APPENDIX A: Background & Sampling

A.1 NATURE AND WORK OF THE INTERNATIONAL ASSOCIATION FOR THE EVALUATION OF EDUCATIONAL ACHIEVEMENT (IEA)

The International Association for the Evaluation of Educational Achievement (IEA) is a cooperative of leading educational research institutions from countries around the world. With headquarters in Amsterdam, its membership currently numbers more than 50.

In New Zealand, the Ministry of Education is the member institution of IEA. IEA projects are undertaken by the ministry's Research Division.

New Zealand participated in its first IEA study — the Six Subject Survey (Science, Reading Comprehension, Literature, French as a Foreign Language, English as a Foreign Language, Civics Education) — in the early 1970s. Other studies in which New Zealand has taken part and the year of data collection include:

- Second International Mathematics Study (SIMS), 1981;
- Written Composition Study, 1984;
- Computers in Education; 1989;
- Reading Literacy Study; 1990;
- Third International Mathematics and Science Study (TIMSS), 1994 & 1995;
- Third International Mathematics and Science Study - Repeat (TIMSS-R), 1998
- Second Information Technology in Education Study (SITES), 1998;

New Zealand is currently participating in Progress in International Reading Literacy Study (PIRLS), with the main data collection in 2001.

A.2 SAMPLING PROCEDURES USED IN TIMSS-98/99 IN NEW ZEALAND

Population definition for TIMSS 98/99

The international desired population for TIMSS-98/99: *All students enrolled in the upper of the two grades that contain the largest proportion of students of age 13 years (at the time of testing).*

In New Zealand, the national *desired* population was all Year 9 (or form 3) students in October 1998.

Exclusions

In TIMSS-94/95, under certain conditions a country was permitted to exclude up to 10 percent of their total national sample (or design sample). Since most countries achieved exclusion rates of five percent or less, the exclusion rate for TIMSS-98/99 was set at five percent. There were two types of exclusions:

- *School level exclusions* — schools could be excluded for the following reasons: they were special education schools, were very small, they offered a curriculum that was different from the mainstream education system, or were geographically isolated schools.

The resulting population after these exclusions have been made is defined as the national *defined* population.

There was another level to the exclusions, but this was implemented further on in the student selection phase:

- *Within-sample exclusions* — included students that had a physical disability which would prevent them undertaking the test in the specified time; students who were notably intellectually or developmentally delayed; or students that had only limited instruction in the language of the test. In New Zealand's case, the language of the test was English.

Sample design

The sample design used in TIMSS-94/95 and TIMSS-98/99 is generally referred to as a two-stage stratified cluster design. This involved:
- firstly, selecting a sample of schools from all eligible schools; and
- secondly, randomly selecting a mathematics (or science) class from each sampled school, regardless of the ability level of the class.

It was possible to sub-sample students from the selected class if the class was very large, but in New Zealand's case as indeed most countries, it was more straightforward to involve all students within the sampled class.

The minimum effective student sample size to be achieved by each country was 400. That is, the achieved samples were to yield sampling errors no greater than those obtained if countries were to select simple random samples of 400 Year 9 (equivalent) students. Although technically the stratified cluster design approach is a less efficient method of sampling, it does mean than school- and teacher information can be collected. Note that the figure of 400 students give estimates of student population parameters within plus or minus five percent. For a more detailed explanation of the sampling design and procedures the reader should refer to Foy and Joncas (2000), referenced at the end of this report, or go to www.timss.org.

What is the effect on the sampling when students are grouped by ability?

There is tendency for students in classes (particularly in mathematics classes) to be more homogeneous than students randomly selected across a school or from the general population. The coefficient of intra-class correlation (*IC*) is a measure of this effect and was built in as a factor when the sample sizes were determined. As the value of the *IC* increases, the sample size required increases to obtain the equivalent levels of sampling precision. For sampling for TIMSS-94/95, the *IC* for New Zealand was set at 0.3. Using information generated from TIMSS-94/95, it was established that the tendency of grouping students by ability in New Zealand was greater than anticipated — the *IC* computed for New Zealand for mathematics was 0.357. This meant that for TIMSS-98/99 New Zealand was required to sample more classes in 1998 than was the case in 1994 (ie, 156 in 1998 compared with 150 in 1994).

Who managed the sampling?

Each participating country was responsible for developing its own sampling plan, in accordance with the guidelines developed for the study. Statistics Canada was responsible for overseeing the sampling process in each country and computing the sampling weights for each country. The final sampling plans went to an independent sampling referee — Dr Keith Rust, Westat, Inc. — for final approval.

Implications of the sample design

It is important to note that in both TIMSS-94/95 and TIMSS-98/99 the achieved student samples have not come from using simple random sampling techniques. Therefore, it is not appropriate to use statistical formulae that are used when working with simple random samples. For example: it is incorrect to used to estimate the standard error of a sample mean X using the formula $se(X) = S/\sqrt{(n)}$. If this formula is used it will result in a large under-estimation of the standard errors. Instead, the study employs a (numerical) method called *Jackknife Repeating Replication (JRR)* for estimating standard errors.

Also, while the achieved sample of students and schools are representative of the population from which they were drawn, the teachers who took part in TIMSS-98/99 (and in TIMSS-94/95) are not representative sample of teachers. Rather, they are the mathematics and science teachers who teach a representative sample of students and are teaching in a representative sample of schools.

School and student-level sampling in New Zealand

Essentially the sampling approach used in New Zealand in 1994 was used in 1998. The following school type were excluded from the sampling frame:

- special education schools;
- the Correspondence School;
- Kura kaupapa Maori schools (schools where Maori language, culture and values predominate);
- Rudolf Steiner schools; and
- Very small schools with less than 13 Year 9 students.

All remaining schools were included in the sample frame. A sample of schools containing students in New Zealand's *defined* target population was drawn from this frame. Schools were then explicitly stratified on one criterion: whether or not they were 'certain to be selected' (due to their size). Schools not fitting into the 'certainty' category were then implicitly stratified (or sorted) to ensure population coverage in the following areas:

School authority

1) state (including state-integrated)
   a. the school type
      i single-sex girls
      ii single-sex boys
      iii co-educational
   b. Target Funding for Educational Achievement (TFEA) indicator as measure of socio-economic status of the school community
   c. urban or rural locality
2) independent (private)

Using the number of students at the target form level, **schools** were randomly selected with probability proportional to the number of Year 9 students.

The **mathematics groupings** within each school were used as the basis for selecting the student sample at the school level.

One mathematics group or class was **randomly** selected regardless of the ability of the class. All students in that class were selected to take part in TIMSS-98/99. Under specific international criteria mentioned on page 108, students that had insufficient English or a physical or intellectual disability could be excluded from the testing. The decision as to whether or not students be excluded was made by individual schools.

Teacher selection

The mathematics teacher who taught the mathematics group or class was selected to take part. Similarly, the science teacher(s) linked to the selected students were invited to take part.

