Wānangatia te Putanga Tauira
National Monitoring Study of Student Achievement

Technical Information 2014

Social Studies • English: Reading

Wānangatia te Putanga Tauira
National Monitoring Study of Student Achievement

National Monitoring Study of Student Achievement
TECHNICAL INFORMATION 2014 – Social Studies • English: Reading
Report 7

Wānangatia te Putanga Tauira

National Monitoring Study of Student Achievement

Technical Information 2014

Social Studies • English: Reading

Educational Assessment Research Unit
and
New Zealand Council for Educational Research

NMSSA Report 7

National Monitoring Project of Student Achievement Report 7: Technical Report – Social Studies, English: Reading 2014
published by Educational Assessment Research Unit, University of Otago, and New Zealand Council for Educational Research, under contract to the Ministry of Education, New Zealand

ISSN: 2350-3238 (Online) ISBN: 978-1-927286-20-3 (Online only)

National Monitoring Study of Student Achievement
Educational Assessment Research Unit, University of Otago, PO Box 56, Dunedin 9054, New Zealand
Tel: 64 3 479 8561 • Email: nmssa@otago.ac.nz

Contents

Acknowledgements

The NMSSA project team wishes to acknowledge the very important and valuable support and contributions of many people to this project, including:

- Members of the reference groups: Technical, Māori, Pasifika, and Special Education
- Members of the curriculum advisory panels in social studies and English: reading
- Principals and students of the schools where the tasks were piloted and trials were conducted
- Principals, teachers and Board of Trustees members of the schools who participated in the 2014 main study including the linking study
- The students who participated in the assessments and their parents, whānau and caregivers
- The teachers who administered the assessments to the students
- The teachers, senior initial teacher education students and others who undertook the marking
- The Ministry of Education Research Team and Steering Committee

2014 Project Team	EARU	NZCER
Management Team	Alison Gilmore Ros Allan Lynette Jones	Charles Darr Chris Joyce
Design/Statistics/ Psychometrics/Reporting	Alison Gilmore Mustafa Asil Denise Quinlan Megan Anakin	Charles Darr Hilary Ferral Jess Mazengarb
Curriculum/Assessment	Ros Allan Jane White Doris Lancaster	Juliet Twist Linda Bonne Verena Watson Sue McDowell Charles Darr Rose Hipkins
Programme Support	Lynette Jones Linda Jenkins James Rae Judith Gray Gail Morris Lee Baker	Jess Mazengarb Sarah Boyd
External Advisors	Jeffrey Smith – University of Otago Marama Pohatu – Te Rangatahi Ltd	

1 Appendix 1: Sample characteristics for 2014

1. Samples for 2014

A two-stage sampling design was used to select nationally representative samples of students at Year 4 and at Year 8. The first stage involved sampling schools, and the second stage involved sampling students within schools.

A stratified random sampling approach was taken with the intention of selecting 100 schools at Year 4 and 100 schools at Year 8. A maximum of 27 students were randomly selected from each school with two being available as reserves. From each school sample up to 25 students contributed to a national sample of approximately 2200 students at Year 4 and 2200 students at Year 8.

To ensure that the 2014 student sample was nationally representative the MoE 2013 school returns for Year 3 and Year 7 were used to inform the selection of Year 4 and Year 8 schools in 2014.

2. Sampling of schools

Sampling Algorithm

From the complete list of NZ schools select two datasets – one for Year 3 students and the other for Year 7 students.

For the Year 3 sample:

- Exclude:
    - Schools which have fewer than 8 Year 3 students
    - Private schools
    - Special schools
    - Correspondence School
    - Kura Kaupapa Māori.
- Stratify the sampling frame by region and quintile[1].
- Within each region-by-quintile stratum, order the schools by Year 3 roll size[2].
- Arrange the strata alternately in increasing and decreasing order of roll size[3].
- Select a random starting point.
- From the random starting point, cumulate the Year 3 roll.
- Because 100 schools are required in the sample, the sampling interval is calculated as:

$$\frac{\text{Total number of Year 3 students}}{100}$$

- Assign each school to a "selection group" using this calculation:

$$Selection\ group\ =\ ceiling\left(\frac{cumulative\ roll}{sampling\ interval}\right)$$

- Select the first school in each selection group to form the final sample.

Follow the same process for the Year 7 sample.

---

1 Decile 1 and 2 comprises Quintile 1; Decile 3 and 4 comprises Quintile 2; Decile 5 and 6 comprises Quintile 3; Decile 7 and 8 comprises Quintile 4; and Decile 9 and 10 comprises Quintile 5.

2 Roll size refers to the year level in question e.g. roll size for Year 3 students

3 This is done so that when replacements are made across stratum boundaries the replacement school is of a similar size to the one it is replacing.

If a school is selected in both the Year 3 and Year 7 samples, randomly assign it to one of the two samples. Locate the school in the unassigned sample and select a replacement school (next on list). Repeat the process for each school selected in both samples.

2014 NMSSA sample

The sample frames constituted 2307 schools for Year 3 and 2434 schools for Year 7 after exclusions had been applied. No schools were listed in both samples.

Selected schools were invited to participate in 2014. Therefore 'Year 3 schools' became 'Year 4 schools' and similarly 'Year 7 schools' became 'Year 8 schools'. Those that declined to participate were substituted using the following procedure:

- From the school sample frame, select the school one row below the school withdrawn.
- Verify that the substitute school is of similar type, decile, size.
- If this school is not available, re-select by going to one row above the school withdrawn. Verify profile.
- If this school is not available, select the school two rows below the school withdrawn. Continue in this sequence until a substitute is found.

In total, 46 schools (24 at Year 4 and 22 at Year 8) declined to participate. Replacement schools were found for all but for one Year 4 school. One Year 8 school withdrew two days prior to their visit date due to school merger issues.

The achieved samples of schools

The achieved sample of 99 schools at Year 4 represented a participation rate[4] of 99 percent; and the achieved sample of 100 schools at Year 8 represented a response rate of 100 percent.

3. Sampling of students

After schools agreed to participate in the programme, they were asked to provide a list of all Year 4 (or Year 8) students, identifying any students for whom the experience would be inappropriate (e.g. high special needs (ORS), very limited English language (ESOL), Māori Immersion Level 1, would be absent during the visit, had left the school, other health or behavioural issues). A group-administered assessment sample (GA) included up to 22 students per school. A sub-set of eight of these students formed the individual assessment sample (IA). The procedure for selecting students for the GA and the IA samples was as follows:

- Each school provided a list of all students in their school at Year 4 or Year 8 in 2014. The lists were arranged in the order as provided by the school (that is alphabetically by last name). A computer-generated random number between 1 and 1,000,000 was assigned to each student. Students were ranked by their random number from highest to lowest.
- The first 27 non-excluded students in the ordered list were identified as belonging to the full sample. The first eight students were identified as also belonging to the individual sample (Additional individual assessments would be carried out with these students). Where there were more than 25 students in a year level, up to two students next on the list were selected as 'reserves' for potential replacements if required.
- The draft school lists of selected students were returned to schools for approval. Principals and teachers were given a second chance to identify students for whom the NMSSA assessment would be inappropriate. Any identified students were replaced with students up to number 27 from the initial list, resulting in a confirmed list. Letters of consent were sent to the parents of selected students.
- The children of parents who declined to have their child participate were withdrawn from the list.

[4] School participation rate is defined as the number of schools that participated (the achieved sample) as a percentage of the number of schools required for the study.

- Prior to the start of school visits, withdrawn students were not replaced unless they had been omitted at the first stage in which case the student next on the school's student sample list would be included. The replacement strategy continued, if an originally selected student was withdrawn, up until two weeks prior to teacher assessors (TAs) arriving in schools to conduct the assessments. This time schedule was put in place as any later withdrawals meant we would not have had sufficient time to gain consent from parents of substitute students.
- On the day before arrival in each school, TAs checked the final student list.
- On-site replacements of students by TAs were made if:
    - any of students 1 – 8 (the individual sample) were absent or withdrawn (e.g. by principal) on the first day, prior to the start of assessments. They were replaced according to ethnicity / gender criteria.
    - All other students (up to 27) participated in group-administered assessments. However, a maximum of 25 booklets from each school was included in the results.
    - If students were absent or withdrawn (e.g. by the principal) after the start of the assessment programme, no replacements were made.

Achieved samples of students at Year 4

Table A1.1 shows that at Year 4 the intended sample was 2633 randomly selected students. Principals identified 307 students for whom the experience would be unsuitable. The 'eligible' sample was reduced to 2326. A further 158 students were withdrawn from the study by the principal or parents after the sample was drawn. Substitutes were selected for 91 students, and were not available for 9. A further 82 students withdrew late, were absent or did not respond for other reasons during the assessment period. The achieved GA sample included 2177 students representing a participation rate of 82 percent. The achieved IA sample included 791 students representing a participation rate of 98 percent.

Table A1.1 The selection of Year 4 students for the GA sample and IA sample

	GA - N	IA- N
Intended sample of students	2633	800
Students withdrawn by principal before sample selected	307	
Eligible sample	2326	800
Students withdrawn by parents or principal after sampling	158	8
Substitute students used	91	
Students for whom there were no substitutes	9	
Late withdrawals	6	1
Absences/non-responses during assessment period	76	
Achieved sample	2177	791

Table A1.2 contrasts the characteristics of the samples with the population.

Table A1.2 Comparison of GA and IA samples with population characteristics at Year 4

	Population (%)	GA sample n = 2174* (%)	IA sample n = 791* (%)
Gender			
Boys	51	47	47
Girls	49	53	53
Ethnicity**			
European	63	60	62
Māori	23	22	23
Pasifika	12	12	11
Asian	11	12	11
Other	3	7	7
School Quintile			
1-2	17	17	18
3-4	16	14	14
5-6	18	20	20
7-8	20	17	18
9-10	29	31	29
School Type			
Contributing (Year 1-6)	61	64	63
Full Primary (Year 1-8)	36	32	33
Composite (Year 1-10 & 1-13)	3	3	4
MOE Region			
Central North	21	22	22
Central South	18	17	18
Northern	40	41	39
Southern	21	20	20

(Note that rounding to integers means that percentages do not always add up to 100 percent)

* Some students' responses were excluded because their assessment data was not able to be used. E.g. too few questions were attempted to be able to be reliable estimate of their achievement, the video taped response was inaudible.

** Percentages for ethnic groupings do not add to 100%. Non-prioritised ethnicity data is used throughout the NMSSA reports. Non-prioritised ethnicity data is sourced from the Ministry of Education's live enrolments database ENROL, rather than School Roll Returns

Achieved samples of students at Year 8

Table A1.3 shows that at Year 8 the intended sample was 2763 randomly selected. Principals identified 394 students for whom the NMSSA assessment experience would be unsuitable. The 'eligible' sample was reduced to 2369. A further 167 students were withdrawn from the study by the principal or parents after the sample was drawn. Substitutes were selected for 106 students, and were not available for 22. A further 76 students withdrew late, were absent or did not respond for other reasons during the assessment period. The achieved GA sample included 2232 students represented a participation rate of 77 percent. The achieved individual sample included 795 students representing a participation rate of 99 percent.

Table A1.3 The selection of Year 8 students for the GA and IA samples.

	GA - N	IA - N
Intended sample of students	2763	800
Students withdrawn by principal before sample selected	394	
Eligible sample	2369	800
Students withdrawn by principals or parents after sampling	167	2
Supplement students used	106	
Students for whom there were no substitutes	22	3
Late withdrawals	8	
Absences/non responses during assessment period	68	
Achieved sample	2232	795

(Note that rounding to integers means that percentages do not always add up to 100 percent)

* Some students' responses were excluded because their assessment data was not able to be used. E.g. too few questions were attempted to be able to be reliable estimate of their achievement, were absent for the reading assessment but not for the questionnaire session.

** Percentages for ethnic groupings do not add to 100%. Non-prioritised ethnicity data is used throughout the NMSSA reports. Non-prioritised ethnicity data is sourced from the Ministry of Education's live enrolments database ENROL, rather than School Roll Returns

Table A1.4 contrasts the characteristics of the samples with the population.

Table A1.4 Comparison of GA and IA samples with population characteristics at Year 8

	Population (%)	GA sample n = 2190* (%)	IA sample n = 793* (%)
Gender			
Boys	51	51	52
Girls	49	49	48
Ethnicity**			
European	61	59	60
Māori	22	22	23
Pasifika	12	13	12
Asian	10	9	8
Other	3	6	6
School Quintile			
1-2	14	11	12
3-4	16	18	19
5-6	24	24	23
7-8	21	22	22
9-10	24	25	24
School Type			
Full Primary (Year 1-8)	34	32	35
Intermediate	47	51	48
Secondary (Year 7-13)	14	11	11
Composite (Year 1-13 & 7-10)	5	6	6
MOE Region			
Central North	22	20	21
Central South	18	19	19
Northern	38	39	38
Southern	22	22	22

(Note that rounding to integers means that percentages do not always add up to 100 percent)

* Some student responses were excluded because their assessment data was not able to be used. E.g. too few questions were attempted to be able to be reliable estimate of their achievement, the video taped response was inaudible.

** Percentages for ethnic groupings do not add to 100%. Non-prioritised ethnicity data is used throughout the NMSSA reports. Non-prioritised ethnicity data is sourced from the Ministry of Education's live enrolments database ENROL, rather than School Roll Returns

2 Appendix 2: Applying weights to the 2014 NMSSA samples

1. Introduction

NMSSA reports on achievement levels for some key subgroups that are not specifically accounted for in the initial sample stratification (for instance, gender and ethnicity). This means that these key subgroups may not be properly represented in the achieved sample. Applying post-stratification weights can correct for misrepresentation of subgroups. This report lays out the methodology for weighting in NMSSA, describes the work done to investigate the need to apply weights to the NMSSA 2014 sample, and makes recommendations as to whether weights should be applied for the 2014 analyses.

The investigation was carried out in two separate parts. One investigation involves the full sample i.e. those students who completed the group administered reading assessment, and the other involves the smaller subsample of students who completed the social studies assessment. The two samples will be referred to as the GA (group administered assessment) sample and the IA (individual assessment) sample respectively.

An early version of the 2014 data file was used in this analysis. Subsequent processing of the file resulted in updates to a small amount of the demographic data. This means that the numbers of students in different demographic groups quoted in this report are slightly different than those used in the final reporting.

Note about multiple ethnicities

NMSSA data is reported allowing for multiple ethnicities. In applying weights this must be taken into consideration. National probabilities for (multiple) ethnic group membership by year level were provided by ENROL as these figures are not available on the list of July roll returns which is provided to NZCER annually by the Ministry of Education (MoE). The July roll returns provided all other information needed to calculate national probabilities of group membership.

2. Method for investigation of both samples

Post-strata

The achieved NMSSA 2014 sample was post-stratified as follows:

- Quintile (Quintiles 1 - 5)
- Gender (Female/Male)
- NZE/non-NZE, Māori/non-Māori, Pasifika/non-Pasifika, Asian/non-Asian, Other/not-Other

Each ethnic group was treated separately. That is, each sample member was initially assigned five separate sample weights.

For **each** ethnic group a sample member belongs to one of 20 possible strata. See Figure A2.1. Examples are:

- For NZE ethnic group:
    - Quintile 2, Male, NZE
    - Quintile 3, Female, non-NZE
- For Māori ethnic group
    - Quintile 2, Female, Māori
    - Quintile 4, Male, non-Māori

Qunitile	1				2				3				4				5			
Gender	Female		Male		Female		Male		Female		Male		Female		Male		Female		Male	
Ethnic group indicator	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Figure A2.1 Post-strata (20 cells) for one ethnic group

Calculating weights

Weights for each ethnic group were calculated as follows:

$$\text{Weight} = \frac{\text{Stratum probability}_{national}}{\text{Stratum probability}_{sample}}$$

A **final weight** taking an average over all five weights was then calculated. This final weight is suitable to be used for reporting purposes if recommended.

3. Group Administered Assessment (GA) sample

Missing data in the reading assessment

322 students in the Year 4 and Year 8 samples did not complete the reading assessment. The reason for non-participation was very largely due to absences on the day of assessment.

An analysis of these students' demographics provided evidence that missing students are close enough to being missing at random (MAR) to be treated as such. This being the case no further adjustments needed to be made to the sample weights to account for non-response.

In Figures A2.2 to A2.7 below, comparisons are made between the GA sample and the responding sample by various demographics. All percentage differences (see Table A2.1) were less than 1 percent. The magnitude of these differences is unlikely to have any effect on estimates at national or at subgroup level.

