Wānangatia te Putanga Tauira National Monitoring Study of Student Achievement

## Technical Information 2019

－English

Wānangatia te Putanga Tauira
National Monitoring Study of Student Achievement

## Technical Information 2019 <br> English

Educational Assessment Research Unit<br>and<br>New Zealand Council for Educational Research

© 2020 Ministry of Education, New Zealand

U N I V ER S I T Y
Otägo


Te Whare Wānanga o Otāgo
NEW ZEALAND

## ๑๑GNZCER

Rangahau Mātauranga o Aotearoa

Technical Information 2019 English
(all available online at http://nmssa.otago.ac.nz/reports/index.htm)


Wannangatia te putanga tauira
National Monitoring Study Nationai Monitoring Stuay
of Student Achievement

## Contents

Acknowledgements ..... 4
Appendix 1: $\quad$ Sample Characteristics for 2019 ..... 5
Appendix 2: Methodology for the 2019 NMSSA Programme ..... 13
Appendix 3: NMSSA Sample Weights 2019 ..... 18
Appendix 4: Variance Estimation: NMSSA 2019 ..... 20
Appendix 5: Curriculum Alignment of the 2019 NMSSA Speaking in the English Learning Area Assessment ..... 24
Appendix 6: Linking English across Cycle 1 and Cycle 2 ..... 31
Appendix 7: NMSSA Assessment Framework for English 2019 ..... 39


## 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 technical advisory panel
- members of the curriculum advisory panel in the learning area of English
- principals, teachers and students of the schools where the tasks were piloted and trials were conducted
- principals, teachers and Board of Trustees' members of the schools that participated in the 2019 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 and senior initial teacher education students who undertook the marking
- the Ministry of Education Research Team and Steering Committee.


## Appendix 1: <br> Sample Characteristics for 2019

## Contents:

Samples for 2019 ..... 6

1. Sampling of schools ..... 6
Sampling algorithm ..... 6
Substitution procedure ..... 7
2. Sampling of students ..... 7
Achieved samples at Year 4 ..... 8
Achieved samples at Year 8 ..... 10
Tables:
Table A1.1 The selection of Year 4 students for the GAT and InD samples from 98 schools ..... 8
Table A1.2 The composition of the Year 4 samples in comparison with the sample frame by gender, ethnicity, school quintile, school type and education region ..... 9
Table A1.3 The selection of Year 8 students for the GAT and InD samples from 100 schools ..... 10
Table A1.4 The composition of the Year 8 samples in comparison with the sample frame by gender, ethnicity, school quintile, school type and education region ..... 11

## Samples for 2019

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; the second stage involved sampling students within schools.

A stratified random sampling approach was taken to select 100 schools at Year 4 and 100 schools at Year 8. A maximum of 25 students were randomly selected from each school to form national samples at Year 4 and Year 8.

The Ministry of Education July 2018 school returns for Year 3 and Year 7 were used to inform the selection of Year 4 and Year 8 schools in 2019.

## 1. Sampling of schools

## Sampling algorithm

From the complete list of New Zealand schools select two datasets - one for Year 3 students and one for Year 7 students.

For the Year 3 sample:

- Exclude:
- schools which have fewer than eight Year 3 students
- private schools
- special schools
- Correspondence School
- Kura Kaupapa Māori
- trial schools
- Chatham Island schools.
- 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:
$\frac{\text { Total number of Year } 3 \text { students }}{100}$
- Assign each school to a 'selection group' using this calculation:

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

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

Follow the same process for the Year 7 sample.
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.

[^0]
## Substitution procedure

The sampling frames constituted 1,494 schools for Year 3 and 975 schools for Year 7 after exclusions had been applied. No schools selected were listed in both samples.
Selected schools were invited to participate in 2019. 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 sampling frame, select the school one row below the school withdrawn.
- If this school is not available, re-select by going to one row above the school withdrawn.
- 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, 35 schools at Year 4 and 49 schools at Year 8 declined to participate, before a sample of 98 schools was achieved at Year 4 (two schools withdrew two days prior to the visit and could not be replaced) and a sample of 100 schools was achieved at Year 8. Of the 135 Year 4 schools approached, 28 were from the original sample and 7 were replacement schools who also declined. Of the 149 Year 8 schools approached, 30 were from the original sample and 19 were replacement schools who also declined. The participation rate ${ }^{4}$ at Year 4 was 72.6 percent and at Year 8 it was 67.1 percent.

## 2. Sampling of students

Four nested student samples were required for the study:

1. A sample that included up to 25 students per school completed group-administered task (GAT) assessments in writing and reading.
2. A subset of up to 12 students per school formed the sample for the GAT in listening and viewing, and the GAT in viewing: moving assessments.
3. A subset of up to eight students per school formed the sample that participated in the in-depth (InD) assessments in presenting and speaking, and participated in an interview that covered aspects of reading, speaking and presenting.
4. A subset of four students per school formed the sample for the paired-speaking task.

The procedure for selecting students for the samples was as follows:

- Participating schools were asked to provide a list of all students in their school at the relevant year level (Year 4 or Year 8) in 2019, identifying any students who should be excluded for logistical reasons, or because 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, and other health or behavioural issues).
- For each school, a computer-generated random number between 1 and 1 million was assigned to each student and they were then ranked in order of their random number from lowest to highest.
- The first 25 students in the ordered list were identified as belonging to the GAT sample for reading and writing.
- The first 12 students also belonged to the GAT sample in listening and viewing, and viewing: moving.
- The first eight students belonged to the InD samples for assessing speaking and presenting, and interviews. The first four students belonged to the InD sample for the paired-speaking assessment.
- The names of selected students were returned to schools for approval. Principals or contact people were given a second opportunity to identify students for whom the NMSSA assessment would be inappropriate. Any students identified for withdrawal were replaced with students listed 26 onwards from the ordered list. The resultant sample was confirmed and letters of consent were sent to the parents of selected students on our behalf via the schools.
- The children of parents who declined to have their child participate were withdrawn from the sample and were replaced in the same way as above (if there were sufficient eligible students). However, no replacements were added within two weeks of the date of the school visit, as there was insufficient time to seek parental permission.

[^1]- On-site replacements of students by teacher assessors (TAs) were made if any of the students 1-8 (the InD sample) were absent or withdrawn on the first day, prior to the start of assessments. They were replaced by students ranked $9-25$, on a best-match basis (e.g. using the gender/ethnicity replacement priorities).
- If students were absent or withdrawn after the start of the assessment programme, no replacements were made.
The following sections describe the achieved GAT and InD samples of students at Year 4 and Year 8 and contrast their demographic characteristics with those of their respective national populations (through comparison with the sample frame of all students in eligible schools). This allows us to assess the national representativeness of the samples in relation to those characteristics.


## Achieved samples at Year 4

Table A1.1 summarises the achieved samples of students participating in the eight different assessments.
Across the 98 schools participating at Year 4, principals identified 293 students for whom the experience would be unsuitable; a further 158 students were excluded from the school sample after it had been selected and 135 students were substituted.

The initial sample (the first 25 students in each school's list) consisted of 2,302 randomly selected students. Principals or parents withdrew 347 students. Substitute students numbered 263. Another 291 students were withdrawn without sufficient time for replacement, were absent or did not respond for other reasons during the assessment period. The achieved GAT reading and writing sample included 1,950 students, which represents a participation rate of 61 percent ${ }^{5}$. The achieved sample for each assessment is displayed in the bottom row of Table A1.1.

Table A1.1 The selection of Year 4 students for the GAT and InD samples from 98 schools

|  | GAT tasks | InD tasks |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Mode of English <br> Maximum students per school | Reading / Writing <br> Initial sample: | Listening / Viewing <br> 25 | Speaking / Presenting <br> 8 | Speaking (pairs) |
| Students withdrawn by parents <br> or principals after sampling | 2302 |  |  | 4 |
| Substitute students used <br> (replacements for above) | -347 |  |  |  |
| Absences, non-responses and <br> withdrawals during assessment <br> period | -291 |  |  |  |
| Achieved sample: | 1950 | -17 | -10 |  |

[^2]Table A1.2 contrasts the characteristics of the samples with the sample frame across a number of key demographic variables.

