The Relative Achievement of Boys and Girls in New Zealand Primary Schools

Terry Crooks
Educational Assessment Research Unit
University of Otago
Box 56, Dunedin
New Zealand

Paper presented at the 2003 combined annual conference of the
Australian Association for Research in Education and the
New Zealand Association for Research in Education,
Auckland, December 2

Abstract

Over the past eight years, the National Education Monitoring Project has monitored the educational achievement and attitudes of year 4 and 8 students in New Zealand schools, covering 15 curriculum areas twice in four-year assessment cycles. This paper presents accumulated evidence from these assessments about the relative achievement of boys and girls. Averaged across the 15 subjects, boys and girls performed similarly at both year levels in both assessment cycles, with only modest differences in most subjects. The only area showing a strong difference was writing, with girls performing markedly better than boys at year 8 level, and in 1998 at year 4 level (the gap decreased for year 4 students in the assessments conducted in 2002).

This research was a component of the National Education Monitoring Project, which is organised by the Educational Assessment Research Unit, University of Otago, and funded by the New Zealand Ministry of Education.

Introduction

In his fortnightly website column on 17 October 2003, Member of Parliament Simon Power discussed the National Party's education discussion paper, *Schools of Excellence*, and stated that:

> *The crisis of confidence in boys' education is acknowledged. The gap between our boys' and girls' achievement is among the largest in the developed world. We [the National Party] are proposing a Commission of Inquiry to identify the changes in teaching styles, assessment and school management needed to lift our boys' achievement.*

These comments reflect widespread unease within the education profession, the media and the community, in New Zealand and elsewhere, about the school achievement of boys. Just a week ago, Massey University announced that one of its staff, Michael Irwin, is about to start the second phase of his work on boys' underachievement with a study on what motivates boys at school, and is also one of the organisers of a conference dedicated to boys' learning to be held at the Albany campus next July.

Much of the New Zealand achievement information which has led to these concerns about the achievement of boys relative to that of girls has come from national examinations in the latter years of secondary school. For instance, the Education Review Office Report entitled *The Achievement of Boys* (Education Review Office, 1999), drew very heavily on these examination results. This year, the release of the results for last year's NCEA assessments (New Zealand Qualification Authority website, May 2003) provoked considerable debate and concern about the relative achievement of boys and girls. For example, among NCEA candidates in their third year of secondary education, 63 percent of girls gained a Level 1 certificate, compared to 52 percent of boys. On average, girls gained 93 credits while boys gained 82 credits (with 80 required for a Level 1 certificate).

Apart from results from the National Education Monitoring Project (NEMP - see http://nemp.otago.ac.nz), comparatively little trustworthy information has been available on the comparative achievement of boys and girls in New Zealand primary schools. Most of that additional information in the past decade has come from New Zealand's participation in the Third International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS).

The latest TIMMS data (Chamberlain, 2001) showed that year 5 boys and girls performed comparably in mathematics, both in 1994 and 1998. Both gained slightly between 1994 and 1998, with the gain for boys large enough to be statistically significant. In science, exactly the same pattern applied.

The latest PIRLS data (Ministry of Education, 2003) showed very little change in performance, for year 5 boys or girls, between 1990 and 2001. In both years, girls scored statistically significantly higher than boys, with the 2001 results showing 55 percent of girls at or above the overall New Zealand median, compared to 45 percent of boys (a slightly wider gap than in 1990).

The most comprehensive analysis of the relative achievement of boys and girls in New Zealand was published by the Ministry of Education in 2000 (Alton-Lee & Praat, 2000). It used all information available at that time, including NEMP data, and tried to identify contextual variables that might account for the patterns reported. The picture presented is

a complex one, not justifying a simple conclusion that boys are performing poorly relative to girls in New Zealand schools.

This study uses the rich and diverse achievement data available from the NEMP assessments between 1995 and 2002 to explore in some detail the relative achievement of boys and girls at two levels in New Zealand Primary education: year 4 and year 8.

National Education Monitoring Project

Since 1995, the National Education Monitoring Project (NEMP) has provided detailed national assessments of the knowledge, skills and attitudes of primary and intermediate school students at two levels: year 4 (ages 8-9) and year 8 (ages 12-13). It identifies which aspects are improving, staying constant, or declining, allowing successes to be celebrated and priorities for curriculum change and teacher development to be debated. No information is provided about individual students or schools; the focus is on national performance patterns.