Sampling summary

TABLE A.1: A SUMMARY OF NEW ZEALAND'S DESIGNED AND ACHIEVED YEAR 9 STUDENT SAMPLES IN 1998

School sampling and final Year 9 student numbers for New Zealand			
Explicit strata	Total number of schools in original sample (N)	Total number of participating schools (N)	Total number of Year 9 students assessed (N)^
Certainty Schools*	16	14	351
Other Schools	140	138+	3262
Total	156	152	3613

* *The certainty schools could not be replaced if they declined to participate in the study.*

\+ *Included 7 replacement schools.*

^ *Participation in at least one achievement session. The number of sampled students was 3966. A total of 118 were excluded leaving 3848 students eligible for the assessment. However, 235 were absent from school on the day of testing.*

APPENDIX B: Reference Tables for Chapter 3

Note that the 1994 achievement figures presented in these tables are based on the new re-scaled achievement data calculated by Educational Testing Service in the United States. The figures (eg, means for New Zealand girls and boys) will appear inconsistent with figures reported in TIMSS-94/95 publications such as Beaton et al (1996b).

TABLE B.1: PERCENTILES FOR YEAR 9 STUDENTS' MATHEMATICS ACHIEVEMENT IN 1994 AND 1998

Year	Percentiles				
	5th	25th	50th	75th	95th
1994	359	447	504	557	634
1998	341	430	493	554	632

TABLE B.2: YEAR 9 STUDENTS' MEAN MATHEMATICS SCORES IN 1994 AND 1998, BY GENDER

Year 9 students	1994 mean mathematics score (se)	1998 mean mathematics score (se)	1998 standard deviations (se)
Girls	497 (5.3)	495 (5.5)	87 (2.9)
Boys	505 (6.1)	487 (7.6)	91 (3.0)
Overall	501 (4.7)	491 (5.2)	89 (2.3)

TABLE B.3: PERCENTILES FOR YEAR 9 STUDENTS' MATHEMATICS ACHIEVEMENT IN 1994 AND 1998, BY GENDER

Year	Gender	Percentiles				
		5th	25th	50th	75th	95th
1994	Girls	361	445	500	551	620
	Boys	357	449	508	562	646
1998	Girls	345	435	497	556	632
	Boys	337	425	488	551	630

TABLE B.4 *t* STATISTIC FOR COMPARING DIFFERENCES BETWEEN YEAR 9 STUDENTS' MEAN MATHEMATICS SCORES IN 1998, BY ETHNIC GROUPING

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		7.59	6.81	–2.33
Maori	–7.59		2.16	–7.20
Pacific	–6.81	–2.16		–7.29
Asian	2.33	7.20	7.29	

TABLE B.5A MEAN EFFECT SIZES FOR YEAR 5 STUDENTS' MATHEMATICS ACHIEVEMENT IN 1994, BY ETHNIC GROUPING

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		0.73	0.95	0.11
Maori			0.16	–0.56
Pacific				– 0.75

TABLE B.5B. MEAN EFFECT SIZES FOR YEAR 9 STUDENTS' MATHEMATICS ACHIEVEMENT IN 1994, BY ETHNIC GROUPING

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		0.71	1.15	–0.19
Maori			0.43	–0.87
Pacific				– 1.28

TABLE B.5C MEAN EFFECT SIZES FOR YEAR 9 STUDENTS' MATHEMATICS ACHIEVEMENT IN 1998, BY ETHNIC GROUPING

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		0.66	0.96	– 0.32
Maori			0.29	– 0.94
Pacific				– 1.12

$d < 0.35$	Difference between means is small
$0.35 \leq d \leq 0.75$	Difference between mean is of medium-size
$d > 0.75$	Difference between means is large

TABLE B.6: PERCENTILES FOR YEAR 9 STUDENTS' MATHEMATICS ACHIEVEMENT IN 1994 AND 1998, BY ETHNIC GROUPING

Ethnic grouping	Year	Percentiles				
		5th	25th	50th	75th	95th
Pakeha/ European	1994	390	468	520	568	642
	1998	372	452	511	566	636
Maori	1994	334	412	466	518	585
	1998	313	400	454	510	591
Pacific	1994	306	376	430	482	557
	1998	297	366	419	486	600
Asian	1994	394	477	536	588	677
	1998	372	475	543	600	667

TABLE B.7: YEAR 9 STUDENTS' MEAN MATHEMATICS SCORES IN 1994, BY ETHNIC GROUPING AND GENDER

1994 assessment	Mean mathematics scale scores (se)							
	Pakeha/European		Maori		Pacific Islands		Asian	
	girls	boys	girls	boys	girls	boys	girls	boys
	513 (5.0)	521 (5.7)	464 (8.5)	462 (7.5)	431 (9.2)	429 (9.9)	525 (10.8)	542 (18.2)
Overall mean	517 (4.5)		463 (6.4)		430 (6.8)		532 (10.9)	

TABLE B.8: YEAR 9 STUDENTS' MEAN MATHEMATICS SCORES IN 1994, BY TFEA DECILE BAND

Schools' TFEA decile band	1994	
	% of students	Mean mathematics score (se)
Low (deciles 1-3)	25	463 (7.7)
Medium (deciles 4-7)	50	497 (5.6)
High (deciles 8-10)	21	547 (7.9)
No TFEA indicator (ie, private schools)	4	534 (31.8)

TABLE B.9: PERCENTILES FOR YEAR 9 STUDENTS' MATHEMATICS ACHIEVEMENT IN 1998, BY TFEA DECILE BAND

Schools' TFEA decile band	Percentiles				
	5th	25th	50th	75th	95th
Low (deciles 1-3)	311	395	453	515	592
Medium (deciles 4-7)	337	422	479	532	605
High (deciles 8-10)	386	472	537	591	656

TABLE B.10: *t* STATISTIC FOR COMPARING DIFFERENCES BETWEEN YEAR 9 STUDENTS' MEAN MATHEMATICS SCORES IN 1998, BY TFEA DECILE BAND

Schools' TFEA decile band	Low (deciles 1-3)	Medium (deciles 4-7)	High (deciles 8-10)	No TFEA indicator
Low (deciles 1-3)		–1.54	–4.75	–5.63
Medium (deciles 4-7)	1.54		–4.74	–5.76
High (deciles 8-10)	4.75	4.74		–1.39
No TFEA indicator	5.63	5.76	1.39	

TABLE B.11: *t* STATISTIC FOR COMPARING DIFFERENCES BETWEEN YEAR 9 STUDENTS' MEAN MATHEMATICS SCORES IN 1998, BY SCHOOL TYPE

School type	Coeducational	Single-sex boys'	Single-sex girls'
Coeducational		–1.42	–3.45
Single-sex boys'	1.42		–0.74
Single-sex girls'	3.45	0.74	