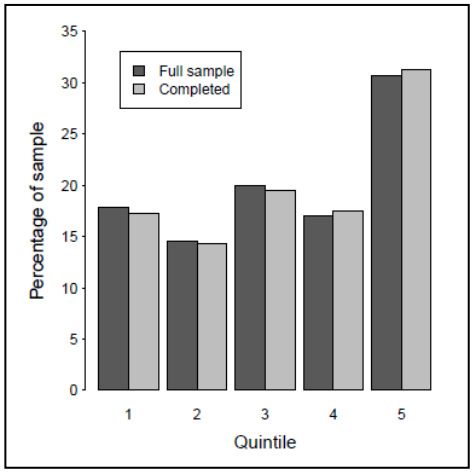

Figure A2.2 Year 4 sample proportions by quintile

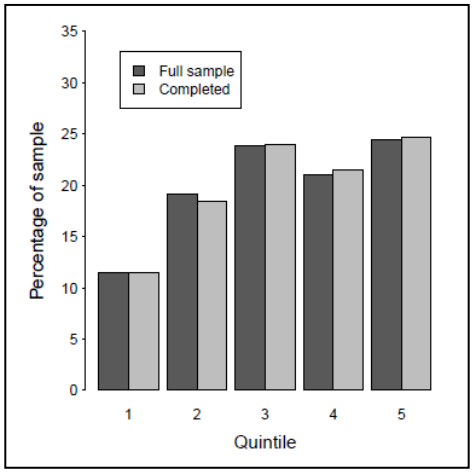

Figure A2.3 Year 8 sample proportions by quintile

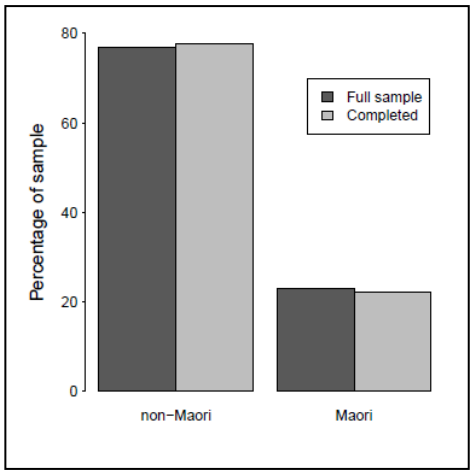

Figure A2.4 Year 4 sample proportions by Māori

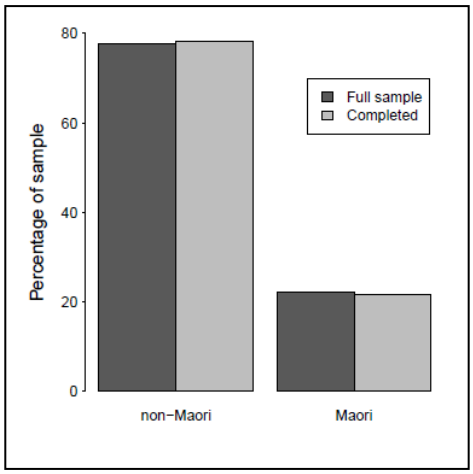

Figure A2.5 Year 8 sample proportions by Māori

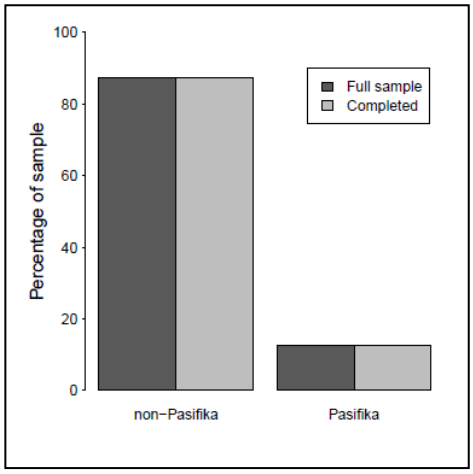

Figure A2.6 Year 4 sample proportions by Pasifika

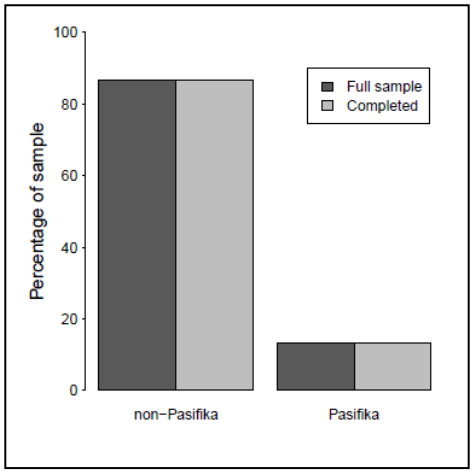

Figure A2.7 Year 8 sample proportions by Pasifika

Table A2.1 Proportional distributions for the GA sample vs. the responding sample

	Year 4		Year 8	
	All (%)	Completed (%)	All (%)	Completed (%)
Quintile 1	17.8	17.3	11.5	11.5
Quintile 2	14.5	14.3	19.1	18.4
Quintile 3	20.0	19.5	23.9	24.0
Quintile 4	17.0	17.5	21.0	21.5
Quintile 5	30.9	31.3	24.5	24.7
Female	53.1	53.3	48.6	49.0
Male	46.9	46.7	51.4	51.0
NZE	62.7	63.2	62.7	62.9
Non-NZE	37.3	36.8	37.3	37.1
Māori	23.1	22.2	22.2	21.8
Non-Māori	76.9	77.8	77.8	78.2
Pasifika	12.6	12.5	13.3	13.3
Non-Pasifika	87.4	87.5	86.7	86.7
Asian	11.8	12.3	9.4	9.5
Non-Asian	88.2	87.7	90.6	90.5

Size of the weights

In general, weights should not be allowed to become too big. A large weight means that only a few sample members are representing a particular post-stratum. This in turn is likely to introduce bias into reported results. Figure A2.8 is a histogram showing the range and distribution of sample weights. The histogram shows that weights are clustered closely around 1. This indicates that the sample has been well selected, and that little bias has been incurred as a result of sampling. Table A2.2 shows the distributional properties of the weights in tabular form.

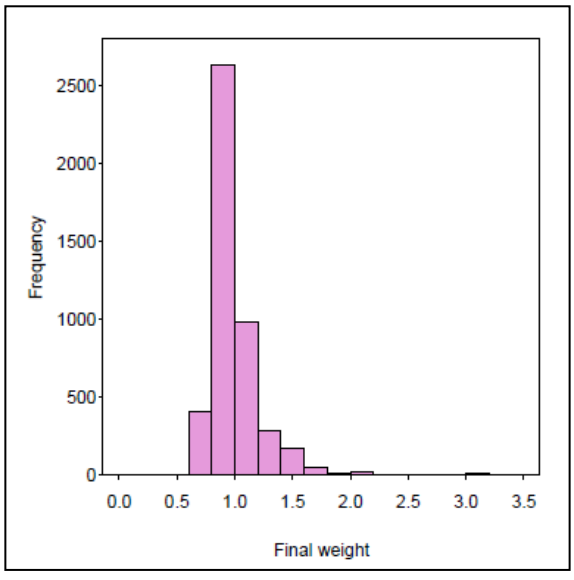

Figure A2.8 Distribution of sample weights

Table A2.2 Distribution of sample weights

	Weight	Meaning
1st percentile	0.75	1% of weights were < 0.75
25th percentile	0.87	25% of weights are < 0.87
50th percentile	0.96	50% of weights are < 0.96
75th percentile	1.06	75% of weights are < 1.06
99th percentile	1.81	99% of weights are < 1.81
Maximum weight	3.18	Year 8, Quintile 1, Female, NZ European (n=12)
Minimum weight	0.62	Year 4, Quintile 1, Female, Māori & Pasifika (n=10)

Do the sample weights change the results?

Figures A2.9 and A2.10 show that the overall distributions of reading achievement at both Year 4 and at Year 8 show very little difference with respect to unweighted or weighted data.

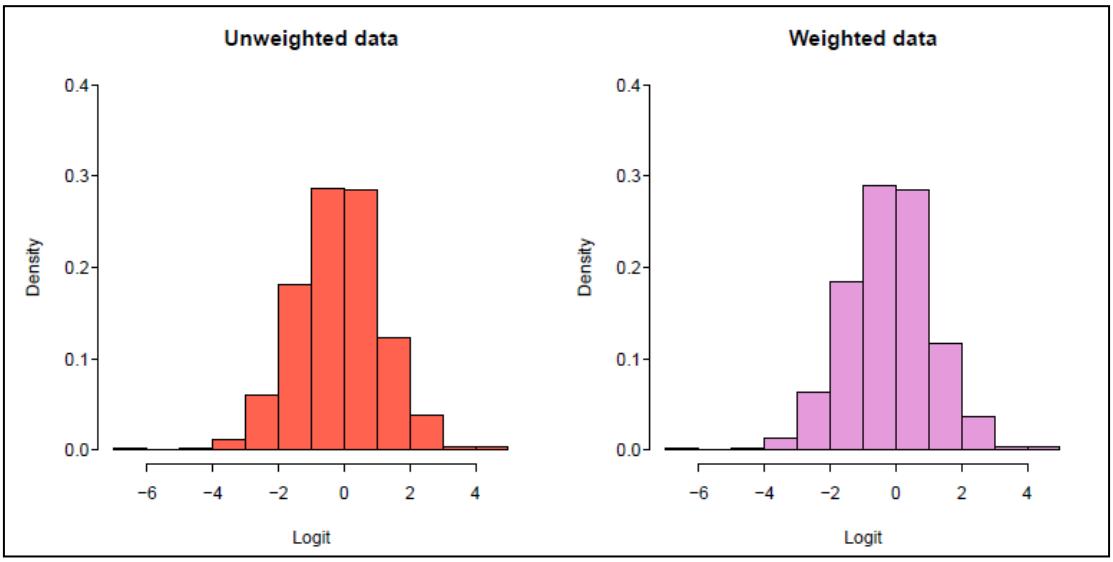

Figure A2.9 Year 4 reading achievement

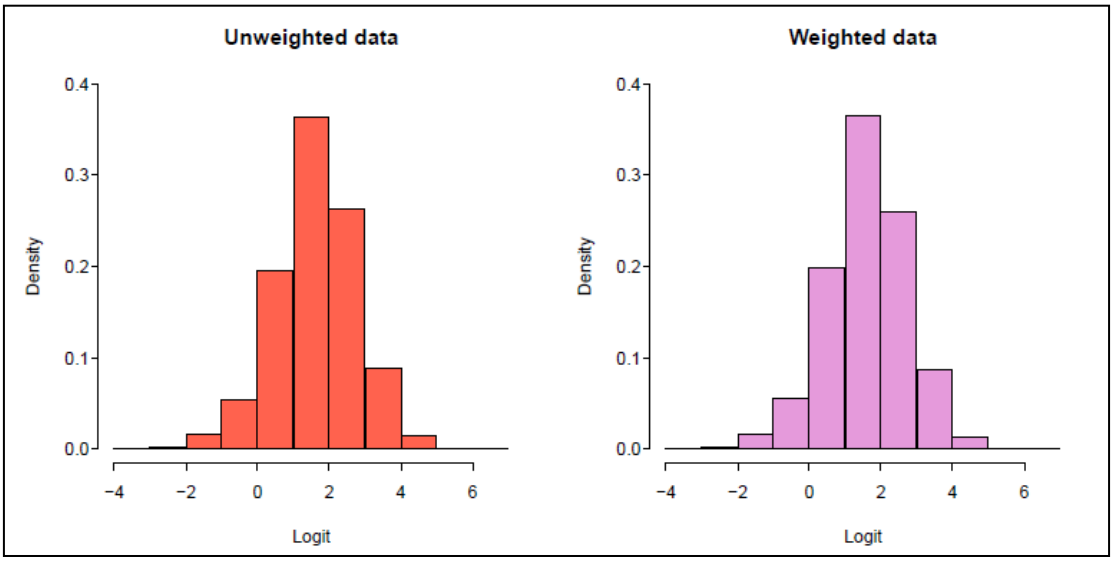

Figure A2.10 Year 8 reading achievement

Very slight differences can be seen across all subgroups in the mean and standard deviation estimates. See Tables A2.3 and A2.4. However, using weighted data would be very unlikely to change any of the inferences made in NMSSA about reading achievement in New Zealand schools. All differences are well within the error bounds given by 95 percent confidence intervals.

Table A2.3 Comparison of means calculated with unweighted and weighted data

NMSSA Year 4 Reading sample (n=2136)						
	Unweighted mean (logits)	**Weighted mean (logits)**	**Difference**	**Unweighted sd (logits)**	**Weighted sd (logits)**	**N**
Overall mean	-0.16	-0.18	-0.02	1.33	1.32	2136
Girls	0.04	0.05	0.01	1.31	1.30	1134
Boys	-0.38	-0.41	0.03	1.32	1.31	1002
NZE	0.06	0.01	-0.05	1.30	1.31	1352
Girls	0.27	0.25	-0.02	1.22	1.22	708
Boys	-0.17	-0.22	-0.05	1.35	1.35	644
Māori	-0.63	-0.60	0.03	1.29	1.28	468
Girls	-0.39	-0.34	0.05	1.34	1.35	254
Boys	-0.90	-0.85	0.05	1.16	1.16	214
Pasifika	-0.78	-0.77	0.01	1.16	1.16	268
Girls	-0.57	-0.51	0.06	1.13	1.11	148
Boys	-1.04	-1.02	0.02	1.14	1.16	120
Asian	0.06	0.04	-0.02	1.22	1.20	262
Girls	0.22	0.24	0.02	1.34	1.32	146
Boys	-0.15	-0.16	0.01	1.02	1.02	116
Quintile 1	-0.85	-0.81	0.04	1.19	1.20	371
Quintile 2	-0.55	-0.56	-0.01	1.34	1.33	305
Quintile 3	-0.15	-0.18	-0.03	1.28	1.28	416
Quintile 4	0.09	0.05	-0.04	1.33	1.34	372
Quintile 5	0.25	0.22	-0.03	1.23	1.22	672

Table A2.4 Social studies achievement for Year 8 students (IA sample)

NMSSA Year 8 Reading sample (n=2146)						
	Unweighted mean (logits)	**Weighted mean (logits)**	**Difference**	**Unweighted sd (logits)**	**Weighted sd (logits)**	**N**
Overall mean	1.63	1.61	-0.02	1.15	1.14	2146
Girls	1.84	1.83	-0.01	1.08	1.07	1057
Boys	1.42	1.40	-0.02	1.17	1.17	1089
NZE	1.86	1.83	-0.03	1.10	1.11	1346
Girls	2.09	2.08	-0.01	0.99	0.98	666
Boys	1.63	1.59	-0.04	1.16	1.17	680
Māori	1.24	1.25	0.01	1.11	1.10	474
Girls	1.48	1.47	-0.01	1.04	1.05	224
Boys	1.02	1.05	0.03	1.12	1.10	250
Pasifika	1.11	1.14	0.03	1.03	1.03	286
Girls	1.37	1.41	0.04	0.97	0.97	142
Boys	0.85	0.86	0.01	1.02	1.03	144
Asian	1.56	1.58	0.02	1.00	0.99	203
Girls	1.64	1.67	0.03	1.03	1.00	104
Boys	1.46	1.49	0.03	0.97	0.97	99
Quintile 1	1.02	1.08	0.06	1.07	1.09	244
Quintile 2	1.24	1.28	0.04	1.16	1.15	396
Quintile 3	1.68	1.66	-0.02	1.10	1.10	513
Quintile 4	1.80	1.80	0.00	1.10	1.09	461
Quintile 5	2.00	1.96	-0.04	1.05	1.06	532

Summary graphics

The following graphics compare weighted and unweighted distributions for NMSSA English: Reading 2014.