Nationally representative samples of approximately 480 students attempt each assessment task. A matrix sampling arrangement distributes three sets of tasks among 1440 students at each year level, so that more tasks can be used without excessive demands on each student. At year 8 level, special samples of 60 students who are learning in the Maori language attempt two of the sets of tasks, translated into Maori.

NEMP operates on a four-year cycle, covering about one quarter of the national curriculum areas for primary and intermediate schools each year. The areas covered are:

Year 1: science; art; graphs, tables and maps
Year 2: reading; speaking; music; technology
Year 3: mathematics; social studies; information skills
Year 4: writing; listening, viewing; health, physical education

In addition, some cross-curricular skills are assessed, such as co-operative skills.

The assessment tasks emphasise aspects of the curriculum which are likely to be of enduring importance to students, and cover a wide range of important skills, knowledge and understandings within the various curriculum strands (Flockton, 1999). They are designed, as far as possible, to interest students and stimulate them to do their best. The use of laptop computers to present video and audio material, and in some cases to record student responses, contributes to this goal, as does the heavy use of other hands-on equipment and resources. Task components vary widely in difficulty, so that all students can enjoy some success and some challenge.

About 45 percent of the tasks are kept constant from one cycle to the next. This re-use of tasks allows trends in achievement across a four-year interval to be observed and reported. The remaining tasks are released, making them available for teacher use and allowing detailed and clear reporting of students' responses.

Four different approaches are used, so that a wide-ranging picture of student capabilities can be built up. Students spend about an hour working in each approach.

- *One-to-one interview.* Each student works individually with a teacher, attempting 15 to 20 tasks, with the whole session recorded on videotape.
- *Team.* Four students work collaboratively, supervised by a teacher and recorded on videotape, on several longer tasks.

- *Stations.* Four students, work independently, moving around a series of stations where tasks have been set up.
- *Independent.* Four students work individually on tasks that involve paper-and-pencil tests or surveys, creating works of art, or demonstrating physical skills and having them videotaped.

In the first two of these approaches, most instructions are given orally and most responses are presented orally or by physical demonstration. This removes difficulties some students would have if they had to read instructions and respond in writing. Teachers also offer help with reading and writing in the other approaches. The team approach allows collegial support, more demanding tasks, and analysis of important social skills.

The tasks are administered to the students by experienced teachers who are relieved from their normal duties for six weeks, specially trained, and then work in pairs to administer the assessments to 60 students in several different schools over a five week period. The marking of the students' performances takes place after all task administration has been completed, with tasks requiring high levels of professional judgement marked by experienced teachers and other tasks marked by senior teacher education students. Each year, participation in task administration provides substantial, highly valued professional development opportunities for 100 teacher administrators, 160 to 200 teacher markers, and 45 student markers.

NEMP gathers very large amounts of data. Across the first four years of national monitoring, for instance, approximately 15,000 hours of video-recorded performances and 240,000 pages of paper responses (including art works) were gathered for marking.

Each year's results are published in four reports, multiple copies of which are sent to all schools and to agencies and individuals with a known interest in education. They are accompanied by an initial response to the reports from a national forum of educators, beginning debate about the meaning and implications of the reported results. Further analysis of the assessment methods, data, and issues raised occurs through financial provision for such work to be undertaken by NEMP staff and other New Zealand researchers, and through periodic reviews by international experts.

Numbers of Assessment Tasks

Table 1 indicates how many tasks were administered in each curriculum area for the first eight years of NEMP (1995-2002). Many tasks were identical for year 4 and year 8, some were adjusted to take account of age appropriateness for the two different year levels, and some were unique to year 4 or year 8. A total of 499 tasks were administered in cycle 1 (1995 - 1998), rising to 555 tasks in cycle 2 (1999-2002). On average, about 23 assessments tasks per subject were used in each year level of cycle 1, rising to about 31 tasks per subject in cycle 2.

Most tasks included several components that were marked separately. Selected components for each task were aggregated to get the overall task score that was used in the statistical analyses reported in this paper.