TABLE B.12: YEAR 9 STUDENTS' MEAN MATHEMATICS SCORES IN 1994, BY SCHOOL TYPE

School type	1994	
	% of students	Mean mathematics score (se)
Coeducational	68	494 (5.7)
Single-sex boys'	18	526 (13.8)
Single-sex girls'	14	502 (12.3)

TABLE B.13: YEAR 9 STUDENTS' MEAN SCORES FOR EACH MATHEMATICS CONTENT REPORTING CATEGORY IN 1998, BY GENDER

Content reporting category	Mean scale scores (se)		
	girls	boys	overall
Fractions & Number Sense	496 (5.6)	490 (6.9)	493 (5.0)
Measurement	494 (5.3)	498 (7.4)	496 (5.3)
Data Representation, Analysis & Probability	502 (7.0)	492 (6.6)	497 (5.0)
Geometry	481 (8.3)	474 (6.7)	478 (4.2)
Algebra	506 (5.6)	487 (6.4)	497 (4.7)

TABLE B.14: TRENDS IN MEAN PERCENT CORRECT IN MATHEMATICS CONTENT REPORTING CATEGORIES FOR 1994 AND 1998*

Year	Mean percent correct scores (se)				
	Fractions & Number Sense	Measurement	Data Rep, Analysis & Probability	Geometry	Algebra
	(17 items)	(6 items)	(8 items)	(6 items)	(11 items)
1994	65 (1.0)	66 (1.2)	70 (1.0)	55 (1.3)	60 (1.2)
1998	63 (1.2)	65 (1.3)	69 (1.3)	51 (1.4)	60 (1.5)

**Based on items common to both assessments.*

SOURCE: MULLIS ET AL, 2000

TABLE B.15: YEAR 9 STUDENTS' MEAN SCORES FOR EACH MATHEMATICS CONTENT REPORTING CATEGORY IN 1998, BY ETHNIC GROUPING

Content reporting category	Mean scale scores (se)			
	Pakeha/ European	Maori	Pacific	Asian
Fractions & Number Sense	508 (4.9)	459 (5.8)	438 (8.1)	535 (11.2)
Measurement	511 (5.0)	466 (5.8)	442 (10.3)	527 (9.9)
Data Representation, Analysis, & Probability	513 (6.9)	466 (7.5)	439 (13.5)	527 (14.1)
Geometry	487 (5.2)	457 (7.2)	441 (14.0)	506 (12.0)
Algebra	508 (4.6)	465 (6.0)	456 (9.0)	543 (11.8)

APPENDIX C: Reference Tables for Chapter 4

Note that the 1994 achievement figures presented in these tables are based on the re-scaled achievement data calculated by the Educational Testing Service in the United States. The figures will appear inconsistent (eg, means for New Zealand girls and boys) with figures reported in TIMSS publications such as Beaton et al (1996a).

TABLE C.1: PERCENTILES FOR YEAR 9 STUDENTS' SCIENCE ACHIEVEMENT IN 1994 AND 1998

Year	Percentiles				
	5th	25th	50th	75th	95th
1994	358	453	514	571	654
1998	348	451	515	574	652

TABLE C.2: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994 AND 1998, BY GENDER

Year 9 students	1994 mean science score (se)	1998 mean science score (se)	1998 standard deviations (se)
Girls	497 (5.6)	506 (5.4)	90 (3.2)
Boys	524 (6.1)	513 (7.0)	96 (3.7)
Overall	511 (4.9)	510 (4.9)	93 (3.1)

TABLE C.3 PERCENTILES FOR YEAR 9 STUDENTS' SCIENCE ACHIEVEMENT IN 1994 AND 1998, BY GENDER

Year	Gender	Percentiles				
		5th	25th	50th	75th	95th
1994	Girls	351	444	501	555	627
	Boys	365	465	526	586	672
1998	Girls	346	452	512	568	643
	Boys	351	451	519	580	660

TABLE C.4 *t* STATISTIC FOR COMPARING DIFFERENCES BETWEEN YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1998, BY ETHNIC GROUPING

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		8.19	8.09	1.73
Maori	–8.19		3.15	–3.68
Pacific	–8.09	–3.15		–5.46
Asian	–1.73	3.68	5.46	

TABLE C.5A MEAN EFFECT SIZES FOR YEAR 5 STUDENTS' SCIENCE ACHIEVEMENT IN 1994, BY ETHNIC GROUPING

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		0.79	1.02	0.44
Maori			0.15	-0.32
Pacific				-0.50

TABLE C.5B MEAN EFFECT SIZES FOR YEAR 9 STUDENTS' SCIENCE ACHIEVEMENT IN 1994, BY ETHNIC GROUPING

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		0.75	1.26	0.42
Maori			0.51	-0.31
Pacific				-0.76

TABLE C.5C MEAN EFFECT SIZES FOR YEAR 9 STUDENTS' SCIENCE ACHIEVEMENT IN 1998, BY ETHNIC GROUPING

Ethnic grouping	Pakeha/ European	Maori	Pacific	Asian
Pakeha/ European		0.73	1.22	0.22
Maori			0.46	-0.47
Pacific				-0.87

$d < 0.35$	Difference between means is small
$0.35 \leq d \leq 0.75$	Difference between mean is of medium-size
$d > 0.75$	Difference between means is large

TABLE C.6 PERCENTILES FOR YEAR 9 STUDENTS' SCIENCE ACHIEVEMENT IN 1994 AND 1998, BY ETHNIC GROUPING

Ethnic grouping	Year	Percentiles				
		5th	25th	50th	75th	95th
Pakeha/ European	1994	397	480	533	587	666
	1998	393	480	537	590	663
Maori	1994	336	416	474	529	601
	1998	324	416	476	530	615
Pacific	1994	296	370	428	488	571
	1998	267	361	428	499	604
Asian	1994	336	440	500	561	647
	1998	356	449	524	583	659

TABLE C.7 YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994, BY ETHNIC GROUPING AND GENDER

1994 assessment	Mean science scale scores (se)							
	Pakeha/European		Maori		Pacific		Asian	
	girls	boys	girls	boys	girls	boys	girls	boys
	519 (5.2)	545 (5.5)	466 (7.4)	478 (6.3)	418 (11.2)	442 (9.8)	485 (11.6)	515 (17.6)
Overall mean	533 (4.4)		472 (5.6)		430 (8.5)		498 (12.0)	

TABLE C8: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994, BY TFEA DECILE BAND

Schools' TFEA decile band	1994	
	% of students	Mean science score (se)
Low (deciles 1-3)	25	470 (8.0)
Medium (deciles 4-7)	50	508 (5.9)
High (deciles 8-10)	21	556 (8.8)
No TFEA indicator (ie, private schools)	4	554 (22.0)

