Acknowledgements

The Project directors acknowledge the vital support and contributions of many people to this report, including:

- the very dedicated staff of the Educational Assessment Research Unit
- Lisa Rodgers and other staff members of the Ministry of Education
- members of the Project's National Advisory Committee
- members of the Project's Science Advisory Panel
- principals and children of the schools where tasks were trialled
- principals, staff, and Board of Trustee members of the 252 schools included in the 2003 sample
- the 2936 children who participated in the assessments and their parents
- the 100 teachers who administered the assessments to the children
- the 45 senior tertiary students who assisted with the marking process
- the 170 teachers who assisted with the marking of tasks early in 2004.

Summary

New Zealand's National Education Monitoring Project commenced in 1993, with the task of assessing and reporting on the achievement of New Zealand primary school children in all areas of the school curriculum. Children are assessed at two class levels: year 4 (halfway through primary education) and year 8 (at the end of primary education). Different curriculum areas and skills are assessed each year, over a four-year cycle. The main goal of national monitoring is to provide detailed information about what children can do so that patterns of performance can be recognised, successes celebrated, and desirable changes to educational practices and resources identified and implemented.

Each year, small random samples of children are selected nationally, then assessed in their own schools by teachers specially seconded and trained for this work. Task instructions are given orally by teachers, through video presentations, on laptop computers, or in writing. Many of the assessment tasks involve the children in the use of equipment and supplies. Their responses are presented orally, by demonstration, in writing, in computer files, or through submission of other physical products. Many of the responses are recorded on videotape for subsequent analysis.

The use of many tasks with both year 4 and year 8 students allows comparisons of the performance of year 4 and 8 students in 2003. Because some tasks have been used twice, in 1999 and again in 2003, trends in performance across the four-year period can also be analysed.

ASSESSING SCIENCE

In 2003, the first year of the third cycle of national monitoring, three areas were assessed: science, art, and the use of graphs, tables and maps. This report presents details and results of the assessments in science. The aims of a science education include the development of knowledge and understanding, skills of scientific investigation, and attitudes on which such investigation depends. A framework for science education and its assessment is presented in **Chapter 2**. This framework highlights the four main content strands of the science curriculum (the living world, physical world, material world, and planet Earth and beyond), and also indicates important scientific approaches, skills and attitudes.

Most students responded with considerable enthusiasm to tasks involving hands-on experimentation, as individuals or as teams. As in the earlier science assessments they did not perform as well in planning experimental work as they did while carrying out their investigations. Recognition of the value of replication (repeating experimental steps, observations and measurements) remains quite limited.

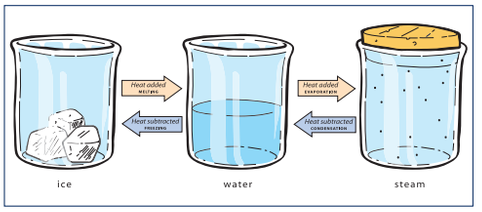
LIVING WORLD

Chapter 3 examines achievement relating to the living world curriculum strand. Averaged across 224 task components used with both year 4 and year 8 students, 10 percent more year 8 than year 4 students produced correct responses. This indicates that, on average, students have made useful progress between year 4 and year 8 in the skills assessed by the tasks. Not surprisingly, students at both levels were less successful in providing explanations for living world phenomena than in demonstrating their knowledge of the phenomena or in their ability to classify and identify observable features of living world phenomena.

Seven trend tasks involving a total of 74 components were administered to year 4 students in both the 1999 and 2003 assessments. Averaged across the 74 components, one percent fewer students succeeded in 2003 than in 1999. This difference clearly is not important. Six trend tasks involving 65 task components were administered to year 8 students in both the 1999 and 2003 assessments. Averaged across the 65 components, one percent more students succeeded in 2003 than 1999. Again, this difference clearly is not important.

MATERIAL WORLD

Chapter 5 reports achievement relating to the material world curriculum strand.