TABLE 1

Number of Assessment Tasks Administered

Subject Area	Year 4	Year 8	Total
Science, 1995	37	39	54
Art, 1995	11	11	16
Graphs, Tables, Maps, 1995	29	31	45
Reading, 1996	17	17	25
Speaking, 1996	13	13	18
Technology, 1996	15	16	22
Music, 1996	22	21	31
Mathematics, 1997	51	46	82
Social Studies, 1997	19	26	35
Information Skills, 1997	21	27	37
Writing, 1998	24	29	34
Listening, 1998	8	9	12
Viewing, 1998	11	14	19
Health, 1998	31	32	39
Physical Education, 1998	25	25	30
Cycle 1 Total	**334**	**356**	**499**
Science, 1999	56	54	70
Art, 1999	13	13	13
Graphs, Tables, Maps, 1999	33	38	51
Reading, 2000	17	19	19
Speaking, 2000	15	16	18
Technology, 2000	22	25	30
Music, 2000	28	28	29
Mathematics, 2001	78	94	101
Social Studies, 2001	36	41	49
Information Skills, 2001	21	28	35
Writing, 2002	29	35	36
Listening, 2002	14	17	18
Viewing, 2002	16	18	19
Health, 2002	31	39	43
Physical Education, 2002	22	24	24
Cycle 2 Total	**431**	**489**	**555**

Method and Results

This paper compares the performances of boys and girls on NEMP tasks administered to individual year 4 and year 8 students between 1995 and 2002. The results in Tables 2, 3, 4 and 5 use the most recent data, from cycle 2. Table 6 then compares the data from cycle 2 with those from cycle 1, to look for trends across time. The analyses cover 15 curriculum areas, and involve 48 national samples of students (8 assessment years, 2 grade levels, and 3 sub-samples of 480 students within each sample).

The first set of analyses focused on year 4 students, and used t-tests to compare the performance of boys and girls, task by task. Because the numbers of students included in the analyses was quite large (approximately 230 boys and 230 girls), the statistical tests were sensitive to small differences. To reduce the likelihood of attention being drawn to unimportant differences, the critical level for statistical significance was set at $p = .01$ so that differences this large or larger among the subgroups would not be expected by chance in more than one percent of cases. The results for individual year 4 assessment tasks were aggregated across all of the tasks for each curriculum area and presented as percentages in Table 2.

TABLE 2

For each subject in NEMP cycle 2 (1999-2002), the percentages of Year 4 tasks on which girls scored statistically significantly higher than boys (G>), there was no statistically significant difference between boys and girls (=), or boys scored statistically significantly higher than girls (B>)

Subject	G>	=	B>
Science	0	72	28
Phys. Ed.	23	27	50
Mathematics	0	88	12
Technology	0	89	11
Social Studies	7	76	17
Graphs/Tables/Maps	0	94	6
Info. Skills	0	100	0
Viewing	6	94	0
Health	11	89	0
Listening	14	86	0
Art	15	85	0
Music	17	83	0
Writing	39	61	0
Reading	53	47	0
Speaking	54	46	0
Mean for Cycle	**14**	**78**	**8**

The results in Table 2 show boys performing slightly better than girls in five subjects, equally or almost equally in three subjects, slightly worse in four subjects, and markedly worse in three subjects (writing, reading, and speaking). Averaged across the 15 subjects, boys did better on 8 percent of tasks, no differently to girls on 78 percent of tasks, and worse than girls on 14 percent of tasks.

The second set of analyses paralleled the first in all respects, but used the assessment results at year 8 level. The results are presented in Table 3.

TABLE 3

For each subject in NEMP cycle 2 (1999-2002), the percentages of Year 8 tasks on which girls scored statistically significantly higher than boys (G>), there was no statistically significant difference between boys and girls (=), or boys scored statistically significantly higher than girls (B>)

Subject	G>	=	B>
Science	0	73	27
Phys. Ed.	26	31	43
Technology	13	70	17
Mathematics	4	93	3
Social Studies	9	85	6
Viewing	11	89	0
Reading	11	89	0
Speaking	14	86	0
Graphs/Tables/Maps	16	84	0
Music	17	83	0
Art	23	77	0
Info. Skills	28	72	0
Listening	29	71	0
Health	46	54	0
Writing	88	12	0
Mean for Cycle	**22**	**72**	**6**

The results in Table 3 show boys performing slightly better than girls in two subjects, equally or almost equally in three subjects, slightly worse in eight subjects, and markedly worse in two subjects (writing and health). Averaged across the 15 subjects, boys did better on 6 percent of tasks, no differently to girls on 72 percent of tasks, and worse than girls on 22 percent of tasks. In the cycle 2 assessments, therefore, girls had a somewhat larger advantage over boys at year 8 level than at year 4 level.