TABLE C.9 PERCENTILES FOR YEAR 9 STUDENTS' SCIENCE ACHIEVEMENT IN 1998, BY TFEA INDICATOR BAND

Schools' TFEA decile band	Percentiles				
	5th	25th	50th	75th	95th
Low (deciles 1-3)	307	412	475	537	615
Medium (deciles 4-7)	341	444	504	559	637
High (deciles 8-10)	400	489	550	600	678

TABLE C.10: *t* STATISTIC FOR COMPARING DIFFERENCES BETWEEN YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1998, BY TFEA DECILE BAND

Schools' TFEA decile band	Low (deciles 1-3)	Medium (deciles 4-7)	High (deciles 8-10)	No TFEA indicator
Low (deciles 1-3)		–2.12	–5.49	–6.21
Medium (deciles 4-7)	2.12		–4.59	–5.49
High (deciles 8-10)	5.49	4.59		–1.22
No TFEA indicator	6.21	5.49	1.22	

TABLE C.11: *t* STATISTIC FOR COMPARING DIFFERENCES BETWEEN YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1998, BY SCHOOL TYPE

School type	Coeducational boys	Coeducational girls	Single-sex boys'	Single-sex girls'
Coeducational boys		1.42	–1.35	–2.11
Coeducational girls	–1.42		–3.01	–3.13
Single-sex boys'	1.35	3.01		–0.07
Single-sex girls'	2.11	3.13	0.07	

TABLE C.12: YEAR 9 STUDENTS' MEAN SCIENCE SCORES IN 1994, BY SCHOOL TYPE

School type	1994	
	% of students	Mean science score (se)
Coeducational	68	503 (5.9)
Single-sex boys'	18	546 (12.8)
Single-sex girls'	14	504 (12.9)

TABLE C.13: YEAR 9 STUDENTS' MEAN SCORES FOR SCIENCE CONTENT REPORTING CATEGORIES IN 1998, BY GENDER AND OVERALL

Content reporting category	Mean scale scores (se)		
	girls	boys	overall
Earth Science	499 (8.6)	510 (7.9)	504 (5.8)
Life Science	506 (6.4)	496 (7.7)	501(5.6)
Physics	494 (4.9)	504 (6.4)	499 (4.7)
Chemistry	497 (7.6)	509 (6.3)	503 (4.9)
Environment & Resource Issues	499 (6.7)	506 (6.6)	503 (5.2)
Scientific Inquiry & Nature of Science	530 (6.6)	513 (11.4)	521 (6.8)

TABLE C.14: YEAR 9 STUDENTS' MEAN SCORES FOR EACH SCIENCE CONTENT REPORTING CATEGORY IN 1998, BY ETHNIC GROUPING

Content reporting category	Mean scale scores (se)			
	Pakeha/ European	Maori	Pacific	Asian
Earth Science	521 (6.6)	477 (8.9)	452 (10.4)	513 (10.3)
Life Science	524 (5.6)	466 (7.7)	429 (12.5)	499 (10.2)
Physics	517 (4.8)	465 (6.1)	441 (11.5)	516 (12.1)
Chemistry	522 (6.1)	476 (6.6)	439 (11.7)	497 (13.9)
Environment & Resource Issues	523 (5.2)	475 (7.3)	433 (14.9)	496 (18.2)
Scientific Inquiry & Nature of Science	538 (8.3)	497 (6.6)	469 (12.6)	524 (10.5)

APPENDIX D: Reference Tables for Chapter 5

TABLE D.1: MEAN ACHIEVEMENT SCORES FOR YEAR 9 STUDENTS IN 1994 REPORTING TO HAVE BEEN BORN IN NEW ZEALAND

Country of birth	Mean achievement scores in 1994		
	% of students	mathematics score (se)	science score (se)
Born in NZ	88	500 (4.6)	512 (4.6)
Not born in NZ	12	510 (8.1)	505 (9.5)

TABLE D.2: MEAN ACHIEVEMENT SCORES FOR YEAR 9 STUDENTS IN 1998 REPORTING TO HAVE BEEN BORN IN NEW ZEALAND, BY ETHNIC GROUPING

Ethnic grouping	Born in New Zealand			Not born in New Zealand		
	% of students	Mean mathematics score (se)	Mean science score (se)	% of students	Mean mathematics score (se)	Mean science score (se)
Pakeha/ European	92	507 (5.2)	533 (4.6)	8	521 (6.8)	543 (8.6)
Maori	97	454 (4.9)	473 (5.9)	3	455 (20.0)	458 (30.0)
Pacific	73	433 (10.7)	441 (12.2)	27	419 (15.3)	401 (18.2)
Asian	29	528 (15.5)	541 (12.9)	71	537 (10.7)	505 (10.5)

TABLE D.3: YEAR 9 STUDENTS' MEAN ACHIEVEMENT SCORES IN 1994, BY THE DEGREE TO WHICH ENGLISH WAS SPOKEN IN THE HOME

Speaking English in the home	Mean achievement scores in 1994		
	% of students	mathematics score (se)	science score (se)
Always/almost always	91	505 (4.5)	519 (4.6)
Sometimes/never	9	477 (9.4)	449 (9.8)

TABLE D.4: THE DEGREE TO WHICH YEAR 9 STUDENTS REPORTED SPEAKING ENGLISH IN THE HOME IN 1994 AND 1998, BY ETHNIC GROUPING

Ethnic grouping	Always/almost always (% of students)		Sometimes/never (% of students)	
	1994	1998	1994	1998
Pakeha/ European	98	98	2	2
Maori	90	92	10	8
Pacific	66	60	34	40
Asian	42	43	58	57

TABLE D.5: YEAR 9 STUDENTS' MEAN ACHIEVEMENT SCORES IN1998 BY THE DEGREE TO WHICH STUDENTS REPORTED SPEAKING ENGLISH IN THE HOME, BY ETHNIC GROUPING

Ethnic grouping	Always/almost always		Sometimes/never	
	Mean mathematics score (se)	Mean science score (se)	Mean mathematics score (se)	Mean science score (se)
Pakeha/ European	510 (5.3)	535 (4.6)	479 (14.9)	500 (18.9)
Maori	457 (5.0)	476 (5.9)	419 (11.6)	430 (11.4)
Pacific	447 (10.9)	460 (10.2)	401 (11.2)	386 (16.1)
Asian	534 (13.4)	539 (12.8)	536 (10.9)	498 (10.2)

TABLE D.6: MEAN ACHIEVEMENT SCORES FOR STUDENTS AT EACH LEVEL OF THE HOME EDUCATIONAL RESOURCES (HER) INDEX IN 1998/99

Level of the HER Index	New Zealand			International		
	Proportion of students (%)	Mean mathematics score (se)	Mean science score (se)	Mean proportion of students (%)	Mean mathematics score (se)	Mean science score (se)
Low	6	418 (9.3)	422 (11.2)	19	431 (1.2)	431 (1.5)
Medium	76	484 (4.8)	503 (4.5)	72	487 (0.8)	487 (0.8)
High	18	546 (6.5)	567 (5.9)	9	559 (2.3)	558 (2.0)

SOURCE: MARTIN ET AL (2000) AND MULLIS ET AL (2000).