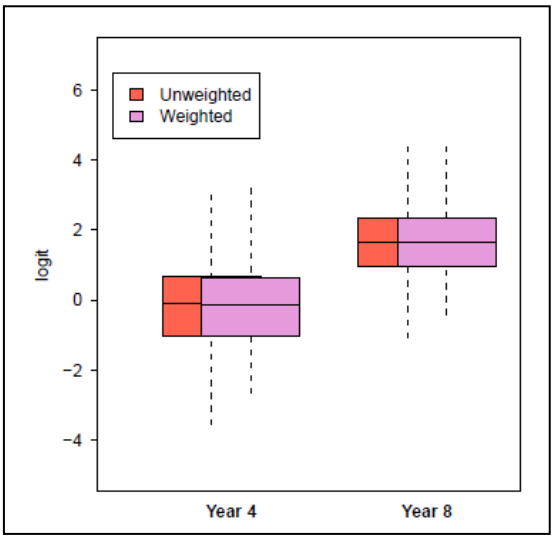

Figure A2.11 Comparison of weighted and unweighted reading scores

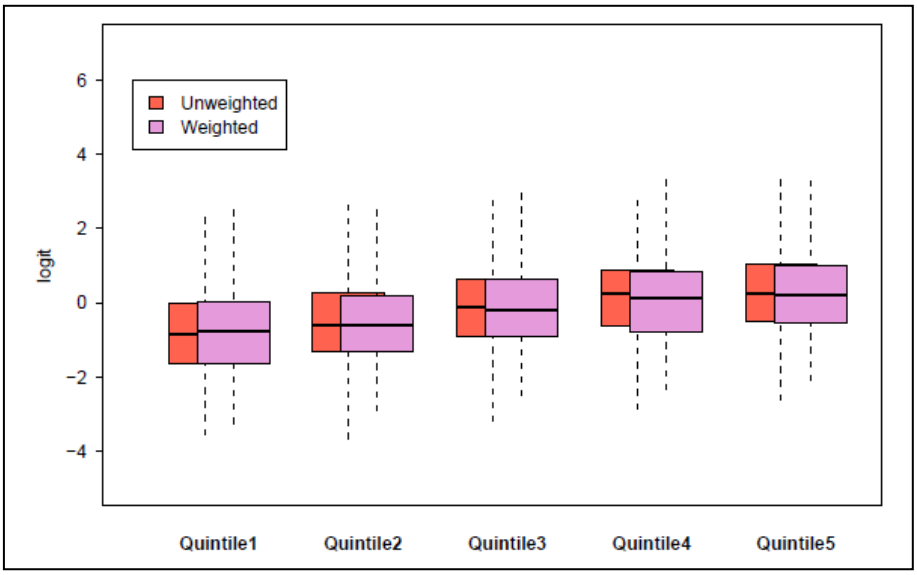

Figure A2.12 Comparison of Year 4 scores by quintile

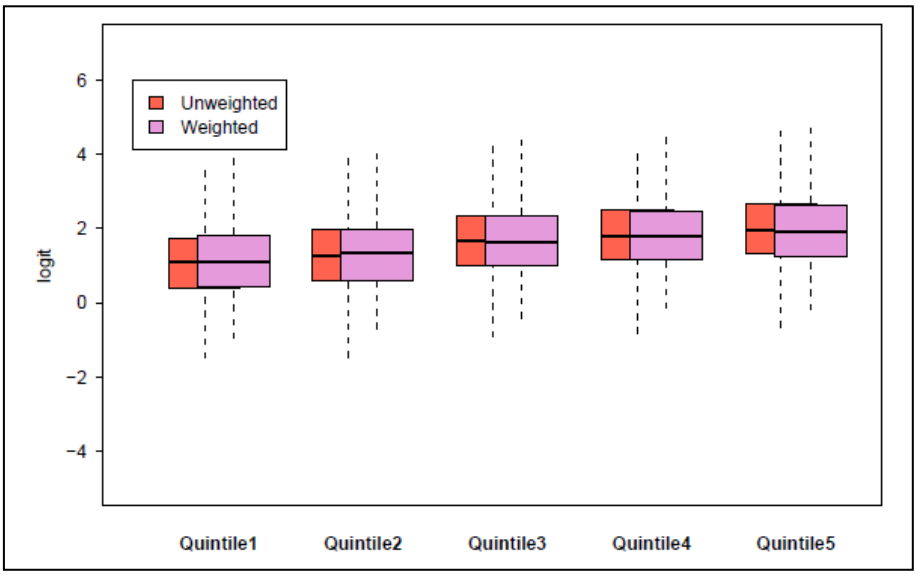

Figure A2.13 Comparison of Year 8 scores by quintile

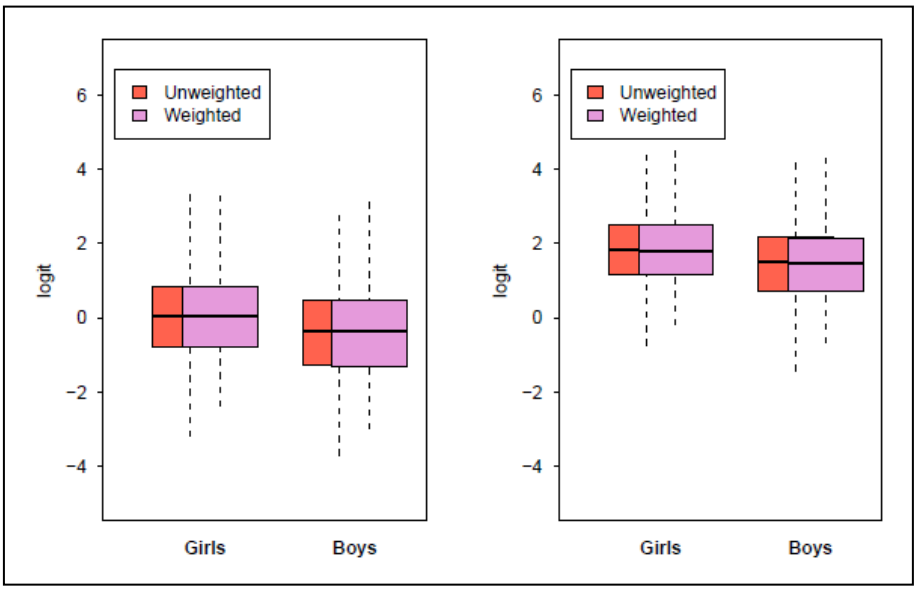

Figure A2.14 Comparisons by gender

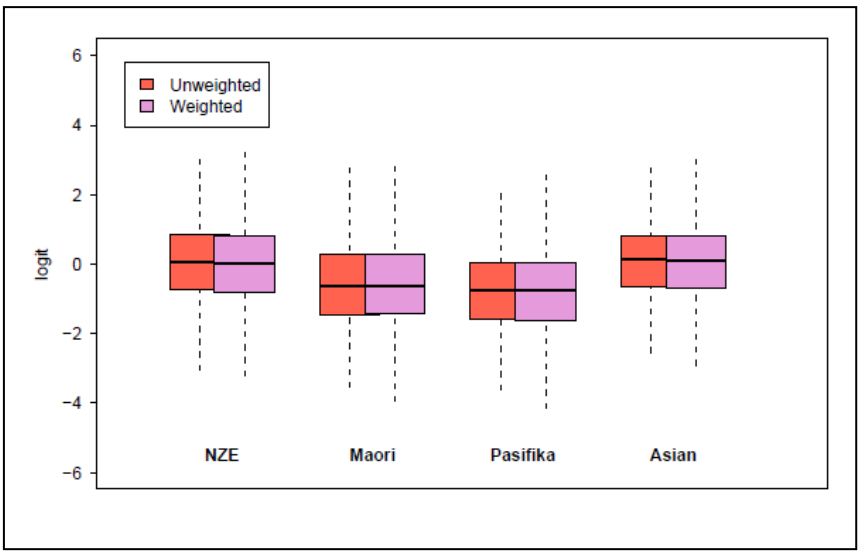

Figure A2.15 Comparisons by ethnicity - Year 4

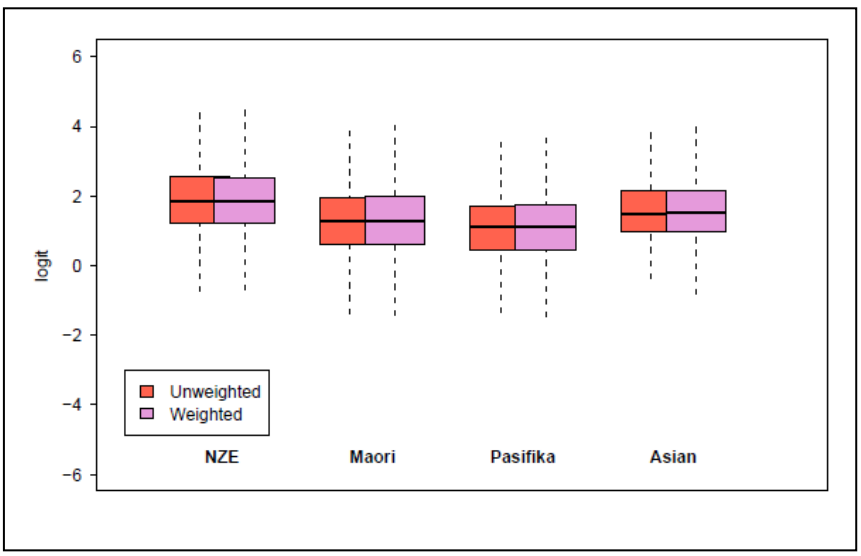

Figure A2.16 Comparisons by ethnicity - Year 8

4. Individual Assessment (IA) sample

The estimated social studies achievement distribution is based on the randomly selected subsample of eight students from each selected school. It is important to assess how weighting would affect the outcomes for the IA sample overall, and the subgroups of interest (e.g. Māori, Pasifika).

Weights were calculated for the subsample on the same basis as they were for the GA sample - that is by using post-strata of quintile, gender and ethnicity.

Missing data was not an issue for the subsample as students who were unable to participate were replaced with other students with similar characteristics.

Figure A2.17 shows that there is a comparatively narrow range of weights as there is for the GA sample. That is, weights are closely clustered around 1.

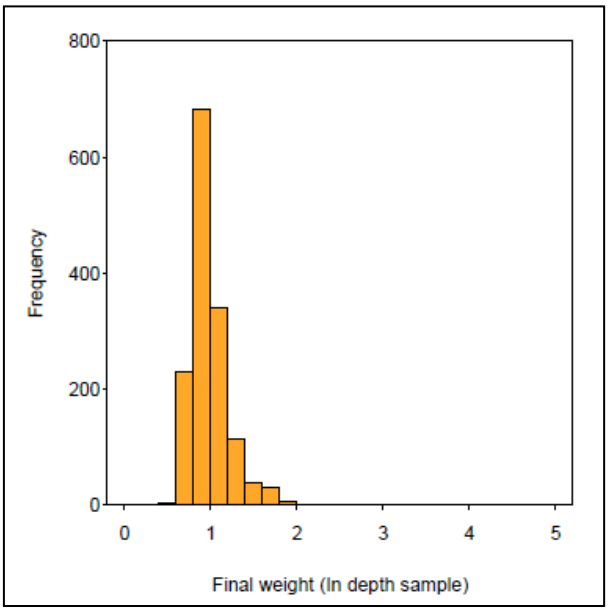

Figure A2.17 Distribution of weights calculated for the IA sample

Table A2.5 Distribution of weights for the IA sample

	Weight	Meaning
1st percentile	0.71	1% of weights were < 0.71
25th percentile	0.86	25% of weights are < 0.86
50th percentile	0.95	50% of weights are < 0.95
75th percentile	1.03	75% of weights are < 1.03
99th percentile	1.73	99% of weights are < 1.73
Maximum weight	3.86	Year 8, Quintile 1, Female, NZ European (n=2)
Minimum weight	0.58	Year 4, Quintile 1, Female, Māori & Pasifika (n=3)

As with the GA sample, comparisons of achievement scores for the IA sample (Figures A2.18 and A2.19) show very little difference between weighted and unweighted data.

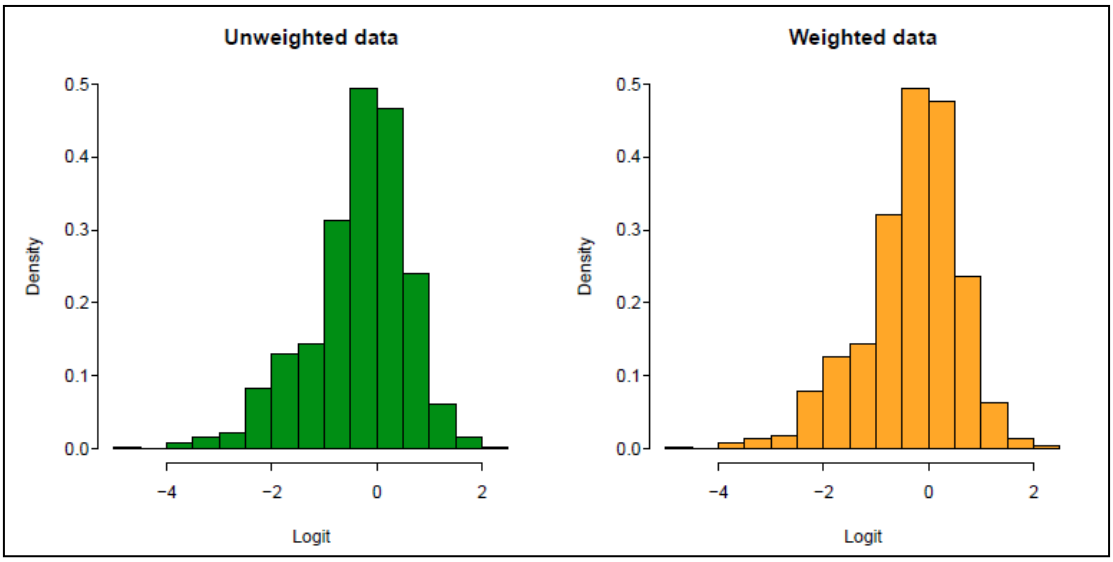

Figure A2.18 Year 4 social studies achievement

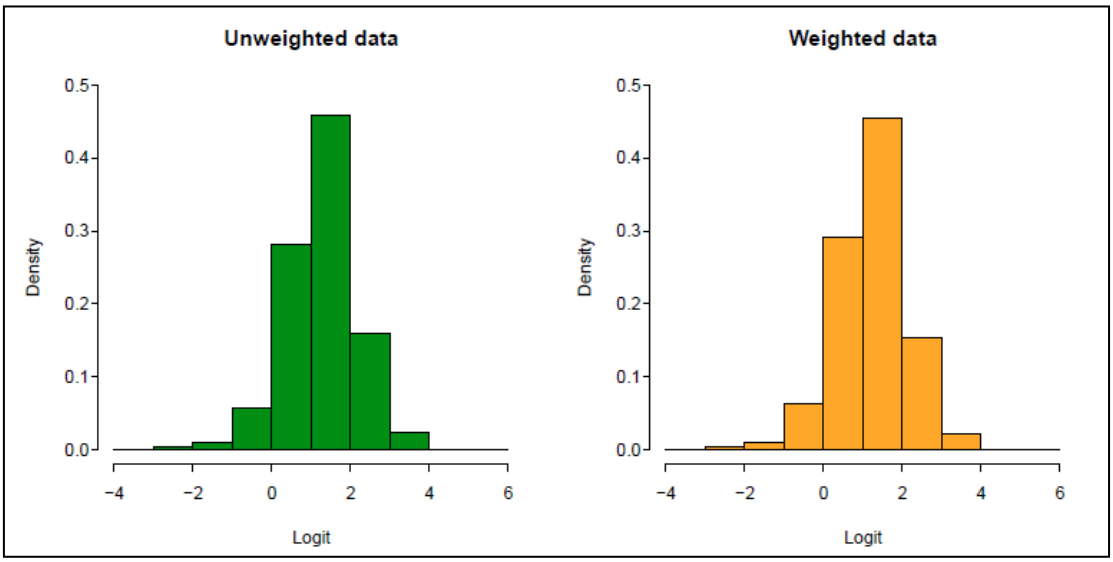

Figure A2.19 Year 8 social studies achievement

Tables A2.6 and A2.7 compare social studies means and standard deviations for the key subgroups in Year 4 and Year 8. As can be seen differences are small in general. Where differences are more noticeable, the sample sizes are generally small (e.g. Māori/Pasifika by gender), and the differences are well within reported 95 percent confidence intervals.

Table A2.6 Social studies achievement for Year 4 students (IA sample)

NMSSA Year 4 Social Studies sample (n=718)						
	Unweighted mean (logits)	**Weighted mean (logits)**	**Difference**	**Unweighted sd (logits)**	**Weighted sd (logits)**	**N**
Overall SS score	-0.33	-0.31	0.02	0.98	0.96	718
Girls	-0.31	-0.26	0.05	1.04	1.00	403
Boys	-0.36	-0.36	0.00	0.90	0.90	315
NZE	-0.05	-0.06	-0.01	0.77	0.77	446
Girls	0.01	0.05	0.04	0.81	0.81	245
Boys	-0.13	-0.14	-0.01	0.71	0.71	201
Māori	-0.68	-0.59	0.09	1.12	1.10	154
Girls	-0.61	-0.49	0.12	1.16	1.11	98
Boys	-0.81	-0.73	0.08	1.05	1.07	56
Pasifika	-1.06	-1.00	0.06	1.09	1.08	77
Girls	-1.01	-0.9	0.11	1.12	1.10	45
Boys	-1.13	-1.09	0.04	1.07	1.06	32
Asian	-0.46	-0.46	0.00	0.94	0.92	84
Girls	-0.49	-0.47	0.02	1.03	1.00	51
Boys	-0.43	-0.43	0.00	0.79	0.80	33
Quintile 1	-0.99	-0.88	0.11	1.18	1.14	135
Quintile 2	-0.61	-0.56	0.05	0.94	0.93	99
Quintile 3	-0.24	-0.29	-0.05	0.89	0.90	148
Quintile 4	-0.08	-0.09	-0.01	0.77	0.78	132
Quintile 5	0.03	0.02	-0.01	0.79	0.79	204

Table A2.7 Social studies achievement for Year 8 students (IA sample)

NMSSA Year 8 Social Studies sample (n=721)						
	Unweighted mean (logits)	**Weighted mean (logits)**	**Difference**	**Unweighted sd (logits)**	**Weighted sd (logits)**	**N**
Overall SS score	1.26	1.23	-0.03	0.92	0.92	721
Girls	1.30	1.26	-0.04	0.89	0.89	353
Boys	1.22	1.21	-0.01	0.95	0.94	368
NZE	1.45	1.41	-0.04	0.82	0.82	430
Girls	1.53	1.48	-0.05	0.76	0.76	204
Boys	1.37	1.34	-0.03	0.87	0.87	226
Māori	0.90	0.89	-0.01	1.02	1.02	161
Girls	0.87	0.81	-0.06	0.95	0.96	78
Boys	0.92	0.97	0.05	1.10	1.08	83
Pasifika	0.86	0.91	0.05	0.93	0.92	84
Girls	0.94	0.99	0.05	0.98	0.97	45
Boys	0.78	0.82	0.04	0.87	0.86	39
Asian	1.27	1.29	0.02	0.75	0.76	63
Girls	1.35	1.39	0.04	0.70	0.70	37
Boys	1.17	1.16	-0.01	0.81	0.82	26
Quintile 1	0.66	0.69	0.03	1.11	1.05	84
Quintile 2	0.89	0.91	0.02	0.93	0.93	126
Quintile 3	1.23	1.23	0.00	0.77	0.76	170
Quintile 4	1.45	1.45	0.00	0.80	0.79	153
Quintile 5	1.66	1.64	-0.02	0.81	0.80	188

5. Summary graphics

The following graphics compare weighted and unweighted distributions for NMSSA Social Studies 2014.