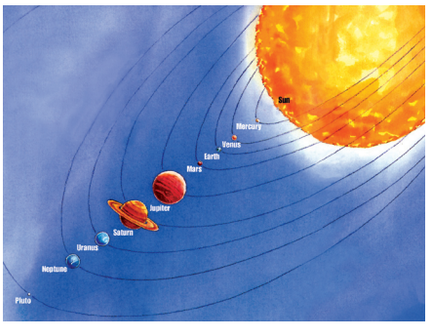


Averaged across 99 task components used with both year 4 and year 8 students, 15 percent more year 8 than year 4 students produced correct responses. Once again, the largest gains generally occurred for task components requiring explanations of phenomena.

Three trend tasks involving a total of 22 components were administered to year 4 students in both the 1999 and 2003 assessments. Averaged across the 22 components, two percent fewer students succeeded in 2003 than in 1999. This difference is not large enough to be judged important. Four trend tasks involving 31 task components were administered to year 8 students in both the 1999 and 2003 assessments. Averaged across the 38 components, four percent more students succeeded in 2003 than 1999.

PLANET EARTH AND BEYOND

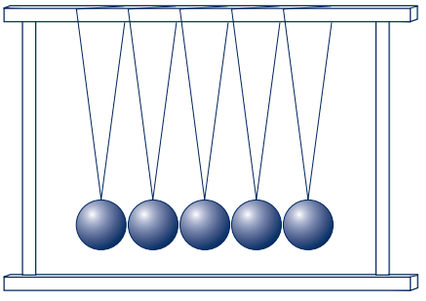
Chapter 6 examines achievement relating to the planet earth and beyond curriculum strand. Averaged across 115 task components used with both year 4 and year 8 students, 12 percent more year 8 than year 4 students produced correct responses.



Four trend tasks involving a total of 57 components were administered to year 4 students in both the 1999 and 2003 assessments. Averaged across the 57 components, the same percentage of students succeeded in 2003 as in 1999. Four trend tasks involving 62 task components were administered to year 8 students in both the 1999 and 2003 assessments. Averaged across the 62 components, three percent more students succeeded in 2003 than 1999.

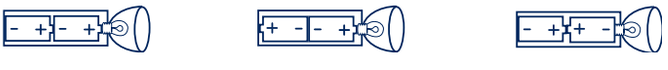
PHYSICAL WORLD

Chapter 4 examines achievement relating to the physical world curriculum strand.



Averaged across 91 task components used with both year 4 and year 8 students, 15 percent more year 8 than year 4 students produced correct responses. The largest gains generally occurred for task components requiring explanations of physical world phenomena.

Six trend tasks involving a total of 34 components were administered to year 4 students in both the 1999 and 2003 assessments. Averaged across the 34 components, four percent fewer students succeeded in 2003 than in 1999. Six trend tasks involving 38 task components were administered to year 8 students in both the 1999 and 2003 assessments. Averaged across the 38 components, one percent fewer students succeeded in 2003 than 1999. This difference clearly is not important.



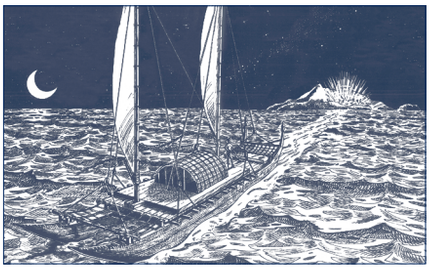
SURVEY

Chapter 7 presents the results of the science surveys, which sought information from students about their curriculum preferences and their perceptions of their achievement and potential in science. Students were also asked about their involvement in science related activities within school and beyond.

Students were asked to indicate their first three preferences from a list of six class science activities. Two activities ("doing things like experiments" and "going on field trips") were strong first preferences at both year levels, with year 4 regarding both similarly and year 8 strongly favouring experiments.

Year 4 students were generally very positive about doing science at school. A majority chose the highest rating for doing science at school and also would like to do more science at school. Almost half wanted to keep learning about science when they grew up. They were less confident that they learned a

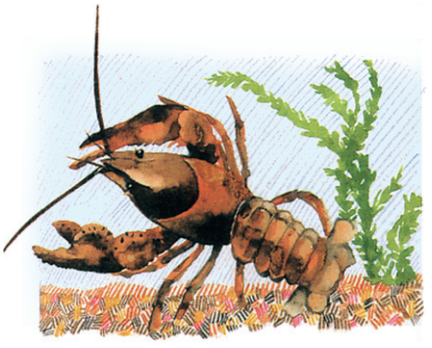
lot of science at school, with about one quarter saying that they learned “heaps” and only 12 percent saying that their class did really good things in science “heaps”. As in 1999, there were indications that much science in school is book work, with practical work, field trips, visits and experiments less common. There was a small but perceptible downward trend between 1999 and 2003 on most items involving science at school, suggesting that both the quantity and attractiveness of science activities may have declined a little over that period. On the other hand, there was a marked increase in the reported attractiveness of doing science things in their own time (perhaps in compensation).