Table A1.2 The composition of the Year 4 samples in comparison with the sample frame by gender, ethnicity, school quintile, school type and education region

|  | Sample frame$\begin{gathered} N=61,674 \\ \% \end{gathered}$ | GAT samples |  | In-depth samples |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Reading / Writing $N=1950$ | Listening / Viewing $N=1146$ | Speaking / Presenting $N=779$ | Speaking (pairs) $N=387$ |
| Gender |  |  |  |  |  |
| Boys | 51.2 | 48.1 | 46.2 | 47.1 | 48.8 |
| Girls | 48.8 | 51.9 | 53.8 | 52.9 | 51.2 |
| Ethnicity* |  |  |  |  |  |
| European | 50.9 | 47.9 | 48.5 | 49.0 | 49.9 |
| Māori | 20.9 | 18.7 | 19.8 | 21.2 | 20.8 |
| Pacific | 11.6 | 12.2 | 12.7 | 11.5 | 10.7 |
| Asian | 13.2 | 15.4 | 14.0 | 13.3 | 13.1 |
| Other | 3.5 | 5.8 | 5.1 | 5.1 | 5.5 |

## Quintile

| 1 | 16.8 | 15.5 | 16.8 | 17.6 | 17.3 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| 2 | 16.9 | 14.9 | 15.4 | 15.1 | 15.5 |
| 3 | 16.4 | 14.5 | 14.6 | 14.2 | 14.5 |
| 4 | 21.6 | 26.2 | 25.6 | 25.7 | 25.8 |
| 5 | $\mathbf{2 8 . 3}$ | 29.0 | 27.7 | 27.3 | 26.9 |

## School type

| Contributing | $\mathbf{6 0 . 9}$ | 64.2 | 62.1 | 61.2 | 61.8 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Full primary | $\mathbf{3 6 . 4}$ | 33.5 | 35.9 | 37.0 | 36.2 |
| Composite (Year 1-10) | $\mathbf{2 . 7}$ | 2.4 | 2.0 | 1.8 | 2.1 |

## Region

| Auckland | 35.3 | 35.5 | 34.0 | 33.9 | 33.1 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Bay of Plenty/Waiariki | 7.8 | 7.2 | 7.4 | 7.6 | 8.3 |
| Canterbury | 11.7 | 10.7 | 11.5 | 11.2 | 11.4 |
| Hawkes Bay/Tairāwhiti | 4.9 | 6.0 | 6.1 | 6.2 | 6.2 |
| Nelson/Marlborough/ West Coast | 3.5 | 3.4 | 3.1 | 3.2 | 3.1 |
| Otago/Southland | 6.2 | 6.9 | 7.2 | 7.3 | 7.2 |
| Northland/Tai Tokerau | 3.9 | 3.9 | 4.2 | 4.0 | 4.1 |
| Taranaki/Whanganui/ Manawatu | 6.5 | 6.2 | 6.2 | 6.3 | 6.2 |
| Waikato | 9.0 | 8.3 | 8.9 | 9.2 | 9.3 |
| Wellington | 11.3 | 11.8 | 11.3 | 11.2 | 11.1 |

Note: Ministry of Education July 2019 school returns for Year 4 were used for the population percentages.

* Ethnicity is based on the Ministry of Education's prioritised ethnicity.


## Achieved samples at Year 8

Across the 100 schools participating at Year 8, principals identified 317 students for whom the experience would be unsuitable; a further 189 students were excluded from the school sample after it had been selected and 128 students were substituted.
The initial sample (the first 25 students in each school's list) consisted of 2,301 randomly selected students. Principals or parents withdrew or excluded 292 students. Substitute students numbered 227. A further 300 students were withdrawn without sufficient time for replacement, were absent or did not respond for other reasons during the assessment period. The achieved GAT sample for reading and writing included 1,957 students, which represents a participation rate of 63 percent. The achieved sample for each assessment is displayed in the bottom row of Table A1.3.

Table A1.3 The selection of Year 8 students for the GAT and InD samples from 100 schools

|  | GAT tasks |  | InD tasks |  |
| :--- | :---: | :---: | :---: | :---: |
| Mode of English <br> Maximum students per school | Reading / Writing <br> 25 | Listening / Viewing <br> Initial sample: | Speaking / Presenting <br> 8 | Speaking (pairs) <br> 4 |
| Students withdrawn by parents <br> or principals after sampling | $\mathbf{2 3 0 1}$ |  |  |  |
| Substitute students used <br> (replacements for above) | -292 |  |  |  |
|  <br> withdrawals during assessment <br> period | 227 |  |  |  |
| Achieved sample: | -300 | -30 | -15 | -9 |

Table A1.4 contrasts the characteristics of the Year 8 samples with the sample frame across a number of key demographic variables.

Table A1.4 The composition of the Year 8 samples in comparison with the sample frame by gender, ethnicity, school quintile, school type and education region


## School type

| Intermediate | $\mathbf{3 0 . 9}$ | 43.0 | 45.4 | 45.9 | 45.5 |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Full primary | $\mathbf{4 8 . 1}$ | 40.8 | 38.7 | 38.0 | 38.4 |
| Sec (Year 7-10 \& 7-15) | $\mathbf{1 6 . 1}$ | 11.1 | 10.9 | 11.2 | 11.0 |
| Composite (Year 1-10) | $\mathbf{5 . 0}$ | 5.1 | 4.9 | 5.0 | 5.1 |

## Region

| Auckland | 33.7 | 34.8 | 33.7 | 33.1 | 33.8 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Bay of Plenty/Waiariki | 8.1 | 8.4 | 8.8 | 9.0 | 9.0 |
| Canterbury | 12.1 | 13.4 | 12.7 | 13.0 | 13.3 |
| Hawkes Bay/Tairāwhiti | 5.0 | 5.2 | 5.9 | 5.9 | 6.1 |
| Nelson/Marlborough/ <br> West Coast | 3.5 | 3.2 | 3.1 | 3.1 | 2.6 |
| Otago/Southland | 6.6 | 7.2 | 7.3 | 7.1 | 7.2 |
| Northland/Tai Tokerau | 4.1 | 2.9 | 3.6 | 3.9 | 3.1 |
| Taranaki/Whanganui/ Manawatu | 6.7 | 5.0 | 6.2 | 6.8 | 7.2 |
| Waikato | 8.7 | 9.1 | 8.2 | 7.9 | 8.2 |
| Wellington | 11.5 | 10.8 | 10.4 | 10.2 | 9.7 |

[^3]At both year levels there was a reasonably close match between the student samples and their respective population (as represented by the sample frame) in relation to ethnicity, quintile and region. A slightly greater proportion of girls and a corresponding smaller proportion of boys was evident for both samples. Some variability was evident for school type. For some of the assessments at Year 4, contributing schools were slightly under-represented and full primary schools were over-represented. At Year 8, intermediate and secondary schools (Year 7-10 \& 7-15) were slightly under-represented and full primary schools were over-represented.
Overall, we have confidence the samples are nationally representative.

## Appendix 2: <br> Methodology for the 2019 NMSSA Programme

## Contents:

1. The 2019 English assessment programme ..... 14
Development and trialling of English tasks ..... 15
Administration of the assessment tasks in 2019 ..... 15
2. Marking ..... 15
3. Creating the achievement scales for English ..... 15
Standardising the scales ..... 16
Scale descriptions ..... 16
4. Reporting achievement against curriculum levels ..... 16
5. Development of questionnaires for examining contextual information ..... 16
Student questionnaire ..... 16
Teacher questionnaire ..... 17
Principal questionnaire ..... 17
Measurement scales for the questionnaires ..... 17
6. Administration of the questionnaires ..... 17

This appendix outlines the methodology for the 2019 English study undertaken by the National Monitoring Study of Student Achievement (NMSSA).

## 1. The 2019 English assessment programme

The 2019 English assessment programme built upon the NMSSA assessment framework for English (see Appendix 7, page 43). We used three assessment approaches to assess the different modes of English. The first approach involved group-administered tasks (GATs) delivered to about 2,000 students at Year 4 and 2,000 students at Year 8 for reading and writing; and to about 1,200 students at Year 4 and 1,200 students at Year 8 for listening and viewing. The second approach included a set of in-depth group tasks, undertaken by a subset of 800 students at each year level, in speaking. The third approach consisted of one-to-one interview tasks conducted with 800 students at each year level. During the interview, a presenting and a speaking task were administered, along with additional in-depth questions related to reading. The in-depth components of assessment, including interviews, provided an opportunity to explore students' knowledge and understanding of presenting, reading and speaking in English without the limitations inherent in requiring students to communicate responses in written form.
Table A2.1 summarises the key differences between the assessment programmes for the English learning area in Cycle $1^{6}$ and Cycle 2. See Appendix 7 for the 2019 assessment framework.