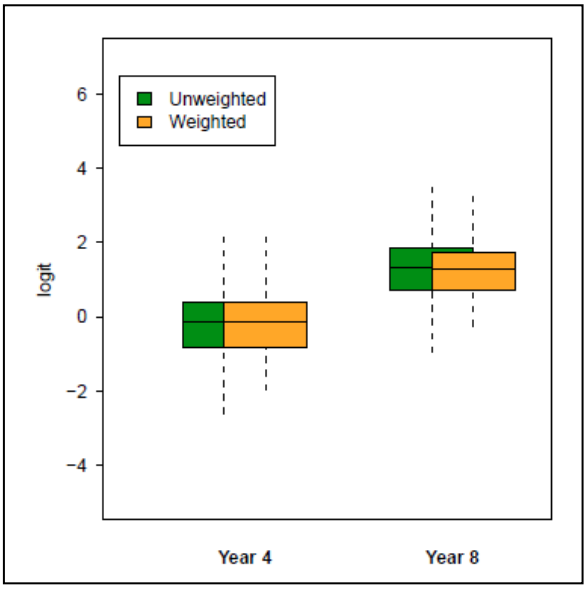

Figure A2.20 Comparison of weighted and unweighted social studies scores

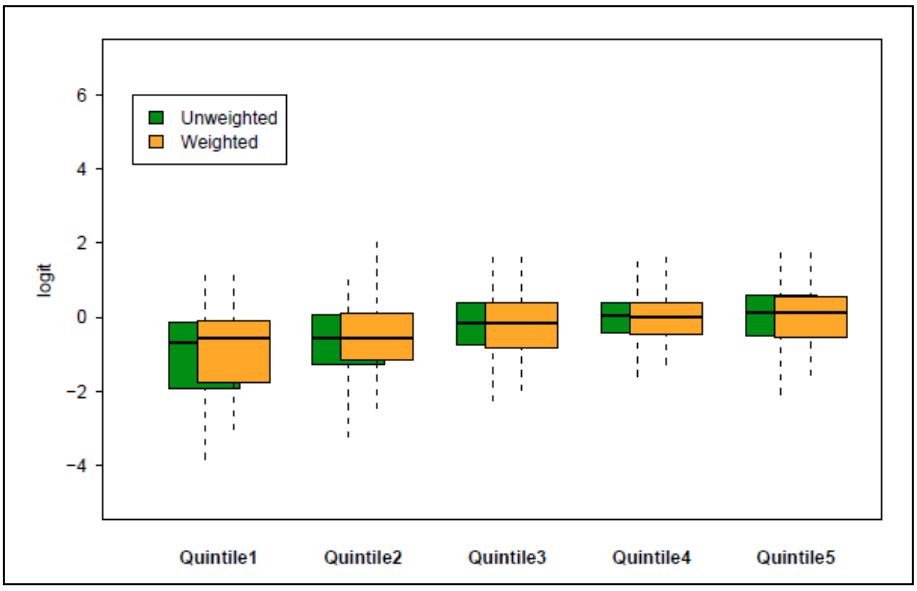

Figure A2.21 Comparison of Year 4 scores by quintile

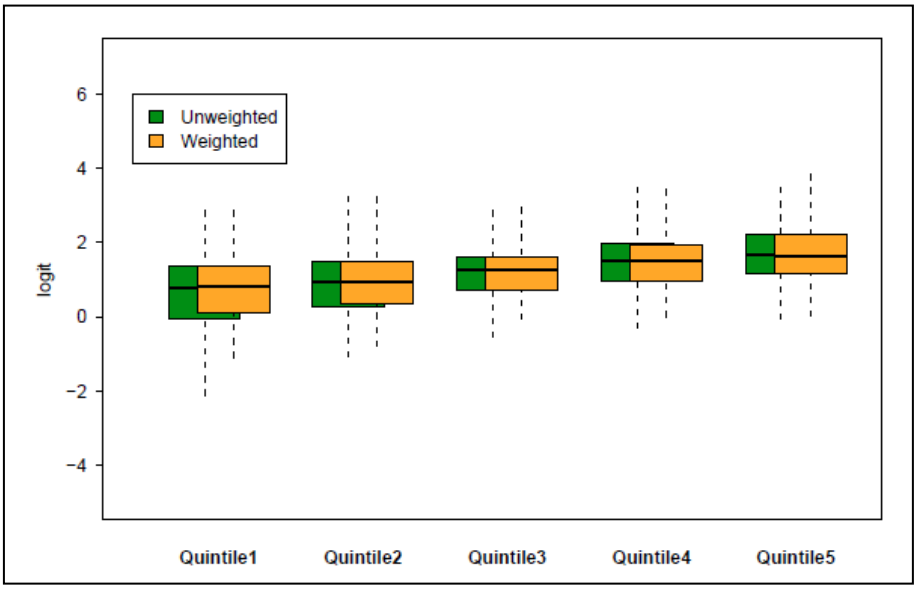

Figure A2.22 Comparison of Year 8 scores by quintile

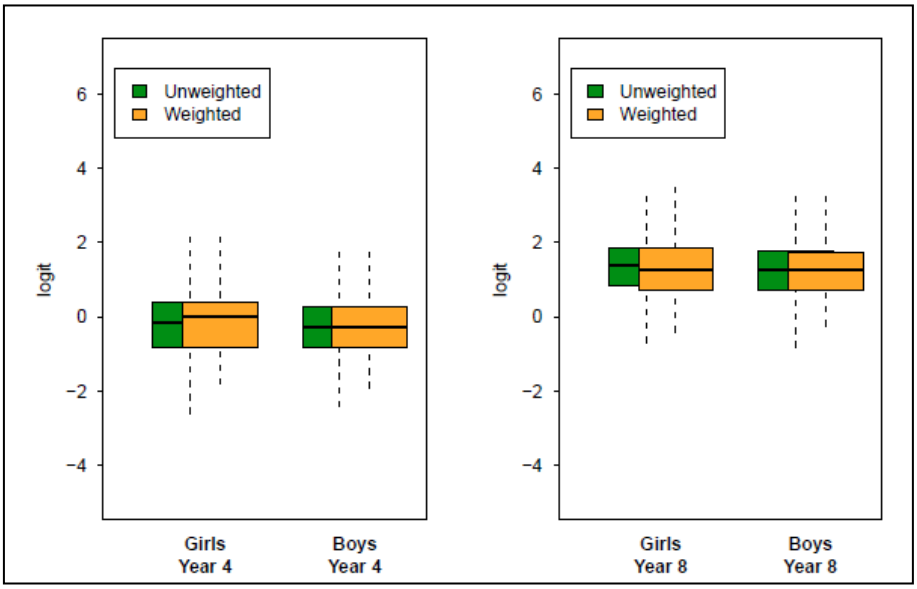

Figure A2.23 Comparisons by gender

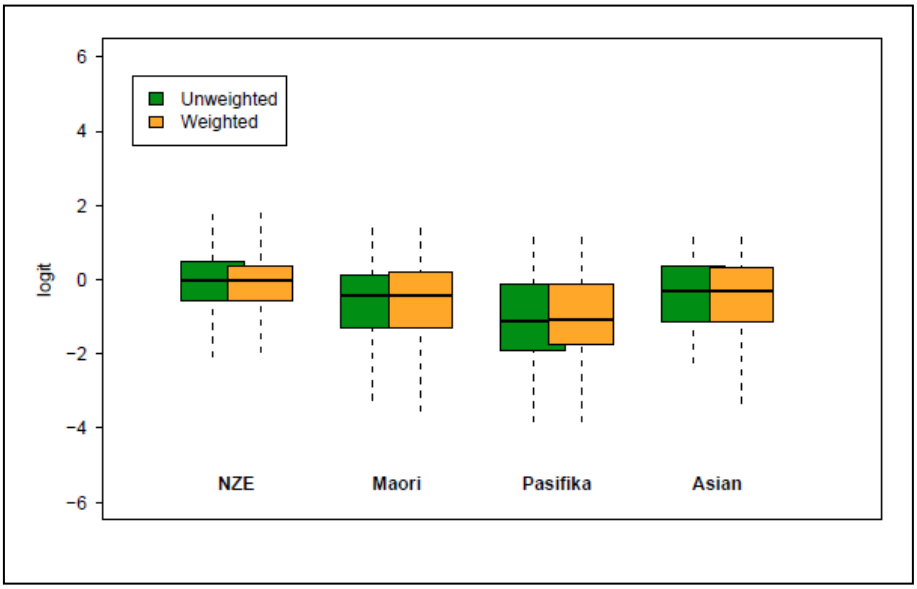

Figure A2.24 Year 4 social studies achievement by ethnicity

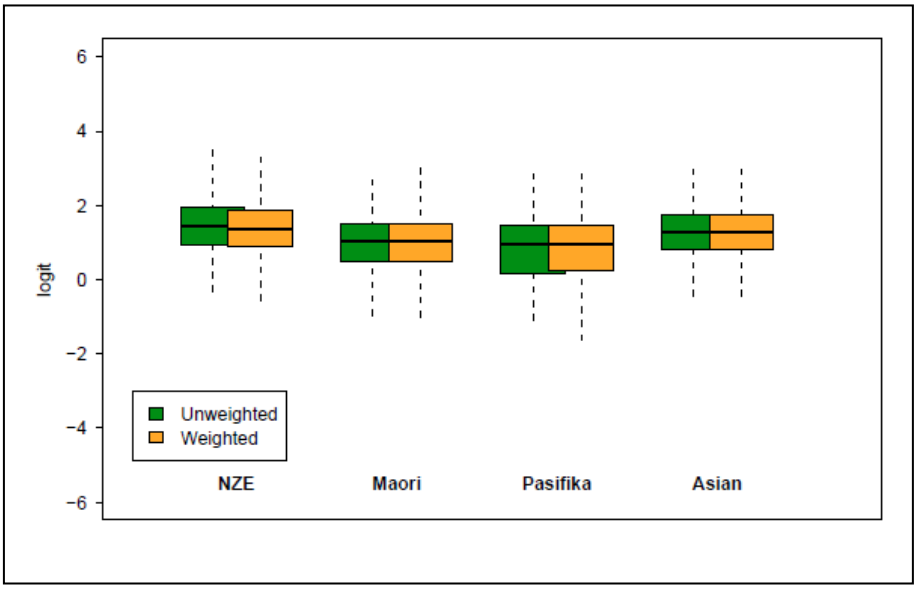

Figure A2.25 Year 8 social studies achievement by ethnicity

6. Conclusion

The cost of introducing analysis methods to account for sample weights would probably outweigh any gains made by including them. Estimates calculated with weighted data all fell well within the error bounds for unweighted estimates. Including weights in the analysis is very unlikely to change any national inferences. Therefore, it was recommended that NMSSA should **not** incorporate weights into the calculations of estimates for NMSSA 2014.

3 Appendix 3: Design effects for the 2014 NMSSA samples

1. Introduction

This report sets out the work undertaken to ascertain whether a robust approach to variance estimation in the 2014 NMSSA sample would have a significant impact on reported outcomes.

The report:

- introduces the concept of a design effect;
- describes the complex NMSSA sample and proposes a post-stratification design;
- discusses a range of possible variance estimation methods;
- compares results for key estimates calculated assuming a simple random sample versus a complex sample; and
- makes recommendations for how NMSSA should incorporate design effects into the 2014 analysis.

An early version of the 2014 data file was used in this analysis. Subsequent processing of the file resulted in updates to a small amount of the demographic data. This means that the numbers of students in different demographic groups quoted in this report are slightly different than those used in the final reporting.

Design effects

A design effect is the ratio of the variance of an estimate calculated for a complex sample design compared to the variance calculated as if the sample was a simple random sample.

$$d = \frac{Var(\theta)_{Complex}}{Var(\theta)_{SRS}}$$

A design effect may be understood as an inflation factor – the percentage increase in the variance of an estimator due to complex sampling methods compared to a simple random sample.

Design effects were calculated for all the main estimates used in NMSSA reporting.

Effective sample size

The design effect tells us the extent of the loss of efficiency in variance estimation caused by the complex sample design. This loss of efficiency can be couched in terms of an *effective sample size.* In a simple random sample (SRS) the sample size influences the precision (efficiency) with which estimates can be calculated. A decrease in the sample size leads to a decrease in efficiency, and subsequently an increase in the variance of an estimate. Using the design effect we can calculate the effective sample size, the size of a SRS that would give us the same efficiency as our complex sample.

$$n_{eff} = \frac{n_{complex}}{d(\hat{\theta})}$$

where n_{eff} = the effective sample size

$n_{complex}$ = the sample size selected under the complex design

d = design effect

θ = the estimate in question

2. Variance estimation for complex survey data

The NMSSA sample is a stratified cluster sample. Schools are the primary sampling unit and are stratified implicitly by region, decile, and size. 100 schools at each of year levels 4 and 8 are selected. Within selected schools, up to 25 students are systematically (randomly) selected rendering an equal probability sample of students representing the NZ student population.

For reporting purposes key variables are year level, decile, gender, and ethnicity. As student level variables gender and ethnicity are not catered for in the original stratification of schools, a post-stratification of the student sample incorporating these variables needed to be done to ensure that these groups were properly represented by the sample. Sample misrepresentation can be corrected by the use of weights (See Appendix 2: Weighting the NMSSA 2014 Samples).

Post-stratification and collapsing rules

The sample was post-stratified by quintile, gender and ethnic group. Throughout the NMSSA analysis care was taken to allow for multiple ethnicities. In this context, allowing for multiple ethnicities results in many very small post-strata. Approximately 11 percent of students at Year 4, and 9 percent of students at Year 8 reported belonging to multiple ethnicities.

For the purposes of variance estimation, Heeringa, West, & Berglund (2010, p.43) suggest that collapsing post-strata so that each contains a minimum of 15-25 members is advisable.

The Year 4 and Year 8 samples were treated separately, and the following collapsing rules were applied in order to small post-strata (i.e. less than 15 members). After each step strata were re-assigned and stratum size re-calculated. If there were remaining small strata, the next step was applied.

1. Remove 'other' classification from students who already belong to NZE/Māori/Pasifika/Asian
2. Small strata containing dual ethnicities are collapsed into prioritised ethnicity groups:
   Māori → Pasifika → Asian → NZE.
   Example: A small stratum specified by Quintile3-Female-Māori/Asian would be collapsed into the Quintile 3-Female-Māori stratum.
3. Collapse remaining small ethnicity strata into the appropriate gender group.
   Example: A small stratum identified by Quintile 4-Male-Pasifika would be collapsed into a Quintile 4-Male stratum
4. Remaining small strata are collapsed into the appropriate quintile stratum.
   Example: A small stratum identified by Quintile 1-Asian would be collapsed into a Quintile 1 stratum
5. Finally any small strata left make up a "mop-up" stratum, with no specific quintile, gender or ethnic identification.

A list of final post-strata for variance estimation can be found in Tables A3.1 and A3.2 of Section 5: Additional Tables.

3. Choosing a variance estimation method for NMSSA

Options

- Treat the sample as a simple random sample (SRS)
- Taylor Series Linearisation method (TSL)
- Jacknife method (JKn)
- Bootstrap method

Treating the sample as SRS

This option would be difficult to justify given that the NMSSA sample is not a simple random sample. Variances would be under-estimated.

Taylor Series Linearistaion (TSL)

Taylor series approximations of complex sample variances for sample estimates of means and proportions have been widely used since the 1950s (Heeringa et al., 2010). It is not a replication method like the jackknife and the bootstrap, but uses Taylor series approximations to estimate variances. It is most useful when using weighted data, but can be used with unweighted data. When the sample is reasonably standard the TSL method generally offers results similar to the Jackknife.