TABLE D.7: YEAR 9 STUDENTS' EDUCATIONAL ASPIRATIONS AND MEAN ACHIEVEMENT SCORES IN 1998

Educational aspiration	Mean achievement scores in 1998		
	% of students	mathematics score (se)	science score (se)
Some secondary	3	433 (7.7)	450 (14.5)
Finish secondary^	16	451 (4.9)	473 (6.9)
Vocational/ polytechnic or some university+	16	477 (5.2)	507 (4.6)
Finish university*	52	520 (5.8)	536 (5.7)
Do not know	13	465 (7.0)	473 (8.5)

^ *Complete form 6 or 7.*

+ *Included trade certificate course at polytechnics.*

* *Includes degree courses offered at a polytechnic and colleges of education programmes.*

TABLE D.8: YEAR 9 STUDENTS' PERCEPTIONS OF THE IMPORTANCE OF 'DOING WELL' IN SCHOOL SUBJECTS AND SPORT IN 1998

It is important to …	Student agrees it is important (%)	Student perceives mother thinks it is important (%)	Student perceives friends think it is important (%)
…do well in science at school	97	98	76
…do well in mathematics at school	93	96	67
…do well in English at school	97	98	75
…have time to have fun	98	95	97
…be good at sports	86	84	86

TABLE D.9: THE EXTENT TO WHICH YEAR 9 STUDENTS IN 1998 AGREED WITH ATTITUDE TO MATHEMATICS STATEMENTS

Statement	Dislike a lot or strongly disagree (%)	Dislike or disagree (%)	Like or agree (%)	Like a lot or strongly agree (%)
How much do you like mathematics?	8	20	53	20
I enjoy learning mathematics	5	22	53	20
Mathematics is boring	14	44	31	11
Mathematics is a very easy subject	13	54	27	6
Mathematics is important to everyone's life	2	5	39	55
I would like a job that involved using mathematics	15	36	37	12

TABLE D.10: THE PROPORTIONS OF YEAR 9 STUDENTS AT EACH LEVEL OF THE POSITIVE ATTITUDES TO MATHEMATICS (PATM) INDEX IN 1994, BY GENDER

Level on the PATM Index	Girls	Boys	Overall	
	% of students	% of students	% of students	Mean mathematics score (se)
Low	12	10	11	481 (5.6)
Medium	56	51	54	496 (4.7)
High	33	39	36	519 (5.9)

TABLE D.11: THE PROPORTIONS OF YEAR 9 STUDENTS AT EACH LEVEL OF THE POSITIVE ATTITUDES TO MATHEMATICS (PATM) INDEX IN 1994 AND 1998, BY ETHNIC GROUPING

Ethnic grouping	Proportion (%) of students at each level of the PATM Index in 1994			Proportion (%) of students at each level of the PATM Index in 1998		
	Low	Medium	High	Low	Medium	High
Pakeha/ European	12	55	33	12	56	31
Maori	11	57	33	9	59	32
Pacific	7	40	53	8	44	48
Asian	3	37	61	4	47	49

TABLE D.12: THE EXTENT TO WHICH YEAR 9 STUDENTS IN 1998 AGREED WITH ATTITUDE TO SCIENCE STATEMENTS

Statement	Dislike a lot or strongly disagree (%)	Dislike or disagree (%)	Like or agree (%)	Like a lot or strongly agree (%)
How much do you like science?	9	21	51	19
I enjoy learning science	7	21	50	22
Science is boring	16	47	27	11
Science is a very easy subject	12	55	29	5
Science is important to everyone's life	4	20	51	25
I would like a job that involved using science	21	37	28	15

TABLE D.13: THE PROPORTIONS OF YEAR 9 STUDENTS AT EACH LEVEL OF THE POSITIVE ATTITUDES TO SCIENCE (PATS) INDEX IN 1994, BY GENDER

Level of the PATS Index	Girls	Boys	Overall	
	% of students	% of students	% of students	Mean science score (se)
Low	21	14	17	489 (5.7)
Medium	56	54	55	509 (4.9)
High	23	32	27	536 (6.7)

TABLE D.14: THE PROPORTIONS OF YEAR 9 STUDENTS AT EACH LEVEL OF THE POSITIVE ATTITUDES TO SCIENCE (PATS) INDEX IN 1994 AND 1998, BY ETHNIC GROUPING

Ethnic Grouping	Proportion (%) of students at each level of the PATS Index in 1994			Proportion (%) of students at each level of the PATS Index in 1998		
	Low	Medium	High	Low	Medium	High
Pakeha/ European	18	55	27	17	55	28
Maori	22	54	24	18	60	22
Pacific	12	55	34	12	52	36
Asian	10	53	37	11	53	37

Technical Notes

TN.1: DESIGN EFFECT

As it can be seen from the outline provided in Chapter 1, TIMSS-98 is a complex design. One area of complexity is the sample design. The object of TIMSS is to be able to generalise results to the national population: TIMSS is a population survey. To summarise the sampling process:

- Schools are selected on a probability proportional to size (of target grade) basis.

- A single whole mathematics class within each selected school is randomly selected to participate.

From this process two issues arise:

- Students in selected schools with many mathematics classes are less likely to be selected than students in selected schools with few mathematics classes. This issue is dealt with simply through weighting the sample appropriately.

- Students clustered within an intact classroom tend to have more similar characteristics than children randomly selected from the population. This is often called the cluster effect.

Note that by selecting intact mathematics classes the clustering effect will be different when examining mathematics achievement and science achievement: it is often the case at this level that students with similar abilities in mathematics are grouped together for mathematics lessons, but these students are regrouped into different science classes on the basis of ability in science. Hence, taking intact mathematics classes as a basis of the sample, students within the class will tend to have more similar mathematics abilities (ie, there will be a stronger clustering effect) than they will science abilities.

This clustering effect is in essence reducing the efficiency of the sample. However, this reduction in sample efficiency can be quantified through calculating the design effect. The **global design effect** (one that accounts for inefficiencies in the sampling design as well as measurement error in the imputation process) is calculated with the following formula:

$$DEFF = \frac{se^2_{jk(1\text{-}5)}}{se^2_{srs(1)}}$$

Where:

$se_{jk(1\text{-}5)}$ is the standard error of the scale score obtained using the jackknife replication method with all five imputed values (see TN.2 and TN.3)

$se_{srs(1)}$ is the standard error of the scale score obtained assuming simple random sample with the first imputed value

TN.1a: effective sample size

Using the design effect, an effective sample size can be obtained with the following formula:

$$n_{effective} = \frac{n}{DEFF}$$

For example, there was a total of 3613 New Zealand Year 9 students who participated in TIMSS-98/99. Adjusting this figure for the clustering effects and measurement error yielded effective sample sizes of 287 for mathematics and 340 for science. Essentially, this means that we can be as confident about the results in TIMSS-98/99 as if we had selected 287 or 340 Year 9 students totally at random from the population to examine their abilities in mathematics and science respectively.