Table A2.1 The key features of the Cycle 1 and Cycle 2 English assessment programmes

|  | $\begin{gathered} \text { Cycle } 1 \\ (2012,2014 \text { and 2015) } \end{gathered}$ | Cycle 2 <br> (2019) |
| :---: | :---: | :---: |
| Reading | In 2014, English reading was assessed in two parts. Part 1 used text passages with accompanying selected-response questions, and cloze ${ }^{7}$ passages. The GATs were answered by all Year 4 and Year 8 students (approximately 2,000 at each year level). <br> Part 2 involved in-depth interview questions about reading. These were answered by 800 students at each of Year 4 and Year 8. | In 2019, English reading was assessed in two parts. Part 1 were GATs with text passages and items used in 2014 plus some new items. Cloze passages were replaced with text passages. The reading GATs were answered by all Year 4 and Year 8 students (approximately 2,000 at each year level). <br> Part 2 involved the same in-depth interview questions as for 2014. These were answered by about 800 students at each of Year 4 and Year 8. |
| Listening | In 2015, English listening was assessed with GATs that were answered by all Year 4 and Year 8 students. | In 2019, the same English listening GATs as 2015 were used, and answered by about 1,200 students at Year 4 and 1,200 students at Year 8. |
| Viewing | In 2015, English viewing was assessed using GATs of static images, which were answered by all Year 4 and Year 8 students. | In 2019, English viewing was assessed with the same paper-and-pencil static image GATs used in 2015 (with updated scoring rubrics). Moving images tasks were added. The moving images (short video clips) were presented on laptops. The English viewing GATs were answered by about 1,200 students at Year 4 and 1,200 students at Year 8. |
| Writing | In 2012, English writing was assessed with five prompts modelled on e-asTTle writing. | In 2019, English writing used the same prompts as 2012. All Year 4 and Year 8 students wrote on one prompt. |
| Speaking | Speaking was not assessed in Cycle 1. | English speaking was assessed with four speaking tasks. Approximately 800 students from each of Year 4 and Year 8 participated in two speaking tasks that involved speaking in a group situation and one task in a one-toone interview. One task presented to a group involved a 'conversation' between two characters. This was undertaken by approximately 400 students at each year level. All tasks were video recorded. |
| Presenting | Presenting was not assessed in Cycle 1. | In 2019, English presenting was assessed with a single task that asked students to design a visual text to convey a message to children in their school. This was undertaken by about 1,200 students at each of Year 4 and Year 8 . Approximately 800 students at each year level were interviewed by a teacher about features of their poster. |

NB *A task is an assessment context. Each task has several questions.

[^4]
## Development and trialling of English tasks

The NMSSA team reviewed all previously used English reading, listening and viewing tasks for possible inclusion in the 2019 assessment programme. Some tasks were retained in their original format to be used as link tasks, necessary for making comparisons between Cycle 1 and Cycle 2.

New and modified tasks were piloted in local schools before being used in a NMSSA trial in March 2019 involving schools in Dunedin. The student responses from the pilots and the trial were used to refine the tasks and support the development of appropriate marking rubrics. An Item Response Theory (IRT) model ${ }^{8}$ was applied to the trial data to help refine the tasks, inform the selection of tasks for the main study and explore the development of the reporting scales.

## Administration of the assessment tasks in 2019

The 2019 study was carried out in Term 3 of 2019. Twelve teacher assessors were trained in the administration of tasks during a five-day training programme prior to the main study. During the study, the teacher assessors were carefully monitored and received feedback to ensure consistency of administration. Student responses were captured on video and paper and stored electronically for marking (responses on paper were scanned).

## 2. Marking

Marking for reading, listening, viewing, speaking and presenting occurred immediately after the administration stage had concluded. Teacher markers, some of whom had been teacher assessors, and third-year University of Otago College of Education students were employed to mark the tasks. All markers were trained, and quality assurance procedures were used to ensure consistency of marking. The marking schedules were refined as necessary to ensure they reflected the range of responses found in the main study. Students' scores were entered directly by the markers into the electronic database.

The writing scripts were scored by trained markers in a marking exercise that occurred in Wellington in November. The exercise was carefully structured so that double marking of student scripts could be used to link the different markers across the range of writing prompts.

## 3. Creating the achievement scales for English

The Rasch IRT model was applied to student responses from the study to construct scales associated with achievement in each of the language modes. This approach included analysing the items used in the assessments for any differential item functioning (DIF) with respect to year level, gender and ethnicity. Items that showed DIF were examined by the task developers, and if their inclusion could not be defended, they were not included in the scale. In the case of DIF related to year level, the affected items were sometimes split into separate year 4 and Year 8 items. Very few items showed DIF. In writing, a multifaceted Rasch model was used that adjusted for differences between markers.

The IRT approach allowed sets of plausible values to be generated for each student involved in the study related to achievement on each of the scales. An exception to this was the scale for writing. The software used to apply the multifaceted model in writing did not allow plausible values to be generated.

Plausible values account for the imprecision associated with scores in an assessment, which can produce biased estimates of how much achievement varies across a population. Each set of plausible values represents a random sample of the possible scores a student might reasonably be expected to attain given their responses to the assessment items. Plausible values provide more accurate estimates of population and subgroup statistics, especially when the number of items answered by each student is relatively small.

The six scales developed for the 2019 English learning area were:

- Writing in the English Learning Area (WELA)
- Reading in the English Learning Area (RELA)
- Listening in the English Learning Area (LELA)
- Viewing in the English Learning Area (VELA)
- Speaking in the English Learning Area (SELA)
- Presenting Task Scale (only one task was included in this assessment).

[^5]The WELA, RELA and LELA scales represented a continuation of scales developed in Cycle 1. The RELA scale involved a reconstruction of the original 2014 reading scale based on plausible values which had not been used for reading in Cycle 1. The remaining scales (VELA, SELA and the Presenting Task Scale) were all new for 2019. Viewing was assessed in Cycle 1, and some tasks were retained and used in Cycle 2. However, the 2019 viewing assessment made use of updated rubrics and included new items based on moving images. These differences meant that a new scale was constructed for viewing based on the 2019 data and that linking the 2019 scale to the 2015 viewing scale was not appropriate. Further information about the process used to link the Cycle 1 and Cycle 2 scales can be found in Appendix 5 (page 27).

## Standardising the scales

When NMSSA scales are constructed they are standardised so that:

- the mean of Year 4 and Year 8 students combined is equal to 100 scale score units
- the average standard deviation for the two year-levels is equal to 20 scale score units.

Achievement on the scales generally ranges from about 20 to 180 units.

## Scale descriptions

The scales for each English language mode were described to indicate the range of knowledge and skills assessed.
To create the scale descriptions, the scoring categories for each item (e.g. $0,1,2$ or 3 ) were located on the respective scales. This meant identifying where the students who scored in each category were most likely to have achieved overall on the scale. Once this had been done for all items, the NMSSA team identified the competencies exhibited as the scale locations associated with the different scoring categories increased, and students' responses became more sophisticated. The result was a multi-part description for each scale, providing a broad indication of what students typically know and can do when achieving at different places on the scale.

The descriptions were provided to give readers of NMSSA reports a strong sense of how the English learning area was assessed. The scale descriptors were not written to necessarily 'line up' with curriculum levels or achievement objectives. They were a direct reflection of what was assessed and how relatively hard or easy students found the content of the assessments.

## 4. Reporting achievement against curriculum levels

The curriculum alignment exercises carried out in Cycle 1 for writing (2012), reading (2014) and listening (2015) allowed the linked results in 2019 to also be reported against curriculum levels (see Appendix 6, page 35). An alignment exercise had also been carried out for viewing in Cycle 1 (2015). Because scale linking was not possible for viewing, a decision was made to set the curriculum level cut scores on the 2019 viewing scale so that the proportions of students achieving curriculum expectations at Year 4 in Year 8 in viewing were the same for both Cycle 1 and Cycle 2.

A curriculum alignment exercise for speaking in the English learning area, which was not assessed in Cycle 1, was carried out in 2020 (see Appendix 5, page 27). This allowed achievement in English speaking to be reported against curriculum levels 2, 3 and 4.

The Presenting Task Scale was not used to report achievement against curriculum levels. This scale was based on a single task and could not support a robust interpretation in terms of achievement against the curriculum.

## 5. Development of questionnaires for examining contextual information

In order to gain a better understanding of student achievement in New Zealand, NMSSA collects contextual information through questionnaires to students, teachers and principals.

## Student questionnaire

The student questionnaire included sections about: their background (e.g. amount of English spoken at home) and reading activities at home; their attitude to (e.g. I like listening to stories and poems) and confidence in learning (e.g. I am good at understanding the ideas in the things I watch-things like movies, presentations, or posters) in each mode; and how often a number of activities happen at school (e.g. in class, we view things like movies, presentations, or posters).
Two IRT scales were constructed from the student questionnaire data:

- Confidence in Reading
- Confidence in Writing.


## Teacher questionnaire

The teacher questionnaire included sections about: their background and experience in teaching (e.g. gender, ethnicity, years of experience teaching, leadership); their confidence to teach each English mode; students' opportunities to experience a number of activities in each English mode at school; and professional development and support, and resources available for teaching each English mode.
A scale for Confidence in Teaching English was constructed.

## Principal questionnaire

The principal questionnaire consisted of four sections: percentage of students in their school with English as their second or alternative language; learning and teaching in the English learning area; professional support for English teaching; and purposes for teaching the English learning area.