Jackknife (JKn)

The Jackknife is a replicate method. Sample replicates are formed by removing one school at a time within stratum. Weights in the affected stratum are scaled up to account for the deleted sample members. Sample estimates (for example, means) are then calculated for each replicate. The estimate of the variance is then a simple function of the differences between the replicate estimates and the estimate for the complete sample. (Heeringa et al., 2010, p. 77)

$$var(\hat{\theta}) = \sum_{all\ replicates} (\hat{\theta}_r - \hat{\theta})^2$$

Bootstrap

The Bootstrap is another replicate method commonly employed for variance estimation. It is generally useful when samples are small, or the underlying distributions are irregular. For the NMSSA data, neither of these situations applies, so the bootstrap was not investigated further.

4. Methods and process

Design effects for all major estimates were calculated using both TSL and JKn. Both methods produced very similar results. The results presented in this document are for the jackknife procedure. Details are in Tables A3.3 and A3.4 in Section 5: Additional Tables. The final decision to use the Jackknife method was taken after the decision to estimate of population statistics with unweighted data (See Appendix 2: Weighting the NMSSA2014 Samples).

5. Results and recommendations

Design effects varied between 1 and 3, mostly in the range of about 1.3 - 1.8 (i.e. a 30 to 80 percent inflation of variances due to complex sampling). While the design effects in some cases were fairly large (over 2.0 in a few cases), the effect on confidence intervals in practice was not. For the most part 95 percent confidence intervals around estimates increased in width by around 1 **scale score point**. A few confidence intervals increased more, but these tended to be already wide intervals around estimates for small sub-samples. For example, see reading achievement for Year 4 Pasifika males (Table A3.3), or reading achievement for Year 8 Pasifika females (Table A3.4) in Section 5.

The NMSSA technical team recommended that for ease of calculation, and to absorb most of the variance bias caused by the NMSSA complex sample design that a factor or multiplier of **0.67** should be used to reduce the sample size in standard error calculations. This assumes a design effect of 1.5 which is close to most design effects calculated. The effective sample size has been used for calculations of all confidence intervals around estimates of means, proportions, and effect sizes in NMSSA 2014.

Example: Calculate the standard error of a NMSSA mean

m_x = estimated mean of variable x

Under a simple random sample we would use

s_m = standard error of the mean = $\frac{s}{\sqrt{n}}$

Applying the recommended factor to account for a complex sample design we use

s_m* = standard error of the mean* = $\frac{s}{\sqrt{n*0.67}}$

6. Additional tables

Table A3.1 Year 4 post-stratification after collapsing

Quintile	Gender	NZE	Māori	Pasifika	Asian	Other	Cluster size
1	F	0	0	0	1	0	24
1	F	0	0	1	0	0	59
1	F	0	1	0	0	0	90
1	F	1	0	0	0	0	29
1	M	0	0	0	1	0	15
1	M	0	0	1	0	0	52
1	M	0	1	0	0	0	81
1	M	1	0	0	0	0	24
2	F	0	0	0	1	0	16
2	F	0	0	1	0	0	37
2	F	0	1	0	0	0	40
2	F	1	0	0	0	0	67
2	F	1	1	0	0	0	15
2	M	0	0	1	0	0	22
2	M	0	1	0	0	0	43
2	M	1	0	0	0	0	55
2	X	0	0	0	0	0	16
3	F	0	0	0	1	0	26
3	F	0	1	0	0	0	38
3	F	1	0	0	0	0	139
3	F	1	1	0	0	0	15
3	M	0	0	0	0	0	16
3	M	0	0	0	1	0	24
3	M	0	1	0	0	0	21
3	M	1	0	0	0	0	117
3	M	1	1	0	0	0	15
4	F	0	0	0	1	0	22
4	F	0	1	0	0	0	18
4	F	1	0	0	0	0	133
4	F	1	1	0	0	0	15
4	M	0	0	0	1	0	17
4	M	0	1	0	0	0	29
4	M	1	0	0	0	0	121
4	X	0	0	0	0	0	25
5	F	0	0	0	0	0	16
5	F	0	0	0	1	0	56
5	F	0	1	0	0	0	31
5	F	1	0	0	0	0	242
5	M	0	0	0	1	0	47
5	M	0	1	0	0	0	31
5	M	1	0	0	0	0	245
9	X	0	0	0	0	0	30

Table A3.2 Year 8 post-stratification after collapsing

Quintile	Gender	NZE	Māori	Pasifika	Asian	Other	Cluster size
1	F	0	0	0	0	0	20
1	F	0	0	1	0	0	54
1	F	0	1	0	0	0	41
1	M	0	0	1	0	0	46
1	M	0	1	0	0	0	63
1	M	1	0	0	0	0	18
2	F	0	0	0	1	0	18
2	F	0	0	1	0	0	44
2	F	0	1	0	0	0	56
2	F	1	0	0	0	0	83
2	M	0	0	0	0	0	17
2	M	0	0	1	0	0	40
2	M	0	1	0	0	0	66
2	M	1	0	0	0	0	77
3	F	0	0	0	1	0	16
3	F	0	0	1	0	0	18
3	F	0	1	0	0	0	45
3	F	1	0	0	0	0	157
3	F	1	1	0	0	0	18
3	M	0	0	0	1	0	19
3	M	0	0	1	0	0	17
3	M	0	1	0	0	0	37
3	M	1	0	0	0	0	172
3	M	1	1	0	0	0	16
4	F	0	0	0	0	0	30
4	F	0	0	0	1	0	29
4	F	1	0	0	0	0	140
4	F	1	1	0	0	0	16
4	M	0	0	0	1	0	23
4	M	0	0	1	0	0	18
4	M	0	1	0	0	0	45
4	M	1	0	0	0	0	162
5	F	0	0	0	0	0	22
5	F	0	0	0	1	0	27
5	F	0	1	0	0	0	35
5	F	1	0	0	0	0	200
5	M	0	0	0	1	0	27
5	M	0	1	0	0	0	26
5	M	1	0	0	0	0	196
9	X	0	0	0	0	0	36

Table A3.3 Comparison of variance estimation methods for Year 4 statistics

	Mean (SRS)	Mean (Jkn)	SE (SRS)	SE (Jkn)	CI (SRS) (lower)	CI (SRS) (upper)	CI (Jkn) (lower)	CI (Jkn) (upper)	Design effect	CI width increase	Variance inflation	N	Effective N
Y4.mean	85.6	85.6	0.48	0.64	84.6	86.4	84.3	86.7	1.7	0.64	31%	2174	1259
Y4.NZE	89.8	89.8	0.64	0.81	88.7	90.9	88.3	91.2	1.5	0.64	23%	1232	812
Y4.M	78.0	78.0	0.97	1.13	76.1	79.8	75.8	80.1	1.4	0.64	19%	482	339
Y4.P	73.4	73.4	1.29	1.93	70.6	75.9	69.5	77.1	2.0	2.26	40%	170	87
Y4.A	89.5	89.5	1.29	1.29	87.1	91.9	86.9	92.1	1.2	0.32	8%	247	211
Y4.female	88.8	88.8	0.64	0.81	87.5	90.0	87.2	90.3	1.6	0.64	28%	1128	690
Y4.male	82.1	82.1	0.64	0.97	80.8	83.5	80.5	83.8	1.7	0.64	31%	975	564
Y4.female.NZE	93.0	93.0	0.81	0.97	91.4	94.5	91.2	94.8	1.3	0.48	16%	655	486
Y4.female.M	81.3	81.3	1.29	1.61	78.7	83.8	78.2	84.5	1.4	1.13	19%	262	186
Y4.female.P	76.3	76.3	1.77	2.26	72.9	79.7	71.9	80.5	1.6	1.77	28%	96	59
Y4.female.A	91.4	91.4	1.77	1.93	87.9	95.0	87.7	95.0	1.1	0.16	3%	144	136
Y4.male.NZE	86.3	86.3	0.97	1.13	84.5	88.0	84.0	88.3	1.5	0.81	22%	577	386
Y4.male.M	74.0	74.0	1.29	1.45	71.6	76.4	71.1	76.9	1.4	0.97	16%	220	163
Y4.male.P	69.5	69.5	2.09	3.54	65.3	73.7	62.7	76.3	2.6	5.16	62%	74	28
Y4.male.A	86.7	86.7	1.61	1.93	83.7	90.0	83.0	90.6	1.5	1.29	23%	103	69
Y4.Q1	74.5	74.5	0.97	1.29	72.6	76.4	72.1	76.9	1.6	0.97	25%	374	239
Y4.Q2	79.5	79.5	1.29	1.61	77.1	81.7	76.3	82.5	1.8	1.61	33%	311	176
Y4.Q3	85.6	85.6	0.97	1.13	83.5	87.5	83.2	87.9	1.4	0.64	17%	411	299
Y4.Q4	89.6	89.6	1.13	1.29	87.5	91.7	87.2	92.1	1.3	0.64	14%	380	293
Y4.Q5	92.4	92.4	0.81	0.97	90.9	93.8	90.6	94.1	1.4	0.64	17%	668	484

Table A3.4 Comparison of variance estimation methods for Year 8 statistics

	Mean (SRS)	Mean (Jkn)	SE (SRS)	SE (Jkn)	CI (SRS) (lower)	CI (SRS) (upper)	CI (Jkn) (lower)	CI (Jkn) (upper)	Design effect	CI width increase	Variance inflation	N	Effective N
Y8.mean	114.4	114.4	0.32	0.48	113.6	115.3	113.3	115.6	2.0	0.64	41%	2190	1104
Y8.NZE	118.6	118.6	0.48	0.64	117.7	119.6	117.3	119.9	1.8	0.64	33%	1255	712
Y8.M	107.8	107.8	0.81	0.97	106.2	109.5	105.9	109.8	1.4	0.64	18%	464	335
Y8.P	104.5	104.5	0.97	1.61	102.4	106.4	101.2	107.7	2.5	2.42	59%	237	93
Y8.A	114.6	114.6	1.29	1.29	112.0	117.2	112.0	117.2	1.0	0.00	-1%	159	163
Y8.female	117.8	117.8	0.48	0.81	116.7	118.8	116.4	119.3	1.9	0.81	39%	1069	550
Y8.male	111.1	111.1	0.64	0.81	109.9	112.2	109.6	112.7	1.9	0.81	38%	1085	570
Y8.female.NZE	122.2	122.2	0.64	0.81	120.9	123.5	120.6	123.8	1.6	0.64	28%	614	377
Y8.female.M	111.7	111.7	1.13	1.29	109.5	114.0	109.0	114.4	1.4	0.97	16%	211	156
Y8.female.P	107.7	107.7	1.45	2.74	104.9	110.4	102.2	113.2	3.9	5.48	97%	116	30
Y8.female.A	114.8	114.8	1.77	1.61	111.2	118.2	111.7	117.8	0.7	-0.81	-13%	90	120
Y8.male.NZE	115.3	115.3	0.81	0.97	113.8	116.7	113.5	117.2	1.6	0.81	28%	641	390
Y8.male.M	104.5	104.5	1.13	1.29	102.4	106.7	101.9	107.0	1.4	0.81	16%	253	187
Y8.male.P	101.2	101.2	1.45	1.61	98.3	104.1	98.3	104.3	1.1	0.16	4%	121	112
Y8.male.A	114.4	114.4	1.93	2.26	110.6	118.2	109.9	119.0	1.4	1.45	19%	69	48
Y8.Q1	104.6	104.6	1.13	1.77	102.4	106.9	101.2	108.0	2.3	2.26	52%	242	105
Y8.Q2	108.3	108.3	0.97	1.13	106.6	110.1	106.1	110.6	1.5	0.97	22%	401	268
Y8.Q3	114.9	114.9	0.81	0.97	113.3	116.4	113.0	116.9	1.6	0.81	25%	515	328
Y8.Q4	117.3	117.3	0.81	1.29	115.7	119.0	114.9	119.8	2.1	1.61	45%	463	220
Y8.Q5	120.4	120.4	0.64	0.81	119.1	121.9	118.8	122.2	1.5	0.64	22%	533	355

References

Heeringa, West, Berglund, 2010, *Applied Survey Data Analysis*, Taylor and Francis Group, LLC

Lumley, T., (2004) Analysis of complex survey samples. Journal of Statistical Software 9(1): 1-19

Software

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Lumley, T., (2014) *Survey: Analysis of complex survey samples*. R package version 3.30.

4 Appendix 4: The NMSSA Social Studies assessment framework

1. Assessment Framework for the Nature of Social Studies assessment

The assessment of the Nature of Social Studies was derived from the achievement objectives and focused on three constructs: students' conceptual understanding, active participation in society and values/perspectives across the strands of the social studies learning area.

Definition of constructs

The three constructs were defined as follows:

Conceptual Understanding These are big ideas that students develop about social studies concepts e.g. customs, change, continuity, choices. The concepts relate to the four conceptual strands of social studies in the New Zealand Curriculum: identity, culture and organisation; place and environment; continuity and change; and the economic world. Students were assessed on the extent to which they would be able to:

- understand the big ideas about social studies concepts;
- transfer the concept to a different context; and
- make connections between concepts.

Active Participation in Society is to be constructively involved in participating in, or observing, critically informed actions in relation to local or global issues. Students were assessed on the extent to which they would be able to:

- recognise the contribution of themselves or others in society;
- identify an issue or a problem;
- identify how they themselves or others can take action or make decisions (process); and
- identify opportunities for themselves or others to participate.

Values/Perspectives These are deeply held beliefs about what is important or desirable. They are expressed through the ways in which people think and act. Students were assessed on the extent to which they would be able to:

- expresses their own values position and provides a justification;
- recognise diverse values and perspectives; and
- consider values and the actions based on them.

Coverage map for the Nature of Social Studies assessment

Many questions in the NSS assessment covered more than one construct. Each question was scored with a partial credit scoring rubric (0, 1, 2). Table A4.1 shows the coverage of the NSS framework by task, strand, concept, construct, setting, and assessment approach. The Graffiti task, for example, covered all three constructs. The marking schedule for Graffiti is presented in Table A4.2. It describes the way that constructs overlap within each task.

Table A4.1 Coverage map for the Nature of Social Studies assessment

	Strand/Concept/s		Construct				
Task Title	**Strand**	**Concepts**	**Conceptual Understanding**	**Active Participation in Society**	**Values/ Perspectives**	**Setting**	**Assessment approach**
Fudge for the school fair	Economic world	Factors affecting pricing. Profit Factors influencing people's purchasing decisions	✓		✓	NZ	Interview
Volunteering	Identity, culture & organisation	Common good	✓	✓	✓	NZ Global	Interview
Kai Moana	Identity, culture & organisation Place & environment	Status of Māori as tangata whenua Use of resources Sustainability	✓		✓	NZ	Interview
Graffiti	Identity, culture & organisation Place & environment	How formal and informal groups make decisions How people view and use places differently	✓	✓	✓	NZ	Interview
Moving Here	Identity, culture & organisation Place & environment Continuity & change	Cultural diversity and interaction Effect of people on the environment Cultural interaction can change culture over time	✓		✓	Global NZ	Interview
School Garden	Identity, culture & organisation	How formal and informal groups make decisions Roles and responsibilities	✓	✓		NZ	Interview
Cultural Symbols	Identity, culture & organisation	What constitutes culture How symbols communicate identity	✓			Asia	Performance
When Disaster Strikes	Identity, culture & organisation	Social responsibility How people respond individually & collectively	✓	✓		Global NZ	Interview
Needs and Wants (Y4)	Economic world	Needs and wants	✓			Global	Performance
Consumers (Y8)	Economic world	Needs and wants Goods and services Rights and responsibilities	✓			Global NZ	Performance
Shopping	Identity, culture & organisation Continuity & change Economic world	Innovation creates opportunities and challenges for people	✓		✓	Global NZ	Interview

Table A4.2 Marking rubric for the Graffiti task

Title:	**Graffiti**	**Level:**	4 & 8
Code:	**SS14S013**	**Approach:**	Interview
Task Info:	**5 photos**		

Col 1

Q1. Is it okay to graffiti?
Q2 Why do you say that?

CONSTRUCT: Values / Perspectives

SCORE:	0	1	2
Criteria:	No response/don't know/ unsure Response not relevant Yes/No but no explanation	Basic/surface response e.g. No it is rude; you must follow the rules; it doesn't look good; no permission Yes it is artistic	Deep level response e.g. No somebody owns the building - vandalism (own values position taking into account others' values); could upset others; impact on others Yes freedom of expression
Express their own values and provide justification	Unable to identify to their own values position	Provides a basic / surface justification for their position	Provides a complex / deep justification for their position

Col 2

Q3. Why might people graffiti? Try to think of 3 reasons.

CONSTRUCT: Values / Perspectives

SCORE:	0	1	2
Criteria:	No response/don't know/ unsure Response not relevant	Can give surface reasons e.g. bored; art; fun; it looks cool	Can give deep level reasons e.g. to express feelings/anger/protest; mark their place; look cool to their friends; so that people see their message (person and society focussed)
Recognising diverse values and perspectives	Unable to explain others' values positions	Explains others' values on a simple / surface level	Explains others' values on a complex / deep level

Col 3

Q4. Why might people want this graffiti to stay on the school?
Q5. Why might people want to get rid of this graffiti?

