding of how valid conclusions in statistics are made.

An open-ended data-driven activity such as this puts pressure on teachers because they cannot predict what output will be produced. If the higher level objectives of curriculum documents such as the NCTM *Standards* are to be met, however, teachers will need to take on the challenge of these activities. Support and professional development will be required to assist teachers to achieve confidence in this area. There is still a long way to go to achieve the higher order functioning desired in data handling.

Acknowledgement
The authors would like to thank Jonathan B. Moritz for assistance with videotaping, editing, and analysis.

References

Australian Education Council. (1991). *A national statement on mathematics for Australian schools.* Carlton, Vic.: Curriculum Corporation.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

.