## Measurement scales for the questionnaires

The scales associated with the questionnaires were constructed using the Rasch model. Unlike the achievement measures, plausible values were not generated for the contextual scales. Each contextual scale was standardised in the same way as the achievement scales.
To aid interpretation of the contextual scales, they were divided into separate score ranges to provide different reporting categories. For instance, the Confidence in Reading scale was broken down into three score ranges. The 'very confident' part of the scale was associated with students mainly using the 'totally agree' category to respond to each of the questionnaire statements related to attitude, the 'confident' section of the scale was associated with students mainly using either 'agree a lot' or 'agree a little', and the 'not confident' part of the scale was associated with students mainly using 'do not agree at all'.

## 6. Administration of the questionnaires

The student questionnaire was administered to the sample of students who participated in the Viewing GAT (about 1,200 students at Year 4 and 1,200 at Year 8). Up to three teachers from each school were invited to complete the teacher questionnaire. These were the classroom teachers in each school with the most students selected for the study. The principal or a designated school leader (if the principal was unavailable) from each school completed the principal questionnaire.

## Appendix 3:

NMSSA Sample Weights 2019

The methodology for calculating sample weights on an annual basis is detailed in NMSSA Approach to Sample Weighting, available online at https://nmssa.otago.ac.nz/reports/Sample_Weighting_NMSSA.pdf.

Each year we set out a brief summary of the effect of applying sample weights in the analysis of the current year's data, and make a recommendation as to whether weights should be used.

Tables of estimated ${ }^{9}$ means and standard errors calculated with and without sample weights follow. In 2019, NMSSA measured achievement in English. Information about the sample can be found in Appendix 1 (page 5).

Tables 1 and 2 report the NMSSA estimated means and standard errors (in scale score units) for the Year 4 and Year 8 reading samples, respectively.

## Summary

Most weighted estimates were well within one standard error of the estimated unweighted mean, and those that were not were close to one standard error of the estimated unweighted mean.
The recommendation was to proceed with the 2019 analyses without using sample weights.

Table A3.1 NMSSA reading achievement Year 4: Comparison of estimates using unweighted and weighted data

| Year 4 |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Using unweighted data |  | Using weighted data |  | Difference | N |
|  | Mean | SE | Mean | SE |  |  |
| All | 82.0 | 0.6 | 81.1 | 0.6 | 0.9 | 1950 |
| Girls | 85.2 | 0.7 | 84.5 | 0.7 | 0.7 | 1012 |
| Boys | 78.6 | 0.8 | 77.8 | 0.8 | 0.8 | 938 |
| NZE* | 85.1 | 0.7 | 84.6 | 0.7 | 0.5 | 1099 |
| NZE girls | 88.3 | 1.0 | 88.0 | 1.0 | 0.3 | 579 |
| NZE boys | 81.5 | 1.1 | 81.2 | 1.1 | 0.3 | 520 |
| Māori | 75.3 | 1.3 | 74.6 | 1.3 | 0.7 | 431 |
| Māori girls | 78.5 | 1.5 | 78.1 | 1.5 | 0.4 | 227 |
| Māori boys | 71.9 | 2.0 | 71.3 | 2.0 | 0.6 | 204 |
| Pacific | 71.9 | 1.4 | 71.2 | 1.4 | 0.7 | 257 |
| Pacific girls | 75.6 | 1.7 | 75.1 | 1.7 | 0.5 | 146 |
| Pacific boys | 67.1 | 2.4 | 66.6 | 2.4 | 0.5 | 111 |
| Asian | 86.8 | 1.2 | 86.0 | 1.2 | 0.8 | 348 |
| Asian girls | 89.3 | 1.6 | 88.8 | 1.6 | 0.5 | 170 |
| Asian boys | 84.5 | 1.6 | 83.6 | 1.6 | 0.9 | 178 |
| Quintile 1 | 69.7 | 1.4 | 69.3 | 1.4 | 0.4 | 298 |
| Quintile 2 | 75.2 | 1.4 | 74.8 | 1.4 | 0.4 | 291 |
| Quintile 3 | 81.0 | 1.4 | 80.5 | 1.4 | 0.5 | 285 |
| Quintile 4 | 85.6 | 1.0 | 85.4 | 1.0 | 0.2 | 511 |
| Quintile 5 | 89.4 | 1.0 | 89.2 | 1.0 | 0.2 | 565 |

* New Zealand European/Pākehā

[^6]Table A3.2 NMSSA reading achievement Year 8: Comparison of estimates using unweighted and weighted data

| Year 8 |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Using unweighted data |  | Using weighted data |  | Difference | N |
|  | Mean | SE | Mean | SE |  |  |
| All | 117.9 | 0.5 | 117.3 | 0.5 | 0.6 | 1952 |
| Girls | 120.6 | 0.7 | 120.1 | 0.7 | 0.5 | 1001 |
| Boys | 115.2 | 0.7 | 114.6 | 0.7 | 0.6 | 951 |
| NZE* | 122.3 | 0.6 | 122.0 | 0.6 | 0.3 | 1204 |
| NZE girls | 124.8 | 0.9 | 124.7 | 0.9 | 0.1 | 623 |
| NZE boys | 119.5 | 0.9 | 119.3 | 0.9 | 0.2 | 581 |
| Māori | 110.0 | 1.0 | 109.5 | 1.0 | 0.5 | 423 |
| Māori girls | 113.1 | 1.4 | 112.9 | 1.4 | 0.2 | 222 |
| Māori boys | 106.6 | 1.5 | 106.2 | 1.5 | 0.4 | 201 |
| Pacific | 106.6 | 1.3 | 105.9 | 1.3 | 0.7 | 264 |
| Pacific girls | 110.8 | 1.8 | 110.5 | 1.8 | 0.3 | 138 |
| Pacific boys | 101.9 | 1.9 | 101.4 | 1.9 | 0.5 | 126 |
| Asian | 121.3 | 1.4 | 121.0 | 1.4 | 0.3 | 231 |
| Asian girls | 123.9 | 2.0 | 123.7 | 2.0 | 0.2 | 109 |
| Asian boys | 119.0 | 2.0 | 118.8 | 2.0 | 0.2 | 122 |
| Quintile 1 | 105.7 | 1.2 | 105.4 | 1.2 | 0.3 | 272 |
| Quintile 2 | 114.3 | 1.2 | 113.8 | 1.2 | 0.5 | 327 |
| Quintile 3 | 116.2 | 1.2 | 115.6 | 1.2 | 0.6 | 393 |
| Quintile 4 | 122.6 | 1.0 | 122.4 | 1.0 | 0.2 | 479 |
| Quintile 5 | 124.2 | 1.0 | 123.9 | 1.0 | 0.3 | 481 |

[^7]
## Appendix 4: <br> Variance Estimation: NMSSA 2019

## Contents:

1. Summary21
2. Tables of design effects 22

Tables:
Table A4.1 Reading Year 4: Comparison of results for different variance estimation methods 22
Table A4.2 Reading Year 8: Comparison of results for different variance estimation methods

## 1. Summary

This brief summary supports the general NMSSA variance estimation paper ${ }^{10}$ with specific findings relating to data in NMSSA 2019.

Design effects were calculated using the data collected for the NMSSA 2019 English reading assessment. The NMSSA English reading assessment was completed by the majority of the NMSSA sample, and therefore provides the most complete information regarding the clustering of students in schools, and consequently the effect on variance estimation.

Design effects for the whole sample and key sub-groups were investigated.
In general, through experience with calculating design effects each year, it has been noted that reducing the sample size by a factor of 0.7 for calculation of population statistics, accounts for most of the design effect related to the clustered nature of the NMSSA sample.

Design effects in 2019 mostly varied between about 1 and 2. While the design effects in some cases are fairly large, the effect on the width of confidence intervals is small in practice. In all but one case, the increase in width of the 95 percent confidence intervals is less than 1 NMSSA scale score point.

It was recommended that, for ease of calculation and to absorb most of the variance bias caused by the NMSSA complex sample design, the standard multiplier of $\mathbf{0 . 7}$ should be used to form an effective sample size in the calculation of statistics dependent on sample size.

Tables showing the effect of the NMSSA complex sample design on the 2019 English reading assessment follow.