Another way of looking at the differences in task performance between boys and girls is to compute effect sizes for these differences and average across tasks. The statistical significance testing used in the previous analyses identified whether or not there were statistically significant differences in task performance between boys and girls, but did not describe the size of those differences. Effect sizes are ideal for the latter purpose.

Effect sizes were calculated for each task by subtracting the mean performance of girls from the mean performance of boys. This difference was then divided by the pooled standard deviation. An effect size of +.20 indicates that boys scored, on average, two tenths of a standard deviation higher than girls. Conversely, an effect size of -.50 indicates that boys scored, on average, one half of a standard deviation lower than girls.

The effect sizes for the individual assessment tasks were averaged across all of the tasks for each curriculum area, for both year 4 and year 8 students, and the mean effect sizes are presented in Table 4.

TABLE 4

Effect sizes for score differences between Year 4 and Year 8 boys and girls in each subject of the second NEMP cycle (1999-2002)

Subject	Year 4	Year 8
Science	+.15	+.14
Social Studies	+.05	-.03
Physical Education	+.15	+.10
Technology	+.03	+.04
Graphs/Tables/Maps	+.02	-.09
Mathematics	+.10	-.03
Health	-.09	-.17
Listening	-.12	-.19
Art	-.11	-.15
Viewing	-.05	-.07
Music	-.15	-.10
Information Skills	-.06	-.15
Speaking	-.24	-.06
Reading	-.26	-.09
Writing	-.24	-.40
Mean for Cycle	-.05	-.08

These effects sizes present a similar picture to the statistical significance results in Tables 2 and 3. Overall, the gender gap favouring girls was slightly larger at year 8 than at year 4 level, but both average effect sizes were small. The only effect sizes large enough to be noteworthy all favoured girls: for writing, reading and speaking at year 4 level and particularly for writing at year 8 level.

It is common practice where mean effect sizes are based on a substantial and diverse data set to interpret them using percentiles derived from the normal curve. An effect size of +1.00 is interpreted as meaning that an average boy is performing at a level equivalent to the 84th percentile of the performance distribution for girls. In other words, an average boy is doing as well or better than 84 percent of girls. Another example is an effect size of -.20, which would be interpreted as meaning that an average boy is performing at the 42nd percentile of the performance distribution for girls (ie. as well or better than 42 percent of girls).

Following this practice, Table 5 converts the mean effect sizes in Table 4 into percentile ranks, estimating where the average performance for boys would rank in the distribution of performances of girls.

TABLE 5

For each subject in NEMP cycle 2 (1999-2002), mean scores for boys as percentiles of the score distributions for girls

Subject	Year 4	Year 8
Science	56	56
Physical Education	56	54
Mathematics	54	49
Social Studies	52	49
Technology	51	52
Graphs/Tables/Maps	51	46
Viewing	48	47
Information Skills	48	44
Art	46	44
Health	46	43
Listening	45	42
Music	44	46
Speaking	41	48
Writing	41	34
Reading	40	46

In most cases, an average boy scored at a level between the 56th and 43rd percentiles for girls. For writing, reading and speaking at year 4 level, however, the average boy was at about the 40th percentile for girls. The gap was much narrower for reading and speaking at year 8 level, but the writing gaps was wider, with an average boy at the 34th percentile for girls. Such a substantial gap certainly justifies careful consideration.

Table 6 presents again the effect sizes from Table 4, based on the results for the second cycle of NEMP assessments (1999-2002), but this time also with the corresponding effect sizes from the first cycle of NEMP assessments (1995-1999). This allows comparison, for both year 4 and year 8 students, of any changes in the relative performance of boys and

girls over the four-year intervals between the assessments for each subject. Also, it allows comparison of the results for the year 4 students in cycle 1 and the year 8 students in cycle 2, which is of articular interest since this is the same cohort of students assessed twice, four years apart.