Compared to year 4 students, year 8 students were less inclined to use the most positive categories. This pattern has been common in national monitoring surveys. It is a concern that only 26 percent of the year 8 students indicated that their class did really good things in science “heaps” or “quite a lot”. Only about one third of the year 8 students reported the use of experiments in science at school more than “sometimes”, with group work and projects dominating their experience of science. Changes between 1999 and 2003 were generally small, but there was a consistent decline in the proportions of highly positive attitudes to school science programmes, and in the reported frequencies of more exciting science activities such as experiments.

PERFORMANCE OF SUBGROUPS

Chapter 8 reports the results of analyses that compared the task performance and survey responses of different *demographic subgroups*. School type (full primary or intermediate), school size, community size and geographic zone did not seem to be important factors predicting achievement on the science tasks. However, there were statistically significant differences in the performance of students from low, medium and high decile schools on 65 percent of the tasks at both year levels (compared to the 1999 results, about 10 percent higher at year 4 but little changed at year 8).



For the comparisons of boys with girls, Pakeha with Māori, Pakeha with Pasifika students, and students for whom the predominant language at home was English with those for whom it was not, effect sizes were used. Effect size is the difference in mean (average) performance of the two groups, divided by the pooled standard deviation of the scores on the particular task.

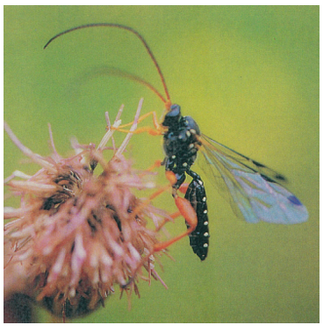
Averaged across all tasks, boys performed a little better than girls, with average effect sizes of 0.08 for year 4 students and 0.09 for year 8 students. In other words, on average boys score about one tenth of a standard deviation higher than girls – a very small advantage. There was a statistically significant advantage for boys on about 15 percent of year 4 tasks and 20 percent of year 8 tasks – somewhat lower percentages than in 1999, when both figures were about 30 percent.

In earlier NEMP reports, the performance of Māori students was compared to that of all other students. Starting with the 2003 reports, three groups are now distinguished: Māori students, Pasifika students, and all other students (described as Pakeha students). This change prevents satisfactory direct comparisons with the results from 1999. Pakeha students scored statistically significantly higher than Māori students on 40 percent of year 4 tasks and 64 percent of year 8 tasks, with average effect sizes of 0.30 and 0.37 respectively. These could be described as moderate effect sizes. Pakeha students scored statistically significantly higher than Pasifika students on 64 percent of year 4 tasks and 70 percent of year 8 tasks, with average effect sizes of 0.57 and 0.62 respectively. These can be described as large effect sizes.

Compared to students for whom the predominant language at home was English, students from homes where other languages predominated scored statistically significantly lower on 52 percent of year 4 tasks, reducing to 34 percent of year 8 tasks. The corresponding mean effect sizes were 0.37 and 0.31, which can be described as moderate effect sizes.

SUMMARY OF PERFORMANCE TRENDS

An indication of overall trends in performance across the four-year period between 1999 and 2003 can be obtained by looking at the patterns of change across the trend tasks for all four of the curriculum strands. Averaged across 186 components of the year 4 trend tasks, one percent fewer students succeeded in 2003 than in 1999. Averaged across 195 components of the year 8 trend tasks, two percent more students succeeded in 2003 than in 1999.



The 1999 science report had revealed trends between 1995 and 1999, with an average gain over that four-year period of one percent on year 4 trend task components, but no change on year 8 trend task components. Taken together, these two sets of trend results suggest little change in science performance overall for the eight year period from 1995 to 2003.