[^8]2. Tables of design effects

| Year 4 | Mean ${ }^{11}$ $\text { (SRS }{ }^{12} \text { ) }$ | $\begin{gathered} \mathbf{S E} \\ (\mathrm{SRS}) \end{gathered}$ | $\begin{gathered} \mathbf{S E} \\ \left(\mathrm{TSL}^{13}\right) \end{gathered}$ | $\begin{gathered} \text { CI } \\ \text { (SRS) } \\ \text { (lower) } \\ \hline \end{gathered}$ | $\begin{gathered} \text { CI } \\ \text { (SRS) } \\ \text { (upper) } \\ \hline \end{gathered}$ | $\begin{gathered} \mathbf{C l} \\ \text { (TSL) } \\ \text { (lower) } \\ \hline \end{gathered}$ | $\begin{gathered} \mathrm{Cl} \\ \text { (TSL) } \\ \text { (upper) } \\ \hline \end{gathered}$ | Design effect | Cl width increase | Cl width increase | N | Effective N |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| All Year 4 | -0.87 | 0.02 | 0.03 | -0.92 | -0.82 | -0.93 | -0.81 | 1.53 | 0.0118 | 24 | 1950 | 1274 |
| NZE* | -0.67 | 0.04 | 0.04 | -0.74 | -0.60 | -0.75 | -0.59 | 1.34 | 0.0111 | 16 | 923 | 689 |
| Māori | -1.22 | 0.05 | 0.07 | -1.32 | -1.11 | -1.35 | -1.09 | 1.43 | 0.0214 | 20 | 431 | 302 |
| Pacific | -1.49 | 0.08 | 0.10 | -1.64 | -1.34 | -1.68 | -1.30 | 1.62 | 0.0415 | 27 | 155 | 97 |
| Asian | -0.59 | 0.05 | 0.06 | -0.70 | -0.49 | -0.72 | -0.47 | 1.37 | 0.0183 | 17 | 316 | 232 |
| Female | -0.70 | 0.03 | 0.04 | -0.76 | -0.63 | -0.77 | -0.62 | 1.34 | 0.0101 | 16 | 1006 | 755 |
| Male | -1.04 | 0.04 | 0.05 | -1.11 | -0.97 | -1.13 | -0.95 | 1.60 | 0.0194 | 26 | 932 | 585 |
| Female NZE | -0.52 | 0.05 | 0.05 | -0.61 | -0.43 | -0.62 | -0.42 | 1.26 | 0.0115 | 12 | 497 | 396 |
| Female Māori | -1.05 | 0.07 | 0.07 | -1.18 | -0.92 | -1.19 | -0.92 | 1.13 | 0.0082 | 6 | 227 | 202 |
| Female Pacific | -1.33 | 0.09 | 0.11 | -1.51 | -1.15 | -1.54 | -1.12 | 1.30 | 0.0249 | 13 | 90 | 71 |
| Female Asian | -0.48 | 0.08 | 0.09 | -0.62 | -0.33 | -0.65 | -0.30 | 1.46 | 0.0312 | 21 | 154 | 108 |
| Male NZE | -0.85 | 0.05 | 0.06 | -0.95 | -0.74 | -0.96 | -0.73 | 1.28 | 0.0135 | 13 | 426 | 336 |
| Male Māori | -1.40 | 0.09 | 0.10 | -1.57 | -1.23 | -1.60 | -1.20 | 1.46 | 0.0358 | 21 | 204 | 141 |
| Male Pacific | -1.71 | 0.13 | 0.16 | -1.96 | -1.47 | -2.02 | -1.41 | 1.54 | 0.0591 | 23 | 65 | 44 |
| Male Asian | -0.70 | 0.08 | 0.09 | -0.85 | -0.56 | -0.87 | -0.53 | 1.35 | 0.0240 | 16 | 162 | 122 |
| Low decile | -1.45 | 0.05 | 0.06 | -1.54 | -1.36 | -1.57 | -1.33 | 1.56 | 0.0237 | 25 | 475 | 306 |
| Mid decile | -0.88 | 0.04 | 0.05 | -0.97 | -0.79 | -0.97 | -0.79 | 1.06 | 0.0027 | 3 | 563 | 532 |
| High decile | -0.56 | 0.03 | 0.04 | -0.62 | -0.49 | -0.63 | -0.48 | 1.39 | 0.0118 | 18 | 912 | 660 |

[^9]${ }_{11}$ All results in table are quoted in logit units except where indicated.
12 Simple random sample
13 Taylor Series Linearisati
13 Taylor Series Linearisation method
Table A4.2 Reading Year 8: Comparison of results for different variance estimation methods

| Year 8 | Mean ${ }^{14}$ (SRS ${ }^{15}$ ) | $\begin{gathered} \text { SE } \\ \text { (SRS) } \end{gathered}$ | $\begin{gathered} \mathbf{S E} \\ \left(T S L^{16}\right) \end{gathered}$ | Cl <br> (SRS) <br> (lower) | (upper) | Cl <br> (TSL) <br> (lower) | Cl <br> (TSL) (upper) | Design effect | CI width increase | CI width increase | N | Effective N |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| All Year 8 | 1.01 | 0.02 | 0.03 | 0.97 | 1.06 | 0.95 | 1.07 | 1.70 | 0.0137 | 30 | 1952 | 1149 |
| NZE* | 1.23 | 0.03 | 0.04 | 1.18 | 1.29 | 1.17 | 1.30 | 1.50 | 0.0128 | 22 | 1095 | 733 |
| Māori | 0.60 | 0.05 | 0.04 | 0.51 | 0.69 | 0.51 | 0.68 | 0.98 | -0.0009 | -1 | 423 | 432 |
| Pacific | 0.29 | 0.07 | 0.08 | 0.16 | 0.42 | 0.13 | 0.45 | 1.56 | 0.0322 | 25 | 193 | 125 |
| Asian | 1.25 | 0.07 | 0.08 | 1.12 | 1.38 | 1.10 | 1.40 | 1.26 | 0.0163 | 12 | 188 | 150 |
| Female | 1.15 | 0.03 | 0.04 | 1.09 | 1.21 | 1.07 | 1.23 | 1.59 | 0.0161 | 26 | 998 | 630 |
| Male | 0.87 | 0.03 | 0.04 | 0.81 | 0.94 | 0.79 | 0.96 | 1.83 | 0.0230 | 35 | 928 | 508 |
| Female NZE | 1.36 | 0.04 | 0.05 | 1.28 | 1.44 | 1.27 | 1.45 | 1.44 | 0.0156 | 20 | 564 | 394 |
| Female Māori | 0.76 | 0.06 | 0.06 | 0.64 | 0.88 | 0.65 | 0.87 | 0.84 | -0.0105 | -8 | 222 | 266 |
| Female Pacific | 0.52 | 0.09 | 0.10 | 0.35 | 0.69 | 0.33 | 0.71 | 1.29 | 0.0230 | 13 | 107 | 85 |
| Female Asian | 1.36 | 0.10 | 0.11 | 1.17 | 1.55 | 1.15 | 1.57 | 1.24 | 0.0217 | 11 | 86 | 70 |
| Male NZE | 1.10 | 0.04 | 0.05 | 1.02 | 1.18 | 1.00 | 1.20 | 1.53 | 0.0194 | 24 | 531 | 348 |
| Male Māori | 0.42 | 0.06 | 0.07 | 0.29 | 0.54 | 0.28 | 0.55 | 1.12 | 0.0073 | 6 | 201 | 180 |
| Male Pacific | 0.00 | 0.09 | 0.10 | -0.17 | 0.18 | -0.19 | 0.19 | 1.15 | 0.0123 | 7 | 86 | 79 |
| Male Asian | 1.16 | 0.09 | 0.10 | 0.98 | 1.34 | 0.96 | 1.36 | 1.23 | 0.0197 | 11 | 102 | 84 |
| Low decile | 0.49 | 0.05 | 0.06 | 0.39 | 0.58 | 0.37 | 0.60 | 1.46 | 0.0203 | 21 | 370 | 254 |
| Mid decile | 0.98 | 0.03 | 0.04 | 0.91 | 1.05 | 0.90 | 1.06 | 1.59 | 0.0175 | 26 | 873 | 550 |
| High decile | 1.33 | 0.03 | 0.04 | 1.26 | 1.39 | 1.26 | 1.40 | 1.11 | 0.0035 | 5 | 709 | 643 |

[^10][^11]
## Appendix 5: Curriculum Alignment of the 2019 NMSSA Speaking in the English Learning Area Assessment

## Contents:

Introduction and background ..... 25

1. Assessment of speaking in the English learning area ..... 25
Administration of the assessment for speaking in the English learning area ..... 27
2. Alignment to the New Zealand Curriculum ..... 27
Knowledge of the scale ..... 27
Experience of the assessments ..... 27
3. The alignment process ..... 27
Assessment conditions ..... 27
4. Estimating score distributions ..... 28
Establishing the cut-points ..... 28
Level 3 ..... 29
5. Results ..... 29
Figures:
Figure A5.1 Overview of NMSSA assessment development process ..... 26
Figure A5.2a Example of a grid used to estimate a response distribution ..... 28
Figure A5.2b Estimating response distributions: Example of a filled-in grid ..... 28
Figure A5.3 Transforming estimated response distributions to SELA scale score cut-points ..... 29
Figure A5.4 Percentage of students achieving at each curriculum level on the SELA assessment in 2019, by year level ..... 30
Tables:
Table A5.1 The English speaking construct for the 2019 NMSSA study of the English learning area ..... 26
Table A5.2 Number of items and number of students completing each English speaking task, by year level ..... 27
Table A5.3 Judges' expected scores and actual average scores for each English speaking item, by year level ..... 29
Table A5.4 Percentage of students achieving at each curriculum level, by year level (first iteration) ..... 30
Table A5.5 Final curriculum level cut-points for the speaking in the English learning area assessment ..... 30
Table A5.6 Final percentage of students achieving at each curriculum level, by year level* ..... 30

## Introduction and background

The underlying objective of NMSSA is to report on student achievement with respect to the New Zealand Curriculum (NZC). To accomplish this, assessment data in relevant learning areas are collected each year, and achievement scales are constructed. The scales are then aligned with the levels of the NZC.