Table TN.1 gives the achieved sample size and the effective sample size for selected Year 9 student sub-populations in TIMSS-98/99. Note how the sample design has been more efficient in measuring achievement in some sub-populations than in others. For example, the 389 Maori boys in the study were effectively equivalent to a simple random sample of 146 when examined in mathematics. Yet the efficiency of the measurement of all boys was much lower, as the sample of 1799 was effectively equivalent to a simple random sample of only 139 when examined in mathematics. This means that there was relatively greater between-class variation than within-class variation in mathematics ability for Maori boys, than there was for boys as a whole.

TABLE TN.1: SELECTED SAMPLE SIZES AND EFFECTIVE SAMPLE SIZES — YEAR 9 STUDENTS IN TIMSS-98

		Mathematics	Science
	n	$n_{effective}$	$n_{effective}$
Overall	3613	287	340
Girls	1814	243	254
Boys	1799	139	183
Pakeha	2208	250	360
Maori	793	266	192
Pacific	245	74	55
Asian	256	81	85
Pakeha Girls	1104	207	274
Pakeha Boys	1104	125	182
Maori Girls	404	200	144
Maori Boys	389	146	141
Pacific Girls	125	39	28
Pacific Boys	120	35	38
Asian Girls	135	44	51
Asian Boys	121	53	39

TN.2: SUMMARY OF THE SCALING PROCESS

As outlined in the introduction, TIMSS-98/99 makes use of a multiple-matrix sampling whereby students answer subsets of items from a larger pool of test items. Psychometric scaling techniques based on Item Response Theory enable population estimates to be generated even though students do not respond to all of the same achievement items. The Item Response Theory scaling used in TIMSS-98/99 uses the plausible value methodology to produce estimates of student proficiency in mathematics and science.

Essentially, these steps were taken to generate the overall mathematics and science scores as well as the scale scores for the subtopics for each subject:

1. *Scaling* — each assessment item is examined against a model which expresses the probability that respondents with different abilities will achieve a certain score on the item. Three Item Response Theory models are used corresponding to the three types of test items.

   For multiple-choice items a three parameter logistic model was used, which characterises the item in terms of difficulty, discrimination and the possibility of guessing. For dichotomous open-response items a two parameter logistic was used (the possibility of guessing is discounted). For polychotomous items, a generalised partial credit model was used, which factors in the different scores available to respondents (eg, 0, 1, 2 and 3 rather than just right or wrong).

2. *Generating plausible values* — to acknowledge the measurement error surrounding the assessment of an individual's abilities, five separate ability estimates are made for each individual. These estimates are referred to as 'plausible values' and are based on information from the student achievement and background data and the item parameters. Together, these plausible values are used to determine unbiased population estimates.

3. *Transforming the scaled scores on to a reporting scale* — for reporting purposes the scaled scores are transformed onto a more accessible reporting scale with a mean of 500 and standard deviation of 100.

The information in this note has largely been adapted from Johnson (1998). The interested reader is referred to Martin, Gregory and Stemler, 2000, for a comprehensive explanation of the scaling process.

TN.3: STANDARD ERROR AND THE JACKKNIFE REPLICATION METHOD

Because of the complexity of the design of TIMSS-98/99, the calculation of standard errors is not as straightforward as it is for a study which uses simple random sampling. Not only is there the design effect with regard to the sample to be considered (as noted in TN.1), there is the further issue that each student receives five imputed values (see TN.2) for each of the scales used to estimate their underlying ability.

Standard errors must, therefore, be estimated using numerical algorithms, which amount roughly to 'simulating' variations in the sampling of class groups, for example, by successively leaving whole classes out of the sample and calculating how the mean is affected. There are a number of different numerical methods or algorithms available for estimating standard errors. The TIMSS-94/95 and TIMSS-98/99 studies use the Jackknife Repeated Replication method (Gonzalez & Foy, 2000).

TN.4: SIGNIFICANCE TESTING

Throughout this report, differences have been described as statistically significant where $\alpha = 0.05$ (95% confidence). Although some results were highly statistically significant (eg, $p < 0.01$) these have not been explicitly identified as such in this report. Conservative methods have been used to calculate of the significance of differences between mean scores achieved by different population groups. The use of conservative methods increases confidence that where differences are reported as significant they have not occurred by chance. That is to say, there is less likelihood of making a Type I error of rejecting a null-hypothesis of 'no differences between the groups' when it should be accepted. Conversely, however, there is an increased likelihood of making a Type II error — that is accepting a null-hypothesis of 'no differences between the groups' when it should be rejected. This needs to be considered when interpreting the statistics presented in this report.

The *t* statistic was calculated using:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{se_1^2 + se_2^2}}$$

Note that *t* may be slightly inflated for some groups because of the assumption that population groups have been independently sampled. For example, the sample of boys and the sample of girls are not completely independent and it would be more accurate to jackknife the differences between boys and girls. However, due to time constraints on reporting these data, the above formula, assuming sample independence, was applied uniformly (see Gonzalez, 1997). The results reported here are thus consistent with the use of conservative methods when examining differences between groups (there is less likelihood of making a Type I error).

TN.5: MULTIPLE COMPARISONS OF MEANS

When making a comparison between two means the value of *t* must be at least equal to the critical value1.96 for $\alpha = 0.05$ (2-tailed). However, in cases where there are more than two means being compared, the probability of making a Type I error increases unless the number of comparisons being made is taken into account. There are a number of ways to correct for this; in this report the Dunn-Bonferroni procedure has been used. Essentially, this procedure raises the critical value that *t* must reach before the (multiple) comparisons can be considered statistically significantly different at the five percent level.

It was sometimes the case, however, that when the achievement means of two groups were compared, after adjusting the significance test for multiple comparisons the differences were found to be not statistically significant; yet if the same groups' means had been compared in isolation the difference *would* have been considered statistically significant. In other words a Type II error may have occurred whereby the null-hypothesis of 'no difference' is accepted when in fact it is possible that there is sufficient evidence to confidently reject the null-hypothesis.

Take, for example, the science achievement means of Pakeha/European (537) and Asian (512). Analysis of this difference yields a *t* of 2.12. If these two group means were compared in isolation it could be said they are significantly different because *t* exceeds the critical value of 1.96 ($\alpha = 0.05$, 2-tailed). However, in this report, these groups' means are compared alongside two other groups' means and are not found to be significantly different from each other, as the critical value has been raised to 2.64 to compensate for the measurement errors inherent in the six comparisons.

In order to ameliorate for this increased likelihood of the Type II error arising from the use of the Dunn-Bonferroni procedure *t* has, in some cases, been reported in appended tables where appropriate. These tables, in conjunction with Table TN.2 below, allow the reader to interpret the significance of the data from their own perspective.