TABLE 6

Effect sizes for score differences between boys and girls on year 4 and year 8 tasks in each subject, for both the first NEMP cycle (1995-1998) and second NEMP cycle (1999-2002)

Subject	Year 4 Cycle 1	Year 4 Cycle 2	Year 8 Cycle 1	Year 8 Cycle 2
Science	+.08	+.15	+.15	+.14
Social Studies	+.13	+.05	+.13	-.03
Physical Education	+.12	+.15	+.12	+.10
Technology	-.05	+.03	-.06	+.04
Graphs/Tables/Maps	-.07	+.02	-.04	-.09
Mathematics	+.01	+.10	-.06	-.03
Health	-.07	-.09	-.09	-.17
Listening	-.07	-.12	-.07	-.19
Art	-.05	-.11	-.05	-.15
Viewing	-.11	-.05	-.11	-.07
Music	-.11	-.15	-.20	-.10
Information Skills	-.15	-.06	-.16	-.15
Speaking	-.16	-.24	-.17	-.06
Reading	-.17	-.26	-.21	-.09
Writing	-.37	-.24	-.41	-.40
Mean for Cycle	-.07	-.05	-.08	-.08

Looking first at the year 4 results, there is evidence that the small gender gap overall in cycle 1 had narrowed further in cycle 2. Boys gained significant ground in writing, and a little in science, technology, graphs/tables/maps, mathematics, and information skills. At the same time, they lost a little ground in social studies, speaking and reading.

The small gender gap had not changed at year 8 level between cycles 1 and 2. Boys lost a little ground in social studies, health, listening and art, and gained a little in technology, music, speaking and reading.

It is particularly interesting to compare the year 4 results for cycle 1 with the year 8 results for cycle 2 (first and last columns): the same cohort four years apart. Overall, the small gender gap was virtually unchanged, with boys gaining a little in science, technology, speaking and reading, and losing a little in social studies, health, listening and art.

Discussion

The results provide strong evidence, based on NEMP assessment results in 15 different subject areas over an eight-year period, that boys and girls are achieving quite similarly in primary schooling in New Zealand.

On average across all 15 subjects, a boy at the 50th percentile for boys was performing at the 47th percentile for girls (ie. as well or better than 47 percent of girls) in both the year 4 and year 8 assessments conducted between 1995 and 1998, and in the year 8 assessments conducted between 1999 and 2002. An even narrower gap was recorded for the year 4 assessments conducted between 1999 and 2002: in these, a boy at the 50th percentile for boys was performing at the 48th percentile for girls.

In the latest cycle of assessments (1999-2002), boys scored slightly better than girls on 6 of the 15 subjects at year 4 level and 3 of the 15 subjects at year 8 level.

Only one subject showed differences at both year 4 and year 8 levels large enough to raise major concern: girls performed substantially better than boys in writing. The latest results for writing at year 4 level showed the gap narrowing significantly compared to the earlier results at year 4 level (and to both results at year 8 level), but still remaining quite large.

The only other subjects with effect sizes larger than 0.20 (which is equivalent to the average boy being at the 42nd percentile for girls) were year 8 reading in 1996 and year 4 reading and speaking in 2000.

Taken together, these results suggest that current professional and public concern about the poor achievement of boys relative to girls is not justified for boys at primary school level, except perhaps in regard to the development of writing skills. Much of the evident concern has arisen from the comparative results of boys and girls in secondary school qualifications, yet has been generalised to suggest problems with the achievement of boys at all levels of our education system. While there will always be scope to improve the overall quality of education offered in our schools, and to tailor it more appropriately to the needs of particular subgroups, wholesale changes in primary schools designed to address the perceived learning and motivational needs of boys do not seem to be necessary.

References

Alton-Lee, A. & Praat, A. (2000). *Explaining and addressing gender differences in the New Zealand compulsory school sector: A literature review*. Wellington: Ministry of Education.

Chamberlain, G. (2001). *Trends in year 5 students' mathematics and science achievement [Results from TIMMS3]*. Wellington: Ministry of Education.

Education Review Office. (1999). *The achievement of boys*. Wellington: ERO.

Ministry of Education (2003). *Reading literacy in New Zealand [Results from PIRLS]*. Wellington: Ministry of Education.

New Zealand Qualifications Authority (2003). *2002 NCEA statistics.* Wellington: NZQA.