In 2019, the learning area of interest was English. One of the measures developed to assess English was Speaking in the English Learning Area (SELA). The assessment involved four different tasks presented by a teacher assessor.

This appendix describes the curriculum alignment of SELA. An alignment of an achievement scale to the NZC has not been attempted before in this area.

## 1. Assessment of speaking in the English learning area

According to the New Zealand Curriculum (NZC), English is the study, use and enjoyment of the English language and its literature, communicated orally, visually and in writing for a range of purposes and audiences in a variety of text forms (p. 18, NZC).
The English essence statement in the NZC signals that literary texts are central to the English learning area. As with previous NMSSA assessments of writing (2012), reading (2014), listening and viewing (both 2015), NMSSA again focused on literary texts (broadly defined) such as fiction, creative non-fiction, spoken poetry and scripted conversations for the 2019 NMSSA English making and creating meaning assessments, including for the assessment of speaking.

The NZC highlights the importance in the English learning area of: text purposes and audiences; ideas; language features; structure; and processes, such as critical analysis and evaluation. These elements informed the NMSSA English speaking construct.

The Speaking in the English Learning Area (SELA) assessment assessed the extent to which students: construct and convey ideas and information through spoken texts; express ideas using communicative features of spoken text to engage an audience; and analyse and evaluate the effectiveness of their use of spoken text features.

Indicators for each construct (not exclusive) are listed in Table A5.1.

| Creating meaning: Speaking |  |  |
| :---: | :---: | :---: |
| NZC English learning area | NMSSA <br> construct | Indicators |
| - Text purpose and audience <br> - Structure <br> - Ideas | Construct and convey ideas | - Conveys ideas and information through spoken texts for a range of purposes and audiences <br> - Organises and structures ideas in an appropriate sequence to engage the audience <br> - Keeps to the topic <br> - Employs a variety of sentence types and joins ideas in cohesive and interesting ways <br> - Speaks (conveys ideas) audibly and clearly |
| - Ideas <br> - Language features | Express ideas with detail and colour | - Integrates language, text features and ideas with effective spoken. communicative features to enhance meaning and the listening experience <br> - Uses verbal features like similes, metaphor, and makes deliberate language choices to enhance the spoken text <br> - Makes deliberate choices of register <br> - Uses prosodic features (pause, pace, rhythm, tone, intonation, volume) <br> - Uses paralinguistic features (e.g. whisper, giggle or laugh) <br> - Uses non-verbal features (e.g. facial expression) |
| - Critical analysis and evaluation | Critically analyse, reflect on and evaluate own spoken texts | - Analyses spoken texts and impact of presentations, questioning the features used and evaluating their effectiveness <br> - Identifies features of effective speaking <br> - Verbal <br> - Prosodic <br> - Non-verbal <br> - Evaluates their own use of the features of effective speaking |
|  |  | Text types <br> - Personal recount or memoir (prepared talk) <br> - Crafted dialogue / narrative (puppet 'play') <br> - Narrative fiction (retelling) <br> - Poetic text (presenting to an audience) |

## The assessment development process

Figure A5.1 shows an overview of the NMSSA assessment development process. This appendix addresses the transition from 'NMSSA Scales' to 'New Zealand Curriculum'.


[^12]
## Administration of the assessment for speaking in the English learning area

Experienced, specially-trained classroom teachers administered the English speaking assessment tasks during Term 3 in 2019 as part of the NMSSA English learning area study. The collection of four tasks were administered to up to eight students in each school. Table A5.2 shows the tasks, the number of items scored on each task and the number of students who completed each task, by year level.

Table A5.2 Number of items and number of students completing each English speaking task, by year level

| Task | Number of items | Year 4 students (N) | Year 8 students (N) |
| :--- | :---: | :---: | :---: |
| Talk Time | 6 | 739 | 764 |
| Pick a Poem | 4 | 387 | 389 |
| The Aliens Have Landed | 6 | 324 | 320 |
| Birds | 5 | 753 | 761 |

The SELA scale was constructed from student responses to the tasks using Item Response Theory.

## 2. Alignment to the New Zealand Curriculum

Six learning area curriculum experts were invited to participate in the alignment exercise as part of an alignment panel. All members of the panel had classroom and curriculum leadership expertise related to the English learning area. The exercise took the form of a one-day workshop and was held in Dunedin during March, 2020.

## Knowledge of the scale

The panel was presented with detailed information to help them gain a thorough understanding of the English speaking assessment framework, and its relationship to the SELA scale. Questions and discussion were encouraged at all times. Both substantive and psychometric aspects of the scale were examined. This was a critical step in the alignment exercise, and time was spent ensuring that the panel was equipped to make informed judgements about the relationship of the scale to the relevant curriculum levels.

## Experience of the assessments

Each English speaking task was introduced to the panel along with exemplars of student responses. They examined the relative difficulty and cognitive demands of each item and discussed any construct-irrelevant factors that may have affected student performance, such as shyness.

## 3. The alignment process

The alignment panel focused on one task at a time, first considering performance expectations at curriculum Level 2 and then at Level 4. For each level, the panel members conceptualised a diverse group of students, who could be considered, on balance, to be minimally competent at the curriculum level. To assist the panel to conceptualise these groups, 'minimal competence at a curriculum level' was thoroughly discussed and the panel worked towards a common understanding of the concept. The panel then focused on determining how these groups of students with 'minimal competence at a curriculum level' would be expected to perform on each item of a task. It was emphasised that students in the group would be expected to get a range of scores. The panel members were asked to imagine what the distribution of scores for the group would look like.

## Assessment conditions

It was important for panel members to understand the circumstances under which students completed the NMSSA assessments. The operational constraints of NMSSA assessments meant that, in some ways, the demands of this assessment were not completely in line with normal classroom activities. When students are less familiar with a process, and are less supported by teachers and classroom activities, they might well tend to perform at a lower level than they would if the supports were in place.

When thinking about question difficulty and how the conceptualised group of minimally competent students would respond to each question, the panel was reminded to consider the following points:

- students had no teaching support for this assessment
- there was no classroom or peer discussion to help students develop their thoughts or moves
- students had no 'scaffolding' in the form of a class study module, or project.


## 4. Estimating score distributions

The panel members provided their judgements regarding how they would expect the groups to perform using the 'Curriculum Alignment Software (CAS)' which was developed by NMSSA in 2017 specifically for curriculum alignment purposes. Figures A5.2a and A5.2b show an example of the grid that panel members were asked to use to record their judgements, before and after being filled in. The possible scores for this fictitious item of a demo task were $0,1,2$ and 3 . In the example shown, the panel member has estimated that the scores achieved by minimally competent students will be scattered across all available score points with most students scoring at least a 1. Their overall estimate of the average score for the group on the question is $(4 \times 0+14 \times 1+8 \times 2+6 \times 3) /(4+14+8+6)=1.5$.


Figure A5.2a Example of a grid used to estimate a response distribution


Figure A5.2b
Estimating response distributions: Example of a filled-in grid

Panel members worked in pairs, but entered individual judgements on their own grids. This was followed by a more general discussion and a chance to reconsider their estimated distribution of scores. There was no requirement for agreement between panel members. However, throughout the day, care was taken to challenge judgements that varied widely, or that appeared to be wildly inconsistent with assessment results. ${ }^{17}$ Justifying their thinking to each other assisted panel members in deciding whether to update their original judgements. Panel members worked in a different pairing for each task.

## Establishing the cut-points

For each item, the panel members' expected (average) raw scores were calculated, and then averaged across all panel members. The resultant expected raw scores for minimally competent students at Level 2 and Level 4 were then transformed into scale scores. These represented the cut-points on the scale where achievement at curriculum Level 2 and Level 4 started.

The expected scale scores were calculated using the Rasch model with Joint Maximum Likelihood (JML) estimation. The scale locations were based on item calibrations calculated using Marginal Maximum Likelihood (MML). The MML calibrations had been used to generate plausible values to represent achievement on the SELA scale.

[^13]

Figure A5.3 Transforming estimated response distributions to SELA scale score cut-points

## Level 3

Panel members were satisfied that Level 3 would be appropriately placed halfway between the Level 2 and Level 4 cut-points.

## 5. Results

Table A5.3 shows the judges' expected scores for each item at each curriculum level. The table also presents the average score, by year level, actually achieved by students in the 2019 study. For about one half of the items at Year 4, the judges' expected score was less than the average score achieved by Year 4 students. At Level 4, however, all but two items had a judged expected score for minimally competent students that was greater than the average score actually achieved by Year 8 students in the study.