CONSTRUCT: Values / Perspectives Conceptual Understanding

SCORE:	0	1	2
Criteria:	No response/don't know/ unsure Response not relevant	Can give surface reasons e.g. they might/might not like what it looks like; they didn't get permission; artistic; not rude; destroys school property	Can give deep level reasons e.g. not appropriate for young people to view; sets a bad example; may influence others to copy; freedom of expression; not offensive
Recognizing diverse values and perspectives	Unable to explain others' values positions	Explains others' values on a simple / surface level	Explains others' values on a complex / deep level
Conceptual Understanding ***-*** *how people use and see places differently*	Demonstrates no understanding of the concepts	Demonstrates understanding of concepts (surface)	Demonstrates understanding of abstract concepts (deep)

Col 4	Q6. How could the school community **decide** whether to keep the graffiti or remove it? Try to think of **2** different ways they could decide. Q7. Why is _____ a good way to decide? (Insert response to Q6)	CONSTRUCT: Active Participation in Society Conceptual Understanding

SCORE:	0	1	2
Criteria:	No response/don't know/ unsure Response not relevant	Simple process e.g. vote because it is fair/no arguments – majority rules; everyone gets a say; lists people e.g. ask the teachers; let the teachers/principal decide as they are in charge	Complex process e.g. vote but representing the voice of the school; process links to people's opinions and values - debate; process understands the need to include many different groups - survey; decide on what's important to the school – set criteria
Identify how they themselves or others can take action or make decisions based in knowledge and understanding *– process*	Unable to describe a social decision making process to solve an issue or problem	Describes a simple social decision making process	Describes a complex social decision making process
Conceptual Understanding – *how formal and informal groups make decisions*	Demonstrates no understanding of the concepts	Demonstrates understanding of concrete concepts	Demonstrates understanding of abstract concepts

5 Appendix 5: Curriculum alignment in NMSSA Social Studies 2014

1. Introduction and background

The underlying objective of NMSSA is to report on student achievement with respect to the *New Zealand Curriculum* (NZC). Assessment data in different learning areas are collected each year and achievement scales constructed in order for this to be carried out. In 2014, social studies was one of the two learning areas under consideration. This paper describes the process and presents results for the alignment of the NMSSA Nature of Social Studies (NSS) scale with the NZC.

An alignment of an achievement scale to the NZC has not been attempted before in this learning area. The process described here has generated some useful discussion and learning particularly in regard to how conceptual understanding is 'measured' in a national monitoring context.

NMSSA defines social studies as follows:

> *Social studies is about how societies work and how people (including the students themselves) can participate and take action as critical, informed, and responsible citizens. (New Zealand Curriculum)*

Figure A5.1 shows a complete overview of the NMSSA social studies assessment development process.

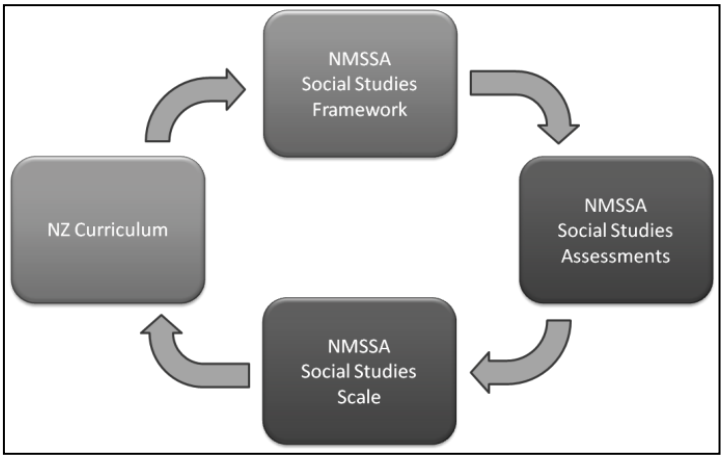

Figure A5.1 Overview of the assessment development for social studies in NMSSA

The NSS framework drew on the NZC document. In particular the NMSSA framework focuses on:

- Conceptual understandings
- Active participation in society
- Values and perspectives

The framework was then used to inform the NSS assessment tasks and scoring rubrics. The assessment mode almost exclusively consisted of one-to-one interviews with students. Verbal responses given by students were captured on video. Scoring was carried out later as a group activity.

The NSS scale was constructed from the scored interview data.

2. Alignment design and method

A panel of six external subject matter experts was invited to participate in the alignment exercise. Each panel member brought curriculum expertise, together with classroom and teaching experience in the social studies learning area. NMSSA researchers and psychometricians also formed part of the alignment team.

The NSS framework was described in detail by NMSSA researchers. A thorough understanding of the framework and knowledge of its construction, appreciation of practical limitations in the NMSSA context, and a good understanding of the NSS achievement scale were all necessary before the panel could make informed judgements about the relationship of curriculum levels to the NSS scale. Some time was spent on each aspect (framework, limitations, and scale) in turn, with questions and discussion encouraged at all times.

Five of the original eleven tasks were selected for the alignment exercise. The five tasks were considered to provide good coverage of the framework, and some varied contexts. Each task was examined separately.

Each task

Each task was described in detail with respect to the following aspects:

- Resources used - for example, pictures/photos, recordings, magnet boards and cards.
- What students were asked to do, and what questions they were asked.
- What the task looked like on video - showing actual footage of responses being given.
- How the responses were scored - scoring rubrics were presented and examined.

Each task was supported by 30 to 40 transcribed responses covering a full range of scores on the task. Panel members were able to see how the total scores for the task were reached. For instance a score of 6 out of a possible 10 on five questions could have been gained with 2+2+2+0+0 or 2+1+1+0+2 or some other combination.

Assessment conditions

The conditions under which students were assessed in social studies was addressed. The assessment was carried out (necessarily) under circumstances that could have detracted from students' ability to demonstrate the breadth of their knowledge and skill.

When thinking about question difficulty and about how certain groups of students could have performed on those questions, the panel was asked to take the following points into consideration.

- Students had reader/writer support and prompts if necessary from teacher assessors.
- Students were asked to respond to someone they did not know.
- Students responded verbally “on the spot”.
- Time to think and consider answers was very limited.
- There was no classroom or peer discussion to help generate ideas.
- Students had no 'scaffolding' in the form of a class module/project.

Minimal competence at different curriculum levels

The panel was asked to imagine a group of 100 students all performing at exactly the same curriculum level. Two levels were considered - borderline curriculum Level 2 and borderline curriculum Level 4.

The concept of 'minimal competence' at a curriculum level was thoroughly discussed until a common understanding was reached. A number of definitions of minimal competence were considered and discussed. Panel members were encouraged to use the one(s) that best helped them visualise a group of students at the appropriate level. The following definitions were discussed for Level 2.

- A minimally competent student **just** (barely) meets the curriculum expectations at Level 2.
- A minimally competent student has **just** enough of the requisite knowledge and skill to perform most of the time according to Level 2 expectations, although their knowledge may be limited.
- A minimally acceptable candidate is **borderline** Level 2.

- A student who is minimally competent at level 2 has done **just enough** to be described as someone performing in/at Level 2.
- Minimally competent students are deemed to be operating at/in level 2, but **only just**.
- Over a number of tasks and contexts, on average this student will produce performances that are overall **just good enough** to mean Level 1 is not an appropriate descriptor.

Distribution of scores

Panel members were asked to make judgements about how they thought the minimally competent group would have fared on a task overall. For each task each panel member filled in a grid to show how they thought the 100 students' scores would be distributed. Figure A5.2 gives an example for one task filled in by one panel member.

Task Score	Level 2	Level 4
8		10
7		10
6		30
5		30
4	10	20
3	40	
2	40	
1	10	
0		

Figure A5.2 Example of a judgement grid

Some panel members preferred to enter their judgements graphically as shown in Figure A5.3. This was also acceptable.

Task Score	Level 2	Level 4
8		●
7		●
6		●●●
5		●●●
4	●	●●
3	●●●●	
2	●●●●	
1	●	
0		

Figure A5.3 Example of how a judgement grid could be filled in

Progress was monitored throughout the day. Judgements were challenged if it was thought that alignment was seriously 'off-target' for a particular task, or if judgements varied widely between panel members. The challenges provided starting points for lively discussion around the social studies curriculum document itself, the NMSSA question objectives, how students respond in particular situations, and the nature of how scores are located on the NSS scale. While complete agreement was not a requirement, some individual judgements were re-considered and updated as a result of these discussions.

3. Results

The estimated distributions were amalgamated, and averaged to find a score on each task. The average scores were summed across the tasks for each curriculum level, and located on the NSS scale. Table A5.1 shows the scale locations for the beginning of Level 2 and the beginning of Level 4. Level 3 was taken to be half-way between Level 2 and Level 4.

Table A5.1 Final curriculum levels aligned with the NSS scale

Beginning of	NSS scale location
Level 2	82.0
Level 3	102.3
Level 4	122.6

6 Appendix 6: Regression analysis: the interaction between ethnicity and decile for social studies

1. Introduction

Reporting on differences between ethnic groups in New Zealand is complicated on two counts. First, a high proportion of Māori and Pasifika students attend lower decile schools, and a much lower proportion attend high decile schools. This creates a skew in the distribution of all ethnic subgroups with respect to decile and means that any score differences between ethnic groups could be explained by general differences between performance levels at different deciles and vice-versa.

The second complication is that students may identify with more than one ethnic group. It is difficult to make useful, robust statistical statements with respect to performance in ethnicity subgroups when there is 'blurring' with regard to group membership.

To explore the performance of ethnic groups on the Nature of Social Studies measure across deciles the following regression analyses were carried out:

- a comparison of Māori and NZ European students' scores on the Nature of Social Studies; and
- a comparison of Pasifika and NZ European students' scores on the Nature of Social Studies.

This paper describes the regression analysis and presents the results.

2. The regression models

For the purposes of the analysis, the school decile band was recoded to quintile[5]. Table A6.1 shows the number of students in each ethnic group by quintile at each year level.

Table A6.1 Numbers in each ethnic group according to quintile and year level

Year	Quintile	NZ European	Maori	Pasifika
4	1	33	70	44
	2	52	40	25
	3	119	29	8
	4	117	22	4
	5	171	18	5
8	1	14	45	34
	2	75	43	30
	3	122	43	10
	4	118	30	15
	5	143	20	7

As evidenced in the table, the ability to precisely assess how Māori or Pasifika students are performing, on average, in higher decile schools (and how NZ European students are performing in lower decile schools) is compromised by the low numbers of students representing the respective ethnic group across the quintiles. The results from this analysis should therefore be interpreted with caution.

[5] Decile 1-2 → Quintile 1, Decile 3-4 → Quintile 2, … , Decile 9-10 → Quintile 5

For each year level and for both Māori and Pasifika subgroups, separate models were run to examine the effect on performance outcomes due to quintile and ethnicity. These are described below.

Models comparing Māori and NZ European subgroups

1. $Y_i = \alpha + \beta1_i$*quintile+ $error_i$

2. $Y_i = \alpha + \beta1_i$*quintile +$\beta2_i$*Māori + $\beta3_i$*NZE + $error_i$

3. $Y_i = \alpha + \beta1_i$*quintile +$\beta2_i$*Māori + $\beta3_i$*NZE +$\beta4_i$*(Māori*quintile) + $\beta5_i$*(NZE*quintile) + $error_i$

Models comparing Pasifika and NZ European subgroups

1. $Y_i = \alpha + \beta1_i$*quintile+ $error_i$

2. $Y_i = \alpha + \beta1_i$*quintile +$\beta2_i$*Pasifika + $\beta3_i$*NZE + $error_i$

3. $Y_i = \alpha + \beta1_i$*quintile +$\beta2_i$* Pasifika + $\beta3_i$*NZE +$\beta4_i$*(Pasifika *quintile) + $\beta5_i$*(NZE*quintile) + $error_i$

In the model statements, Y_i is the Nature of Social Studies scale score for student *i*, and quintile, Māori, Pasifika and NZ European (NZE) are all classification ('dummy') variables.

For each model, there was a strong and statistically significant quintile effect. Average scores overall increased consistently with quintile.

Results for Māori/NZ European models

The R^2 statistic, indicating the proportion of variance in the Nature of Social Studies scores accounted for by each model is presented below.

Table A6.2 R^2 results by model and year level for Māori/NZ European models

Year	Model	R^2	R^2 change*
4	1	.13	-
	2	.17	**.04**
	3	.20	**.02**
8	1	.16	-
	2	.18	**.02**
	3	.20	.02

* R^2 change values in bold indicate statistically significant (p<.05) improvement from the previous model.

Models were compared using the usual F-test[6]. At Year 4, Model 3 was found to be the most parsimonious, whereas at Year 8 it was Model 2 in the context of the variables of interest. The model with additional interaction terms (Model 3) showed a significant, but small improvement for Year 4, but no significant improvement for Year 8 over the main effects model (Model 2).

[6] $F = \frac{\left(\frac{Drop\ in\ SSE}{Number\ of\ added\ terms}\right)}{s^2\ for\ the\ full\ model}$, where SSE = Sum of the squared residuals in the respective model.

Results for Pasifika/NZ European models

The R^2 statistic, indicating the proportion of variance in the Nature of Social Studies scores accounted for by each model is presented below.

Table A6.3 R^2 results by model and year level for Pasifika/NZ European models

Year	Model	R^2	R^2 change*
4	1	.12	-
	2	.21	**.09**
	3	.22	.01
8	1	.10	-
	2	.13	**.03**
	3	.14	.01

* R^2 change values in bold indicate statistically significant ($p<.05$) improvement from the previous model.

The models were compared using the usual F-test. At each of Year 4 and Year 8, Model 2 was found to be the most parsimonious in the context of the variables of interest. The models with additional interaction terms (Model 3) showed no significant improvement over the main effects model (Model 2).

3. Summary

Figures A6.1 and A6.2 show Nature of Social Studies average scores and their associated confidence intervals by quintile, for NZ European students compared to Māori and Pasifika students respectively.

In all cases, the models show that there was an effect due to ethnicity, which remained after accounting for the quintile effect. That is, there was a difference in average Nature of Social Studies scores between Māori and NZ European and Pasifika and NZ European students over and above the difference accounted for by quintile.

At Year 4, the modelled scale scores show, that on average, Māori students scored 11 scale score units lower than NZ European students and at Year 8, eight scale score units lower (Figure A6.1). At both year levels there appears to be a 'closing of the gap' between NZ European and Māori students towards the higher quintiles, although the Year 8 model did not detect the pattern as statistically significant.

At Year 4, the modelled scale scores show, that on average, Pasifika students scored 21 scale score units lower than NZ European students, and at Year 8, 18 scale score units lower (Figure A6.2). The models however, did not detect any statistically significant changes in this difference across quintile. As noted above, the power of the models to detect statistically significant interactions is compromised by very small numbers in some categories (Table A6.1). This is evidenced by the very wide confidence intervals associated with the group means at each quintile, which can be seen in the figures.