For example, when looking at differences between the four main ethnic groupings, six comparisons are made and the resulting critical value for *t* where $\alpha = 0.05$ is 2.64. But the reader may only be interested in comparing one ethnic grouping (eg, Pacific) with the other ethnic groupings and thus are only making three comparisons. In this case a critical value of 2.39 is more appropriate to use to determine $\alpha = 0.05$. Note, however, that using this lower critical value means that nothing can then be said about the differences between the other groups and there is also the increased likelihood of a Type I error occurring (for which the Dunn-Bonferroni procedure was employed to guard against in the first place).

TABLE TN.2: DUNN-BONFERRONI ADJUSTED CRITICAL VALUES FOR *t* FOR SELECTED NUMBERS OF COMPARISONS

Number of groups	Number of comparisons	Critical value of t ($\alpha = 0.05$).
2	1	1.96
~	2	2.24
3	3	2.39
~	4	2.50
~	5	2.58
4	6	2.64
6	15	2.94
8	28	3.12
9	36	3.20
16	120	3.53

TN.6 EFFECT SIZES

As well as determining whether differences in mean achievement are statistically significant it is useful to give an impression of the *magnitude* of the difference. One way of doing this is through the use of effect sizes. There are various ways of calculating and using effect sizes (see Rosenthal, 1994). For the purpose of this report we have used the following method.

Calculating the effect size

Firstly the within, pooled standard deviation (s_w) of the two groups being compared is calculated for each of the five imputed scale scores using:

$$s_w = \sqrt{\frac{\sum W_1 s_1^2 + \sum W_2 s_2^2}{\sum W_1 + \sum W_2}}$$

Where:

W_i is the sample weight of group i

s_i is the standard deviation of the scale score of group i

Then the effect size between the two groups, Cohen's d, is calculated for each of the five imputed scale scores using:

$$d = \frac{\overline{X}_1 - \overline{X}_2}{s_w}$$

Where:

X_i is the mean imputed scale score of group i

The final effect size figure reported in this report is the mean effect size of the five imputed scale scores.

Interpreting the effect size

When interpreting an effect size between two groups, technically, an effect size of 1.0 indicates a relative advantage of one standard deviation on the utilised measure. In other words, the mean of one group will be a whole standard deviation higher than the mean of the other (see Wilkinson et al, 2000, section 1.3.2 for a useful outline of effect sizes and their uses).

TN.7: MISSING STUDENTS

Analyses in this report are based on all students with achievement data. However, there is invariably data missing from students with regard to *context* variables for two reasons:

1. the student was absent from the session in which the context questionnaire was administered; or

2. the student participated in the session but did not answer the specific question(s) under examination.

Typically, between two and four percent of students are missing from contextual analyses reported in this document.

REFERENCES

References (including the TIMSS-98/99 reports) cited in this publication

Alton-Lee, A, and Praat, A (2000). *Explaining and Addressing Gender Differences in the New Zealand Compulsory School Sector: A Literature Review*. Wellington: Research Division, Ministry of Education.

Beaton, A E, Martin, M O, Mullis, I V S, Gonzalez, E J, Smith, T A, and Kelly, D L (1996a). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Beaton, A E, Mullis, I V S, Martin, M O, Gonzalez, E J, Kelly, D L, and Smith, T A (1996b). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Bradby, D, Owings, J, and Quinn, P (1992). *Language Characteristics and Academic Achievement: A Look at Asian and Hispanic Eighth Graders in NELS:88*. Washington: NCES 92-479, US Department of Education.

Chamberlain, G S (1997a). Achievement in Science. In R A Garden (ed), *Mathematics and Science Performance in Middle Primary School: Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Chamberlain, M C (1996a). Achievement in Science — 1. In R A Garden (ed), *Science Performance of New Zealand Form 2 and Form 3 Students: National Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Chamberlain, M C (1996b). Achievement in Mathematics — 1. In R A Garden (ed), *Mathematics Performance of New Zealand Form 2 and Form 3 Students: National Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Chamberlain, M C (1997b). Achievement in Mathematics. In R A Garden (ed), *Mathematics and Science Performance in Middle Primary School: Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Crooks, T and Flockton, L (2000). *Science Assessment Results 1999*. Dunedin: Educational Assessment Research Unit, University of Otago.

Department of Statistics (1992). Standard Classification of Ethnicity. *The Statistics Co–ordination Update*. Wellington: Department of Statistics.

Department of Statistics (1996). Standard Classification of Ethnicity. *The Statistics Co–ordination Update*. Wellington: Department of Statistics.

Foy, P and Joncas, M (2000). TIMSS Sample Design. In M O Martin, K A Gregory, and S E Stemler, (eds) (2000). *The TIMSS 1999 Technical Report*, Chestnut Hill, MA: International Study Centre, Boston College.

Garden, R A (1996a). Conclusion. In R A Garden (ed), *Mathematics Performance of New Zealand Form 2 and Form 3 Students: National Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Garden, R A (1996b). Conclusion. In R A Garden (ed), *Science Performance of New Zealand Form 2 and Form 3 Students: National Results from New Zealand's Participation in the Third International*

Mathematics and Science Study. Wellington: Research and International Section, Ministry of Education.

Gonzalez, E J (1997). Reporting Student Achievement in Mathematics and Science. In M O Martin, and D L Kelly (eds) (1997). *Third International Mathematics and Science Achievement: Technical Report Volume II: Implementation and Analysis, Primary and Middle School Years*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Gonzalez, E J and Foy, P (2000). Estimation of Sampling Variance. In M O Martin, K A Gregory, and S E Stemler (eds) (2000). The TIMSS 1999 Technical Report, Chestnut Hill, MA: International Study Center, Boston College.

Johnson, E (1998). *Scaling the TIMSS-R Data*. Unpublished paper presented to the TIMSS-R National Research Coordinators meeting in Boston 18-22 August, 1998.

Martin, V M (1996a). The Students. In R A Garden (ed), *Mathematics Performance of New Zealand Form 2 and Form 3 Students: National Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Martin, V M (1996b). The Students. In R A Garden (ed), *Science Performance of New Zealand Form 2 and Form 3 Students: National Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Martin, M O, Gregory, K A, and Stemler, S E (eds) (2000). *The TIMSS 1999 Technical Report*, Chestnut Hill, MA: International Study Center, Boston College.

Martin, M O, Mullis, I V S, Beaton, A E, Gonzalez, E J, Smith, T A, and Kelly, D L (1997). *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Mullis, I V S, Martin, M O, Beaton, A E, Gonzalez, E J, Kelly, D L, and Smith, T A (1997). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M O, Mullis, I V S, Gonzalez, E J, Gregory K D, Smith, T A, Chrostowski, S J, Garden, R A, and O'Connor, K M (2000). *TIMSS 1999 International Science Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: International Study Center, Boston College.