Table A5.3 Judges' expected scores and actual average scores for each English speaking item, by year level

| Task | Judges' Level 2 expected score | Judges' Level 4 expected score | Average Year 4 score | Average Year 8 score |
| :---: | :---: | :---: | :---: | :---: |
| S001_02 | 1.23 | 1.83 | 1.12 | 1.48 |
| S001_03 | 1.22 | 1.71 | 1.05 | 1.38 |
| S001_04 | 1.15 | 1.72 | 0.96 | 1.28 |
| S001_05 | 1.10 | 1.53 | 0.87 | 1.16 |
| S001_06 | 1.02 | 1.34 | 1.02 | 1.20 |
| S001_07 | 1.04 | 1.36 | 0.97 | 1.26 |
| S002_01 | 1.17 | 1.68 | 1.29 | 1.33 |
| S002_02 | 1.04 | 1.50 | 1.19 | 1.31 |
| S002_03 | 1.09 | 1.54 | 1.08 | 1.16 |
| S002_04 | 0.98 | 1.48 | 1.02 | 0.94 |
| S002_05 | 0.92 | 1.17 | 1.05 | 1.05 |
| S002_06 | 0.77 | 1.04 | 1.06 | 0.97 |
| S003_02 | 1.03 | 1.67 | 1.29 | 1.53 |
| S003_03 | 0.92 | 1.37 | 0.92 | 1.00 |
| S003_04 | 1.02 | 1.16 | 1.16 | 1.21 |
| S003_05 | 0.65 | 0.86 | 0.40 | 0.59 |
| S004_02 | 1.30 | 2.20 | 1.85 | 2.18 |
| S004_03 | 1.31 | 1.70 | 1.33 | 1.64 |
| S004_04 | 1.18 | 1.64 | 1.21 | 1.46 |
| S004_05 | 0.74 | 1.33 | 1.12 | 1.21 |
| S004_06 | 1.14 | 1.30 | 1.68 | 1.80 |
| Total | 22.03 | 31.14 | 23.63 | 26.19 |

Table A5.4 shows the percentage of students achieving at each curriculum level based on the cut-score estimations provided by the judges.

Table A5.4 Percentage of students achieving at each curriculum level, by year level (first iteration)

| Curriculum level | Percentage of Year 4 students | Percentage of Year 8 students |
| :--- | :---: | :---: |
| Below Level 2 | 44.0 | 21.3 |
| Level 2 | 26.0 | 20.9 |
| Level 3 | 18.7 | 23.0 |
| Level 4 and above | 11.3 | 34.8 |

As a follow-up to the exercise, the proportions shown in Table A5.4 were shared with the judges, who were asked to respond. Three of the five judges were comfortable with the proportions achieving at or above expected curriculum levels. Two of the judges, however, believed the proportion at Year 8 was too low. The discussions resulted in a decision to lower the Level 4 cut-point slightly to reflect the overall view of the judging panel.

Table A5.5 shows the final minimum scale locations associated with curriculum Levels 2 to 4 on the SELA scale as agreed by the judging panel. Table A5.6 shows the proportion of students achieving at each curriculum level, by year level.

Table A5.5 Final curriculum level cut-points for the speaking in the English learning area assessment

| Curriculum level | Level 2 | Level 3 | Level 4 |
| :--- | :---: | :---: | :---: | :---: |
| Speaking in the English Learning Area (SELA) cut-points (SELA units) | 90.74 | 101.73 | 112.71 |

Table A5.6 Final percentage of students achieving at each curriculum level, by year level*

| Curriculum level | Percentage of Year 4 students | Percentage of Year 8 students |
| :--- | :---: | :---: |
| Below Level 2 | 44.0 | 21.3 |
| Level 2 | 23.1 | 17.9 |
| Level 3 | 18.1 | 20.7 |
| Level 4 and above | 14.7 | 40.1 |

* Rounding to one decimal place can mean total does not sum to 100 percent

Figure A5.4 shows the percentage of students achieving at each curriculum level, by year level.


## Appendix 6: <br> Linking English across Cycle 1 and Cycle 2

## Contents:

Introduction ..... 32

1. Writing ..... 32
Linking approach ..... 32
Change in achievement between 2012 and 2019 ..... 32
Linking error ..... 33
Standard error for differences between means ..... 33
Alignment to the New Zealand Curriculum ..... 33
2. Reading ..... 34
Linking approach ..... 34
Change in achievement over time ..... 35
Alignment to the New Zealand Curriculum ..... 36
3. Listening ..... 37
Linking approach ..... 37
Change in achievement over time ..... 38
Alignment to the New Zealand Curriculum ..... 38
Figures:
Figure A6.1 Comparison of item calibrations for English reading in 2014 and 2019 at Year 4 ..... 35
Figure A6.2 Comparison of item calibrations for English reading in 2014 and 2019 at Year 8 ..... 35
Figure A6.3 Scatterplot of the 2015 and 2019 item difficulties for English listening at Year 4 ..... 37
Figure A6.4 Scatterplot of the 2015 and 2019 item difficulties for English listening at Year 8 ..... 37
Tables:
Table A6.1 Percentage of Year 4 and Year 8 students achieving at each curriculum level in 2012 ..... 33
Table A6.2 Curriculum level cut-scores on the 2019 Writing in the English Learning Area (WELA) scale ..... 33
Table A6.3 Curriculum level cut-scores on the 2019 Reading in the English Learning Area scale ..... 36
Table A6.4 Curriculum level cut-scores on the 2019 Listening in the English Learning Area scale ..... 38

## Introduction

This appendix describes the processes used to link results across Cycle 1 and Cycle 2 for writing, reading and listening in the English learning area, for the purposes of comparing student achievement.

## 1. Writing

In 2012, the Writing for a Variety of Purposes (WVP) scale was constructed using data gathered from a groupadministered paper-and-pencil writing assessment. The same assessment was used in 2019.

## Linking approach

In 2019, a Multifacet Rasch analysis of the combined 2012 and 2019 NMSSA writing data was used to generate a measurement scale called the Writing in the English Learning Area (WELA) scale. Because 2012 students' writing achievement was being measured on the same scale as 2019 students' writing achievement, results from 2012 could then be directly compared with results from 2019, circumventing the need for a separate linking exercise. This aproach was possible because:

- Both cycles utilised an assessment design that allowed for Multifacet Rasch analysis with student, marker, task and criteria facets.
- All five writing prompts used in 2019 were also used in 2012.
- The same judgement criteria were used for marking across both cycles.
- A subset of scripts from 2012 were able to be re-marked in 2019.

In 2012, scale construction involved supplementing the Year 4 and 8 NMSSA samples with students in Years 5, 6 and 7 for the purposes of vertical linking. This ensured a strong vertical link between Year 4 and Year 8 students. Each student completed two writing prompts, so that the prompts could be linked with one another.

In 2019, only Year 4 and Year 8 students were assessed, and only one prompt was completed by each student. Within both cycles, a sample of scripts was marked multiple times by different markers to link markers with one another and to ensure a robust marking practice. In order to link markers across cycles, 400 scripts from 2012 were remarked by 2019 markers. Results were then able to be compared.
The reanalysis of 2012 data was done solely for the purposes of analysing change in achievement between 2012 and 2019. It means that results reported in the 2012 NMSSA English: writing report cannot be directly compared with those in the 2019 report. Meaningful comparisons across time are restricted to those reported in the change in achievement analysis sections of the 2019 reports.

## Change in achievement between 2012 and 2019

In 2019, Year 4 students scored lower, on average, than Year 4 students in 2012 by 2 WELA scale score units. The difference was statistically significant ( $\mathrm{p}<.05$ ). There were also some statistically significant differences between cycles at the subgroup level for Year 4. In all cases, the 2019 average was lower than 2012. Subgroups with statistically significant differences were:

- Year 4 boys
- Year 4 New Zealand European/Pākehā
- Year 4 high decile.

There were no statistically significant differences between the cycles at Year 8.
More detailed findings from change over time analysis are included in the NMSSA Report 22: English 2019 - Key Findings.

## Linking error

Because linking was included in the scale construction process (by using the combined 2012 and 2019 NMSSA writing data), a separate linking exercise was not necessary. To produce an estimate of linking error, we calculated the standard error of the (dummy) facet associated with the calendar year (2012 or 2019) in which a script was marked. In this case the linking error was 0.11 scale score units.

## Standard error for differences between means

Change over time analysis involved examining differences between mean scores in 2012 and 2019, for complete year levels and for key subgroups. The formula used for calculating a confidence interval around the observed difference was:

$$
1.96 * \sqrt{s e_{\text {pooled }}^{2}+\text { linking } \text { error }^{2}}
$$

## Alignment to the New Zealand Curriculum

NMSSA has a particular interest in the achievement of Year 4 students against Level 2 of the New Zealand Curriculum (NZC), and the achievement of Year 8 students against level 4 of the curriculum.