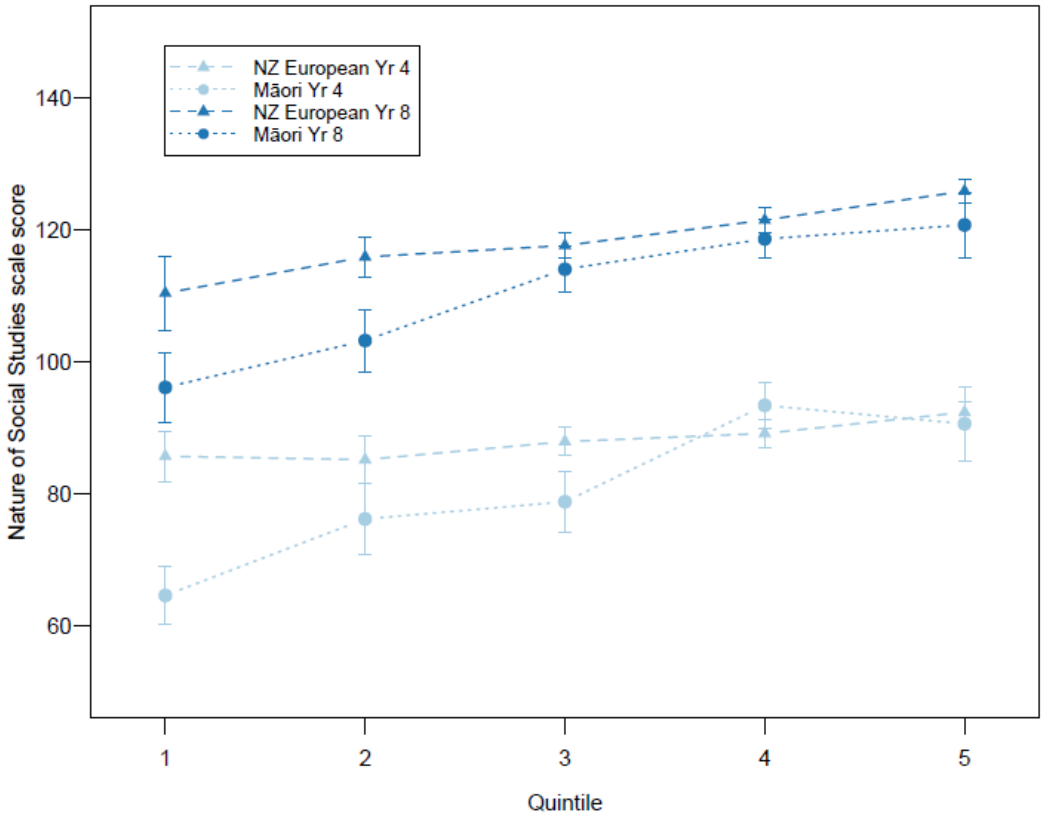

Figure A6.1 Year 4 and Year 8 NZ European and Māori students' Nature of Social Studies scores by quintile

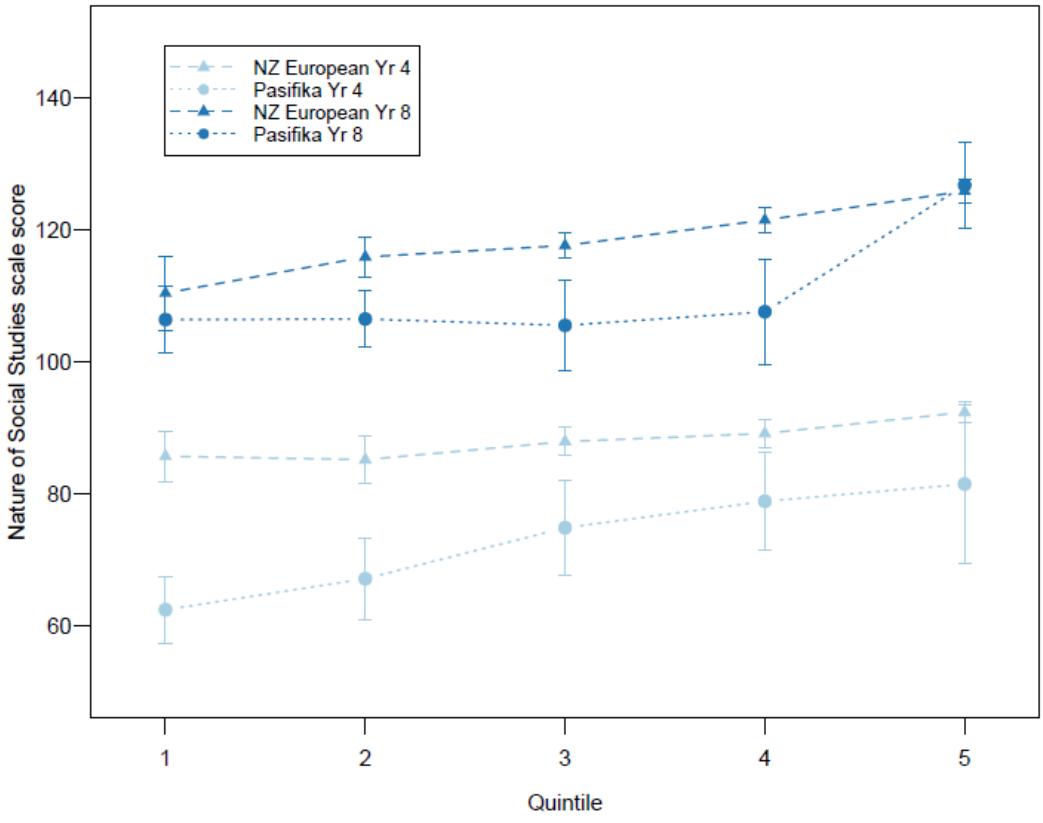

Figure A6.2 Year 4 and Year 8 NZ European and Pasifika students' Nature of Social Studies scores by quintile

7 Appendix 7: The Knowledge and Application of Reading in English assessment framework

1. Introduction

The Knowledge and Application of Reading in English (KARE) is a group-administered assessment designed primarily to monitor students' reading achievement in Year 4 and Year 8. KARE will assess reading as described in the English learning area of the *New Zealand Curriculum* (NZC) in the 2014 NMSSA study. This document locates the KARE assessment in relation to English: reading of the NZC, and provides an overview of the framework that underpins the assessment.

2. Reading in the NZC

English: reading is part of the receptive strand in which students are primarily making meaning of ideas and information they receive. By studying English: reading, students develop the ability to use processes and strategies in order to understand the purposes and audiences, ideas, language features, and structure of written texts. The NZC describes indicators of increasingly complex tasks and texts with which students are expected to engage in progressively greater depth as they develop as readers. An overview of how English: reading is characterised in the NZC and other key New Zealand documents is shown in Table A7.1.

Table A7.1 Summary of reference points in key NZ documents, as applicable to the Knowledge and Application of Reading in English (KARE) assessment

The New Zealand Curriculum	**English in The New Zealand Curriculum**	**The New Zealand Curriculum Reading and Writing Standards for years 1-8**	**The Literacy Learning Progressions**	**National Education Monitoring Project (NEMP)**	**KARE assessment**
The Vision, Principles, and Values: confident; connected; actively involved; lifelong learners, etc. *The Key Competencies:* Thinking; Relating to Others; Using Language, Symbols, and Texts	*To participate in:* • social; • cultural; • political; and • economic life. *of New Zealand and wider world, by having access to:* • understanding; • knowledge; and • skills; *related to*: • text purposes and audiences; • ideas within language contexts; • language features that enhance texts; and • the structure and organisation of texts.	*By the end of Year 4:* Students will read, respond to, and think critically about texts… locate and evaluate information and ideas within texts… *By the end of Year 8:* (as above, reading with increased accuracy and speed, and greater independence) Students need to be confidently and deliberately choosing the most appropriate reading strategies…	*By the end of Year 4:* students use their reading and comprehension strategies to read texts appropriate to this level accurately and fluently … use and integrate a variety of comprehension strategies … to understand, respond to, and think critically …. *By the end of Year 8:* students need to be confidently and deliberately choosing the most appropriate strategies to suit their purposes for reading in different learning areas.	• Recalling and using specific factual information. • Interpreting information. • Making inferences. • Evaluating ideas.	• Locate and recall. • Integrate and interpret. • Critique and evaluate.

What the KARE study will assess

The broad definition of English: reading that underpins the KARE assessment is an individual's ability to understand and use written language forms to learn, to achieve their goals, and to participate in society. More than just decoding text, as reading is sometimes interpreted, comprehension of text has a central role in this assessment.

The KARE assessment will examine reading in relation to three, more specific *cognitive targets* that correspond to the intentions of the *New Zealand Curriculum* (NZC) (Ministry of Education, 2007). The cognitive targets represent different kinds of thinking that underpin reading comprehension (National Assessment Governing Board, 2010) and are students' ability to:

- locate and recall information
- integrate and interpret information
- critique and evaluate information.

These targets have been informed by the cognitive processes for reading assessed by the Progress in International Reading and Literacy Study (PIRLS) (Campbell et al. 2001), the Programme for International Student Achievement (PISA) (OECD, 2000), and the National Assessment of Educational Progress (NAEP) (National Assessment Governing Board, 2010). The targets are consistent with the three aspects of literacy acquisition described in the *Effective Literacy Practice* documents (Ministry of Education, 2003, 2006): *learning the code, making meaning,* and *thinking critically*. The assessment of vocabulary is incorporated across the three cognitive targets.

Students need to know that readers are active and have agency – that it is through their interaction with a text that meaning is made, drawing on their individual and cultural identity to make meaning. For this reason, we have included questions that ask students to interpret and integrate, and critique and evaluate.

Making validity claims about the KARE assessment results

To ensure the KARE assessment of the three cognitive targets allowed us to make valid claims in relation to students' achievement on Reading in English, as described in the NZC, a conceptual assessment framework guided the design and development of the assessment items. The claims and sub-claims shown in Table A7.2 are based on the indicators presented in English in the NZC and informed an assessment blueprint that outlines the relative proportion of items to be developed to represent each target, and the type of questions to be used. A collection of assessment items was developed around appropriate texts. The items included selected response and constructed response questions.

Table A7.2 Draft conceptual assessment framework for the Knowledge and Application of Reading in English assessment

Claims of KARE for NMSSA Through Reading, students: • develop the ability to use processes and strategies; and • understand the purposes and audiences, ideas, language features, and structure of written texts.			
	Sub-claims		**Task/item types and characteristics**
Cognitive targets	**Students will be able to:**	**Students will know:**	
• Use Reading skills to ***locate and recall*** ideas and information across a range of written texts.	• monitor their Reading in order to comprehend; and • match information to either literal or synonymous information.	• how to identify and state facts; • how to identify and state main ideas; • how to identify and state supporting details; and • how to identify and state essential elements of a text (such as characters, time, setting).	• Cloze tasks, in which students draw on context to make meaning and select an appropriate response;. • Multi-choice response items. • Short constructed-response items.
• Use Reading skills to ***integrate and interpret*** ideas and information across a range of written texts.	• process information; and • extend their initial impressions of text.	• when and how to make inferences; • when and how to make comparisons and contrast of information/ideas; • when and how to form generalisations, e.g., author's purpose, implied message of text, audience; • when and how to examine relations across parts of text, or multiple texts; and • when and how to consider alternatives to what is presented in text.	• Cloze tasks, in which students draw on context to make meaning and select an appropriate response. • Multi-choice response items. • Short constructed-response items.
• Use Reading skills to ***critique and evaluate*** ideas and information across a range of written texts.	• consider texts critically; and • evaluate texts.	• when and how to be objective about text; • when and how to assess text from different perspectives; • when and how to synthesise across texts; • when and how to evaluate quality of text, e.g., does it achieve its purpose? • when and how to evaluate effectiveness of textual features; • when and how to evaluate effectiveness of language features; and • when and how to evaluate specific techniques used by authors to convey their intended messages.	• Multi-choice response items. • Short constructed-response items.

The numbers of items included in the group-administered tasks and interviews to assess each of the cognitive targets and vocabulary are shown for Year 4 in Table A7.3 and for Year 8 in Table A7.4.

Table A7.3 Achieved item bank for Year 4 by cognitive target plus vocabulary, and text type

Year 4 Knowledge and Application of Reading in English					
		Cognitive Targets[7]			
Text types	**Vocabulary items**	**Locate and recall**	**Integrate and interpret**	**Critique and evaluate**	**Totals**
Poetry	1 (-)*	4 (1)	7 (1)	(3)	12 (5)
Literary fiction	1 (-)[2]	5 (-)	4 (2)	- (-)	10 (2)
Literary non-fiction	2 (-)	4 (-)	5 (2)	- (1)	11 (3)
Totals	4 (-)	13 (1)	16 (5)	- (4)	33 (10)

* Numbers in parentheses indicate items asked in one-to-one interviews with a sub-sample of Year 4 students.

Table A7.4 Achieved item bank for Year 8 by cognitive target plus vocabulary, and text type

Year 8 Knowledge and Application of Reading in English					
		Cognitive Targets[8]			
Text types	**Vocabulary items**	**Locate and recall**	**Integrate and interpret**	**Critique and evaluate**	**Totals**
Poetry	1 (-)*	2 (-)	8 (1)	- (1)	11 (2)
Literary fiction	2 (-)	4 (-)	7 (1)	- (2)	13 (3)
Literary non-fiction	5 (-)	2 (-)	4 (1)	- (2)	11 (3)
Totals	8 (-)	8 (-)	19 (3)	- (5)	35 (8)

* Numbers in parentheses indicate items asked in one-to-one interviews with a sub-sample of Year 4 students.

References

Ministry of Education. (2003). *Effective Literacy Practice in Years 1 to 4*. Wellington: Learning Media.

Ministry of Education. (2006). *Effective Literacy Practice in Years 5 to 8*. Wellington: Learning Media.

Ministry of Education. (2007). *The New Zealand Curriculum*. Wellington: Ministry of Education.

National Assessment Governing Board. (2010). *Reading framework for the 2011 national assessment of educational progress*. Washington, DC: US Department of Education.

Organisation for Economic Co-operation and Development. (2000). *Measuring student knowledge and skill: The PISA 2000 assessment of reading, mathematical, and scientific literacy*. Paris: Author.

[7] In addition to these items, 37 Cloze items were included in the group-administered assessment for Year 4 students. These items assessed combinations of vocabulary knowledge and cognitve targets.

[8] In addition to these items, 31 Cloze items were included in the group-administered assessment for Year 8 students. These items assessed combinations of vocabulary knowledge and cognitve targets.

8 Appendix 8: Curriculum alignment in NMSSA English: reading 2014

1. Introduction and background

The underlying objective of NMSSA is to report on student achievement with respect to the *New Zealand Curriculum* (NZC). Assessment data in different learning areas is collected each year and achievement scales constructed in order for this to be carried out. In 2014 English: reading was one of the two learning areas under consideration. This appendix describes the process and presents results for the alignment of the NMSSA Knowledge and Application of Reading in English (KARE) scale with the NZC.

Figure A8.1 shows a complete high-level overview of the NMSSA English: reading study process.

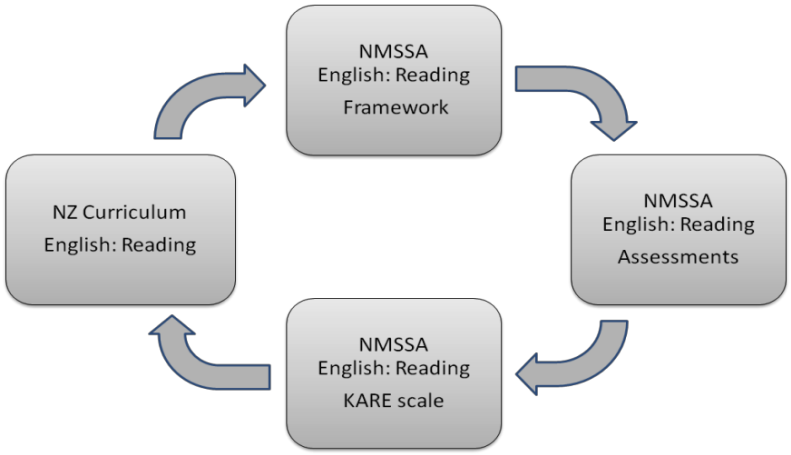

Figure A8.1 Overview of the NMSSA English: reading study

The English: reading framework drew on the NZC document. In particular the NMSSA framework focuses on assessment of the following three cognitive targets:

- locate and recall;
- integrate and interpret; and
- critique and evaluate.

Assessment of the cognitive targets was spread across three text types:

- literary fiction;
- literary non-fiction; and
- poetry.

The framework was then used to inform the construction of the English: reading item bank which consisted of 154 items.

Assessments were delivered mainly in paper and pencil mode, with about 2200 students at each of Year 4 and Year 8 completing an assessment. Each paper and pencil assessment consisted of three or four stimulus texts with questions (mostly multi-choice) attached to each text. Questions were categorised to align with the cognitive targets.

Students also completed two cloze paragraphs where about eight words chosen from a list were inserted into a given text in a way that made the text make sense.

A subset of students answered a small number of additional questions in one-to-one interviews.

A measurement scale, the Knowledge and Application of Reading in English (KARE) scale, was constructed from student responses.

2. Alignment: method

A panel of six external subject matter experts was invited to participate in the alignment. Each panel member brought curriculum expertise, together with classroom and teaching experience in the English: reading learning area. NMSSA researchers and psychometricians also formed part of the alignment team.

The English: reading framework was described in detail by NMSSA researchers. A thorough understanding of the framework, knowledge of its construction, and a good understanding of the KARE achievement scale were all necessary before the panel could make informed judgements about the relationship of curriculum levels to the KARE scale. Some time was spent on each aspect, with questions and discussion encouraged at all times.

It was decided to exclude the cloze exercises from the curriculum alignment activity. It had already been established that the cloze items fitted well on the KARE scale, but the items did not lend themselves so usefully to the alignment method chosen. All other items were considered in the alignment activity, and together represented the complete framework adequately.

Units (stimulus text and all related questions) were presented to the panel one by one so that each could be focussed on without distraction. A bookmarking method was applied to each unit in turn. The method is described in the next section.

Bookmarking method

The stimulus text was read and discussed by panel members. All assessment questions related to the text were ordered by difficulty (easiest to most difficult) and presented in a booklet. Details about scoring the responses were given with each question together with exemplars and annotations where appropriate. A number of questions had been scored with a partial credit scoring rubric (0, 1, 2). These questions appeared twice in the booklets; once where a response would have just scored a '1', and again where the response would have just scored a '2'.

Minimal competence at different curriculum levels

The panel was asked to imagine a group of 100 students all performing at exactly the same curriculum level. Two levels were considered - borderline curriculum Level 2 and borderline curriculum Level 4.