Mullis, I V S, Martin, M O, Gonzalez, E J, Gregory K D, Garden, R A, O'Connor, K M, Chrostowski, S J, and Smith, T A (2000). *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: International Study Center, Boston College.

McKnight, C C, Schmidt, W H, and Raizen, S (1993). *Test Blueprints: A Description of the TIMSS' Achievement Test Content Design*. Unpublished document, IEA reference: ICC 797/NRC 357.

Ministry of Education (1999). *Education Statistics of New Zealand for 1998*. Wellington: Data Management and Analysis Section, Ministry of Education.

Ministry of Education (1992). *Mathematics in the New Zealand Curriculum*. Wellington: Learning Media.

Ministry of Education (1993). *Science in the New Zealand Curriculum*. Wellington: Learning Media.

Nash, R and Harker, R K (1997). *Progress at School: Final Report to the Ministry of Education*. Palmerston North: Educational Research and Development Centre, Massey University.

Nash, R and Harker, R K (1998). *Making Progress: adding value in secondary education*. Palmerston North: Massey University.

Robitaille, D F and Maxwell, B (1996). The Conceptual Framework and Research Questions. In D F Robitaille and R A Garden (eds), *TIMSS Monograph No. 2: Research Questions and Study Design*. Vancouver, BC: Pacific Educational Press.

Robitaille, D F, McKnight, C C, Schmidt, W H, Britton, E, Raizen, S, and Nicol, C (1993). *TIMSS Monograph No 1: Curriculum Framework for Mathematics and Science*. Vancouver, BC: Pacific Educational Press.

Rosenthal, R (1994). Parametric measures of effect size. In H Cooper and L V Hedges (eds), *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

Stocking, M L (1999). Item Response Theory. In G N Masters and J P Keeves (eds), *Advances in Measurement in Educational Research and Assessment*. Oxford, England: Elsevier Science Ltd.

Wagemaker, H (ed) (1993). *Achievement in Reading Literacy: New Zealand's Performance in a National and International Context*. Wellington: Research Section, Ministry of Education.

Walker, M (1998). The Students. In R A Garden (ed), *Mathematics and Science Literacy in the Final Year of Schooling*. Wellington: Research Division, Ministry of Education.

Wilkinson, I A G, Hattie, J A, Parr, J M, Townsend, M A R, Thrupp, M, Lauder, H & Robinson, T (2000). *Influence of peer effects on learning outcomes: a review of the literature*. Wellington: Ministry of Education, Research Division.

International TIMSS-94/95 reports, monographs, and other publications

Beaton, A E, Martin, M O, Mullis, I V S, Gonzalez, E J, Smith, T A, and Kelly, D L (1996a). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Beaton, A E, Mullis, I V S, Martin, M O, Gonzalez, E J, Kelly, D L, and Smith, T A (1996b). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Harmon, M, Smith, T A, Martin, M O, Kelly, D L, Beaton, A E, Mullis, I V S, Gonzalez, E J, and Orpwood, G (1997). *Performance Assessment in IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Howson, A G (1995). *TIMSS Monograph No 3: Mathematics Textbooks: A Comparative Study of Grade 8 Texts*. Vancouver, BC: Pacific Educational Press.

Martin, M O and Kelly, D L (eds) (1996). *Third International Mathematics and Science Achievement: Technical Report Volume 1: Design and Development*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M O and Kelly, D L (eds) (1997). *Third International Mathematics and Science Achievement: Technical Report Volume II: Implementation and Analysis, Primary and Middle School Years*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M O and Kelly, D L (eds) (1997). *Third International Mathematics and Science Achievement: Technical Report Volume III: Implementation and Analysis, Final Year of Secondary School (Population 3)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M O and Mullis, I V S (eds) (1996). *Third International Mathematics and Science Study: Quality Assurance in Data Collection*. Chestnut Hill, MA: Centre for the Testing, Evaluation, and Educational Policy, Boston College.

Martin, M O, Mullis, I V S, Beaton, A E, Gonzalez, E J, Smith, T A, and Kelly, D L (1997). *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Martin, M O, Mullis, I V S, Gonzalez, E J, Smith, T A, and Kelly, D L (1999). *School Contexts for Learning and Instruction: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Mullis, I V S, Martin, M O, Beaton, A E, Gonzalez, E J, Kelly, D L, and Smith, T A (1997). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Mullis, I V S, Martin, M O, Beaton, A E, Gonzalez, E J, Kelly, D L, and Smith, T A (1998). *Mathematics and Science Achievement in the Final Year of Secondary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Centre for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Orpwood, G and Garden, R A (1998). *TIMSS Monograph No 4: Assessing Mathematics and Science Literacy*. Vancouver, BC: Pacific Educational Press.

Robitaille, D F and Garden, R A (eds) (1996). *TIMSS Monograph No 2: Research Questions and Study Design*. Vancouver, BC: Pacific Educational Press.

Robitaille, D F, McKnight, C C, Schmidt, W H, Britton, E, Raizen, S, and Nicol, C (1993). *TIMSS Monograph No 1: Curriculum Framework for Mathematics and Science*. Vancouver, BC: Pacific Educational Press.

Robitaille, D F (ed) (1997). *National Contexts for Mathematics and Science Education: An Encyclopedia of the Education Systems Participating in TIMSS*. Vancouver, BC: Pacific Educational Press.

Robitaille, D F, Beaton, A E, and Plomp Tj (eds) (2000). *Impact of TIMSS on the Teaching & Learning of Mathematics & Science*. Vancouver, BC: Pacific Educational Press.

Schmidt, W H, McKnight, C C, Valverde, G A, Houang, R T, and Wiley, D E (1997). *Many Visions, Many Aims Volume 1: A Cross–National Investigation of Curricular Intentions in School Mathematics*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Schmidt, W H, McKnight, C C, Valverde, G A, Houang, R T, and Wiley, D E (1997). *Many Visions, Many Aims Volume 2: A Cross–National Investigation of Curricular Intentions in School Science*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

New Zealand TIMSS-94/95 national reports

Chamberlain, M C, Chamberlain, G S, and Garden, R A (1998). *Student Performance on Open-ended Questions in the Third International Mathematics and Science Study: New Zealand Results*. Wellington: Research Division, Ministry of Education.

Garden, R A (ed) (1996a). *Mathematics Performance of New Zealand Form 2 and Form 3 Students: National Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Garden, R A (ed) (1996b). *Science Performance of New Zealand Form 2 and Form 3 Students: National Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Garden, R A (ed) (1997). *Mathematics and Science Performance in Middle Primary School: Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research and International Section, Ministry of Education.

Garden, R A (1997b). *Performance Assessment in the Third International Mathematics and Science Study: New Zealand Results*. Wellington: Research and International Section, Ministry of Education.

Garden, R A (ed) (1998). *Mathematics and Science Literacy in the Final Year of Schooling: Results from New Zealand's Participation in the Third International Mathematics and Science Study*. Wellington: Research Division, Ministry of Education.