'Minimal competence' at a curriculum level was thoroughly discussed and a common understanding of the concept was reached.

A number of definitions of minimal competence were considered. Panel members were encouraged to use the definition(s) that best helped them visualise a group of students at the appropriate level. The following definitions were discussed for Level 2.

- A minimally competent student **just** (barely) meets the curriculum expectations at Level 2.
- A minimally competent student has **just** enough of the requisite knowledge and skill to perform most of the time according to Level 2 expectations, although their knowledge and skill may be limited.
- A minimally acceptable candidate is **borderline** Level 2.
- A student who is minimally competent at Level 2 has done **just enough** to be described as someone performing in/at Level 2.
- Minimally competent students are deemed to be operating at/in Level 2, but **only just**.
- Over a number of tasks and contexts, on average this student will produce performances that are overall **just good enough** to mean Level 1 is not an appropriate descriptor.

It was agreed that once Level 2 and Level 4 cut-points had been decided satisfactorily Level 3 would be placed half way between them. An attempt was made to establish a Level 5 (top of Level 4) cut-point. However the item bank was not targeted specifically for Level 5, and there was not enough information to find an agreed Level 5 boundary.

Assessment conditions

It was very important for panel members to understand the circumstances under which students completed the NMSSA assessments. Often the demands of the assessment were not completely in line with normal classroom activities. This is a necessary constraint of NMSSA assessments. When students are less familiar with a process, and less supported by teachers and classroom activities, they will tend to perform at a lower level than they would if the supports were in place.

When thinking about question difficulty, and how the visualised group of students would perform on each question, the panel was asked to take the following points into consideration.

- Students had no teacher support for this assessment.
- Students who completed the interview questions.
    - had limited time to think.
    - had to respond verbally 'on the spot'.
    - were asked to respond to someone they did not know.
- There was no classroom or peer discussion to help think about a text.
- Students had no 'scaffolding' in the form of a class module or project.

When considering how difficult a question was to answer for various groups of students, the panel were also asked to think about the following points.

- How a primary school student **thinks** and **processes information**.
- What any particular question is actually asking them to do.
    - How many pieces of information do they need to process?
    - How many thinking steps does it take to answer correctly?
- Are there abstractions or metaphors to cope with in the text or in the question?
- What depth of inference is required?
- Does the question have a 'reading between the lines' aspect?
- Is the context familiar?
- Text density, length, sophistication of vocabulary.
- What misconceptions might trip students up?
- How hard do the distractors make a multi-choice question?
    - Vocabulary and/complexity of distractors.
    - Are the distractors 'close' to the correct answer?
- Question type: Multi-choice or open-ended?

Establishing the cut-points

The panel members were asked to work through the ordered item booklets from the beginning and mark the first page at which they considered that the group of minimally competent students would have less than a 70 percent chance of answering correctly. Figure A8.2 shows this graphically.

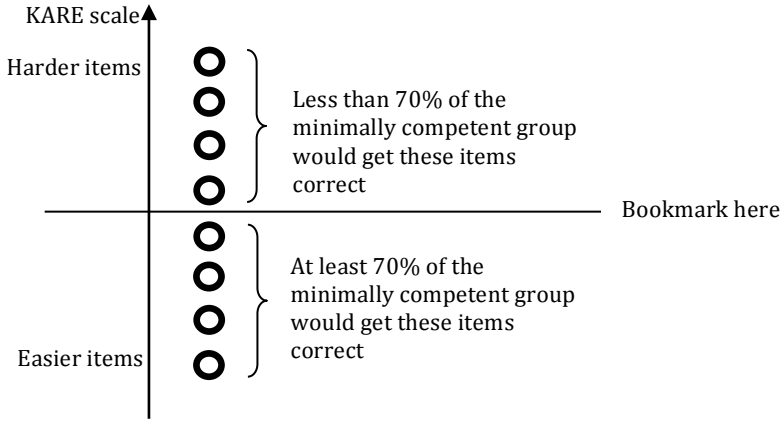

Figure A8.2 Book marking method

Each panel member made individual judgements. This was followed by discussion, and a chance to re-consider the placement of the bookmark. There was no requirement for complete agreement between panel members. However, throughout the day care was taken to challenge judgements that varied widely, or that appeared to be very off-target. Justifying their own thinking to other panel members helped in deciding whether to update original judgements.

3. Results

A cut-point for each curriculum level in each unit for each panel member was calculated as being half way between the difficulty of the question on the bookmarked page and the difficulty of the question on the previous page. The estimated cut-points were averaged to find a scale location for each unit, and then averaged across units to establish the final cut-points for Level 2 and Level 4.

For some texts there was discussion about whether the beginning of Level 2 could be observed at all. In general, agreement was reached about whether or not a Level 2 cut-point could be constructed with some of the more complex texts. As a result three of the six texts examined did not contribute to the Level 2 cut-point.

Table A8.1 shows the final locations on the KARE scale for the beginning of Level 2, Level 3 and Level 4.

Table A8.1 Final curriculum levels aligned with the KARE scale

Beginning of	KARE scale location
Level 2	82.0
Level 3	96.5
Level 4	110.9

9 Appendix 9: Regression analysis: interaction between ethnicity and decile for English: reading

1. Introduction

Reporting on differences between ethnic groups in New Zealand is complicated on two counts. First, a high proportion of Māori and Pasifika students attend lower decile schools, and a much lower proportion attend high decile schools. This creates a skew in the distribution of all ethnic subgroups with respect to decile and means that any score differences between ethnic groups could be explained by general differences between performance levels at different deciles and vice-versa.

The second complication is that students may identify with more than one ethnic group. It is difficult to make useful, robust statistical statements with respect to performance in ethnicity subgroups when there is 'blurring' with regard to group membership.

To explore the performance of ethnic groups on the Knowledge and Application of Reading in English (KARE) measure across deciles the following regression analyses were carried out:

- a comparison of Māori and NZ European students' KARE scores;
- a comparison of Pasifika and NZ European students' KARE scores;

For the purposes of the analysis decile was coded to quintile[9].

For each year level and for both Māori and Pasifika sub-groups, separate models were run to examine the effect on performance outcomes due to quintile and ethnicity.

This paper describes the regression analysis and presents the results.

2. The regression models

For the purposes of the analysis, the school decile band was recoded to quintile[10].

For each year level and for both Māori and Pasifika subgroups, separate models were run to examine the effect on performance outcomes due to quintile and ethnicity. These are described below.

Models comparing Māori and NZ European (NZE) subgroups

1. $Y_i = \alpha + \beta1_i$*quintile+ $error_i$

2. $Y_i = \alpha + \beta1_i$*quintile +$\beta2_i$*Māori + $\beta3_i$*NZE + $error_i$

3. $Y_i = \alpha + \beta1_i$*quintile +$\beta2_i$*Māori + $\beta3_i$*NZE +$\beta4_i$*(Māori*quintile) + $\beta5_i$*(NZE*quintile) + $error_i$

[9] Decile 1-2 → Quintile 1, Decile 3-4 → Quintile 2, … , Decile 9-10 → Quintile 5

[10] Decile 1-2 → Quintile 1, Decile 3-4 → Quintile 2, … , Decile 9-10 → Quintile 5

Models comparing Pasifika and NZ European subgroups

1. $Y_i = \alpha + \beta1_i$*quintile+ $error_i$

2. $Y_i = \alpha + \beta1_i$*quintile +$\beta2_i$*Pasifika + $\beta3_i$*NZE + $error_i$

3. $Y_i = \alpha + \beta1_i$*quintile +$\beta2_i$* Pasifika + $\beta3_i$*NZE +$\beta4_i$*(Pasifika *quintile) + $\beta5_i$*(NZE*quintile) + $error_i$

In the model statements, Y_i is the KARE scale score for student *i*, and quintile, Māori, Pasifika and NZE are all classification ('dummy') variables.

For each model, there was a strong and statistically significant quintile effect. Average scores increased consistently with quintile.

The results from this analysis should be interpreted with caution. The model's ability to precisely assess how Māori or Pasifika students are performing, on average, in higher decile schools (and how NZ European students are performing in lower decile schools) is compromised by the disproportionate numbers of students in those deciles in the national sample with respect to their ethnicity.

Results for Māori/NZ European models:

The R^2 statistic, indicating the proportion of variance in the KARE scores accounted for by each model is presented below.

Table A9.1 R^2 results by model and year level for Māori/NZ European models

Year	Model	R^2	R^2 change*
4	1	.09	-
	2	.11	**.02**
	3	.11	.00
8	1	.08	-
	2	.11	**.03**
	3	.11	.00

R^2 change values in bold indicate statistically significant (p<.05) improvement from previous model.

Models were compared using the usual F-test[11]. At each of Year 4 and Year 8, Model 2 was found to be the most parsimonious in the context of the variables of interest. Models with additional interaction terms (Model 3) showed no significant improvement over the main effects model (Model 2).

Results for Pasifika/NZ European models:

The R^2 statistic, indicating the proportion of variance in the KARE scores accounted for by each model was presented below.

Table A9.2 R^2 results by model and year level for Pasifika/NZ European models

Year	Model	R^2	R^2 change*
4	1	.08	-
	2	.09	**.01**
	3	.09	.00
8	1	.06	-
	2	.09	**.03**
	3	.10	.01

R^2 change values in bold indicate statistically significant (p<.05) improvement from previous model.

The models were compared using the usual F-test. At each of Year 4 and Year 8, Model 2 was found to be the most parsimonious in the context of the variables of interest. Models with additional interaction terms (Model 3) showed no significant improvement over the main effects model (Model 2).

[11] $F = \frac{\left(\frac{Drop\ in\ SSE}{Number\ of\ added\ terms}\right)}{s^2\ for\ the\ full\ model}$, where SSE = Sum of the squared residuals in the respective model.

3. Summary

Figures A9.1 and A9.2 show KARE average scores and their associated confidence intervals by quintile, for NZ European students compared to Māori and Pasifika students respectively.

In all cases, the models showed that there was an effect due to ethnicity which remained after accounting for the quintile effect. That is, there was a difference in average KARE scores between each ethnic subgroup and NZ European students over and above the difference accounted for by quintile. This difference was constant (as far as the model could determine) across all quintiles.

Figures A9.1 and A9.2 show KARE mean scores and confidence intervals by quintile for NZ European students compared to Māori and Pasifika students respectively. Average scores for each ethnic group are shown using dotted lines and symbols.

At both year levels, the modelled scale scores show that on average Māori students scored 8 scale score units lower than NZ European students (Figure A9.1). At Year 4, the modelled scale scores show that on average Pasifika students scored 8 scale score units lower than NZ European students, and at Year 8, 10 scale score units lower (Figure A9.2). The models however did not detect any statistically significant changes in this difference across quintile. As noted above, the power of the models to detect statistically significant interactions is compromised by small numbers in some categories.

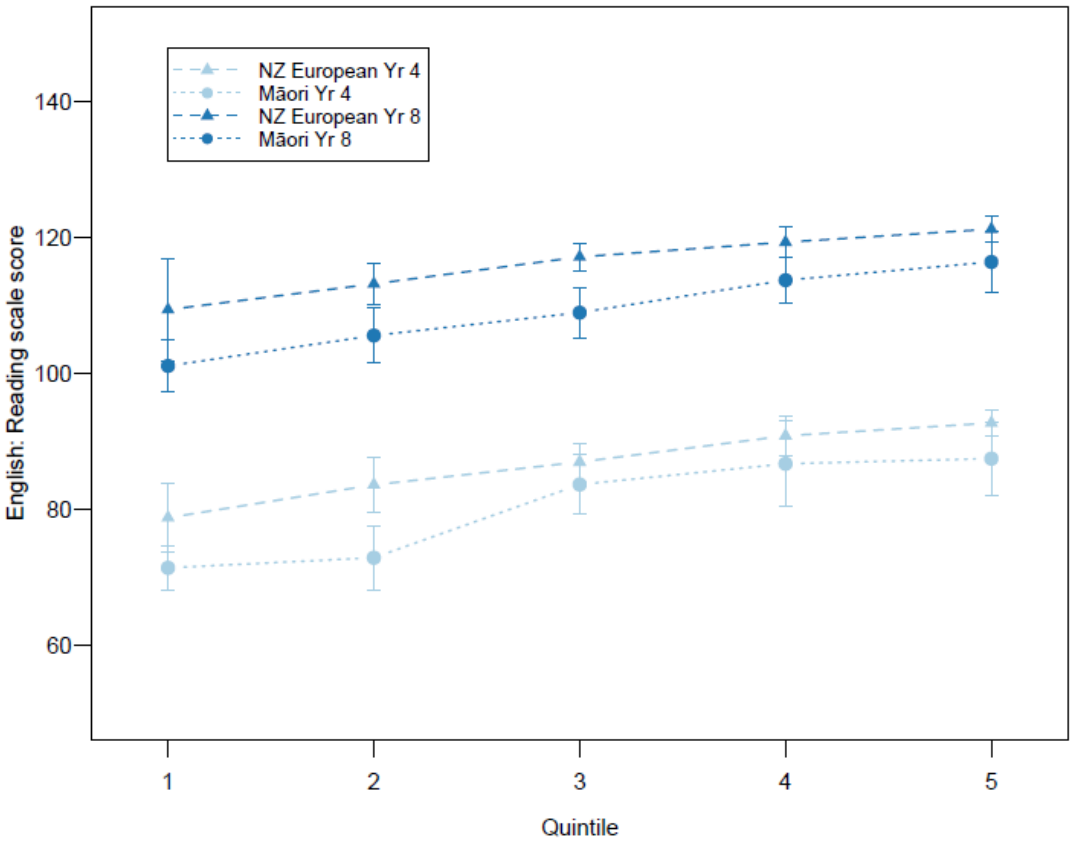

Figure A9.1 Year 4 and Year 8 NZ European and Māori students' KARE scores by quintile

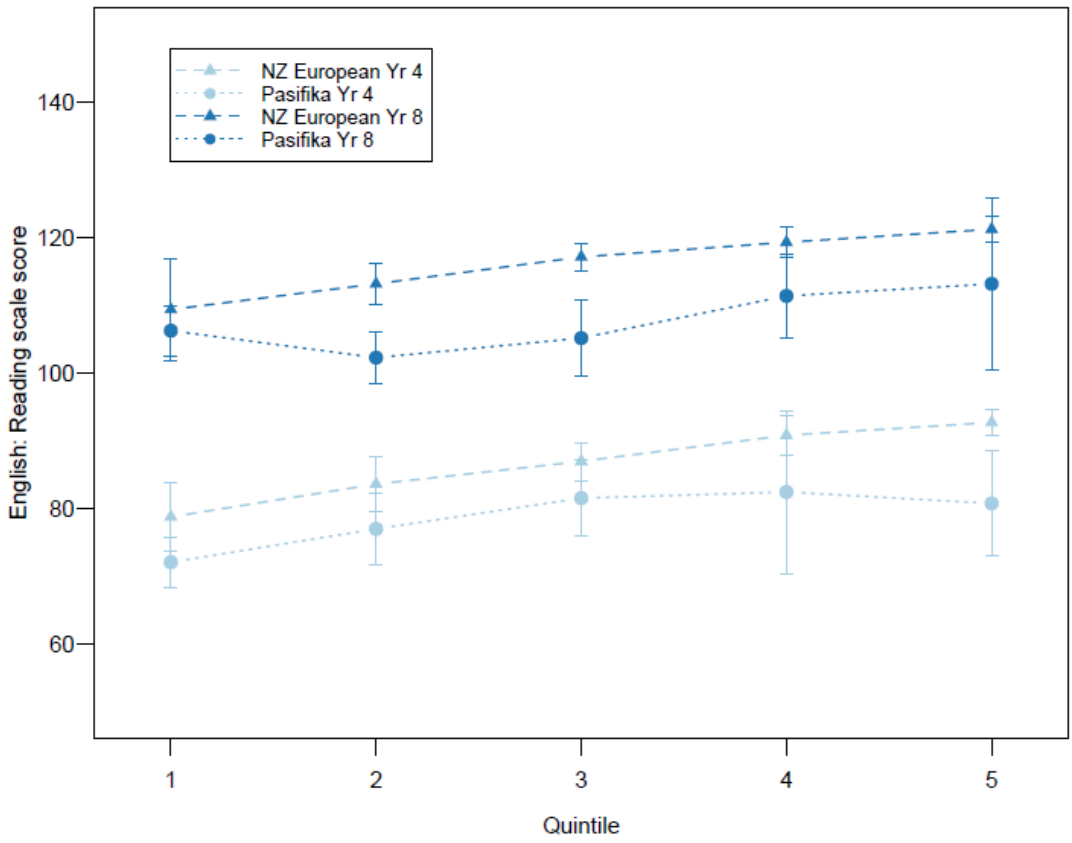

Figure A9.2 Year 4 and Year 8 NZ European and Pasifika students' KARE scores by quintile

ISSN: 2350-3254 (Print) / 2350-3238 (Online)

ISBN: 978-1-927286-20-3 (Online only)