Curriculum alignment carried out in 2012 generated boundaries on the WVP scale to indicate curriculum level cutscores. As the WVP scale was based on the same framework and scoring rubrics as e-asTTle writing, the cut-points for the e-asTTle writing scale were applied to the WVP scale, using a linking exercise that is described in the 2012 English writing report. ${ }^{18}$ These cut-scores were then used to determine how the Year 4 and Year 8 student samples were achieving against year-level curriculum expectations.

Percentile equating was used to locate the curriculum level cut-scores from the 2012 WVP scale on the 2019 WELA scale. Percentile equating assumes that the proportion of 2012 students achieving at or above the expected curriculum level on the WVP scale, should be the same as the proportion of 2012 students achieving at or above the expected curriculum level on the WELA scale.

Table A6.1 shows the percentage of students achieving at each curriculum level, according to 2012 NMSSA reporting.

Table A6.1 Percentage of Year 4 and Year 8 students achieving at each curriculum level in 2012

| Year level | Curriculum level | Cut-score (WELA units) |
| :---: | :---: | :---: |
|  | Level 1 | 35 |
|  | Level 2 | 45 |
|  | Level 3 | 18 |
| 8 | Level 4+ | 2 |
|  | Level 1 | 5 |
|  | Level 2 | 23 |

Cut-scores for curriculum levels on the WELA scale were set so that the percentage of 2012 students achieving at each curriculum level matched, as closely as possible, the percentages reported against the 2012 WVP scale. The resultant cut-scores are shown in Table A6.2

Table A6.2 Curriculum level cut-scores on the 2019 Writing in the English Learning Area (WELA) scale

| Curriculum level | Cut-score (WELA units) |
| :---: | :---: |
| Level 2 | 80.9 |
| Level 3 | 103.2 |
| Level 4 | 121.5 |

[^14]
## 2. Reading

In 2014, the Knowledge and Application of Reading in English (KARE) scale was constructed using items from a group-administered paper-and-pencil assessment, as well as items from in-depth interview tasks. In 2019, NMSSA assessed English reading using a paper-and-pencil assessment made up of a combination of existing items from 2014 and items newly developed for 2019. A subset of the in-depth interview tasks from 2014 were also used in the 2019 assessment.

## Linking approach

## Reconstructing the 2014 scale

In 2014, student scores on the KARE scale were produced using Maximum Likelihood estimation. From 2015, NMSSA started using Plausible Values for calculating population statistics. As generating sets of plausible values involves using Expected A Posteriori estimation rather than Maximum Likelihood estimation, it was necessary to reconstruct the 2014 scale so that data from 2014 and 2019 could be compared. The use of different estimation methods for reporting across the two cycles means that measures provided in the 2014 NMSSA English: Reading report cannot be directly compared with those in the 2019 report. Meaningful comparisons across time are restricted to those reported in the change in achievement over time sections of the 2019 reports.

The reconstruction of the 2014 scale was carried out using the same exclusions and recodings applied in 2014. The original calibration anchored the group-administered items before adding the in-depth items. Calibrating all of the items together gave almost identical results to the original calibration, so this simpler process was used. A set of 50 plausible values for each student was created from the results of the recalibration.

## Relocating curriculum level cut-scores

With a reconstructed scale for 2014, the curriculum level cut-scores needed to be relocated. For each curriculum level boundary, a group of items surrounding the cut-score on the KARE scale was selected. The recalibrated item difficulties were then identified and the new cut-score was located at the average of those item difficulties.

When the original 2014 scale scores were compared with the newly generated ones, there was some slight variation in the tails of the curriculum level distributions. However, the percentage of students achieving at or above expected curriculum levels was quite consistent. Because analyses of achievement over time are concerned primarily with the Level 2 cut-score for Year 4 and the Level 4 cut-score for Year 8, differences in the tails were not of particular concern.

## Analysing 2019 data

The 2014 NMSSA study used results from a sample of Year 6 students to link Year 4 and Year 8 students together vertically on the scale. In 2019, no Year 6 students were assessed. There was also no overlap between Year 4 and Year 8 items in 2019. Many of the items were retained from 2014, but some changes were made that could potentially affect item functioning to varying degrees. The changes included removing all cloze-type tasks and developing a number of new tasks, primarily at Year 8. There were also changes made to rubrics used to score some of the open-ended responses.

Due to the lack of vertical linking in 2019, the preferred method for estimating student scores for 2019 was to use the recalibrated 2014 item difficulties as anchors to construct a scale for 2019. This would make the two scales equivalent and enable direct comparisons.

To determine whether this was appropriate, it was first necessary to create a new scale using the 2019 data, without fixing any of the item difficulties. This meant that item functioning could be compared across the cycles, particularly for those items that had been changed in some way. As there were no common items at Year 4 and Year 8, a separate scale was created for each level. To maintain consistency, the exclusions and recodings from the original 2014 analysis were used throughout.

Figures A6.1 and A6.2 show the item difficulties from the 2014 recalibration plotted against the item difficulties from the new 2019 calibration. In these figures, the red lines represent a confidence interval around the theoretical line, $x=y$. Items that fall outside these lines were considered to have significantly different item function between 2014 and 2019 and were reclassified as 'new' items. For the most part, differences in item functioning across the cycles could be attributed to changes in marking rubrics.


Figure A6.1 Comparison of item calibrations for English reading in 2014 and 2019 at Year 4


Figure A6.2 Comparison of item calibrations for English reading in 2014 and 2019 at Year 8

The final 2019 scale, called the Reading in the English Learning Area (RELA) scale, was constructed by fixing the difficulties of similarly-functioning common items using values extracted from the reconstructed 2014 scale, and leaving the remaining items to 'float'. A set of 50 plausible values for each student was then created for the 2019 sample.

## Change in achievement over time

Between 2014 and 2019, there was no significant difference in average scale score at Year 4 or Year 8 overall ( $\mathrm{p}<.05$ ). However, there was a statistically significant difference for one subgroup:

- Year 8 girls scored, on average, lower in 2019 than 2014.

More detailed findings are included in the NMSSA Report 22: English 2019 - Key Findings.

## Linking error

The reconstructed 2014 scale and the 2019 RELA scale are considered equivalent due to the process of anchoring items. However, it is reasonable to assume some effect on the uncertainty around estimates of using 2014 item parameters with 2019 respondents. In order to approximate a measure of this uncertainty, pairwise differences between the anchored item difficulties and the respective difficulties obtained from the new 2019 scales were used. The linking error was calculated as 0.70 scale score units at Year 4, and 0.86 scale score units at Year 8.

The formula used for calculating linking error was: $\sqrt{\sum_{i=1}^{L}\left(\delta_{i}-\delta_{i}^{\prime}\right)^{2} * \frac{1}{L(L-1)}}$, where $L$ is the number of link items.

## Standard error for differences between means

The formula used for calculating the confidence interval around an observed difference was:

$$
1.96 * \sqrt{s e_{\text {pooled }}^{2}+\text { linking error }}{ }^{2}
$$

## Alignment to the New Zealand Curriculum

The relocated curriculum level cut-scores from the reconstructed 2014 scale were able to be applied directly to the RELA scale for the purposes of 2019 reporting. These cut-scores are presented in Table A6.3.
Table A6.3 Curriculum level cut-scores on the 2019 Reading in the English Learning Area scale

| Curriculum level | Cut-score (RELA units) |
| :---: | :---: |
| Level 2 | 76.6 |
| Level 3 | 98.8 |
| Level 4 | 115.8 |

## 3. Listening

In 2015, the Knowledge and Application of Listening in English (KALE) scale was constructed using items from a group-administered paper-and-pencil assessment. In 2019, the same paper-and-pencil assessment was used.

## Linking approach

The 2015 KALE scale was constructed using results from the Year 4 and Year 8 students in the NMSSA sample, as well as a group of Year 6 students for vertically linking Year 4 and Year 8 together on the scale. The 2019 data collection did not involve any Year 6 students. Because the 2015 and 2019 English listening assessments were identical, it was theoretically possible to simply use the 2015 item difficulties for analysing the 2019 data. To determine whether this was appropriate, a new scale was constructed using the 2019 data, and item function compared with 2015.

The 2019 and 2015 item difficulties were very strongly correlated. At Year 4, r=0.98; and at Year 8, r=0.93. The relationship between the 2015 and 2019 item difficulties for Year 4 items is shown in Figure A6.3, and for Year 8 items, in Figure A6.4. In general, the strength of the relationships supports the use of the 2015 scale for analysis of the 2019 data.



Figure A6.4 Scatterplot of the 2015 and 2019 item difficulties for English listening at Year 8

A final 2019 scale, called the Listening in the English Learning Area (LELA) scale, was constructed by fixing the difficulties of all but two items (one of which had been deleted in the 2015 analysis, and the other that seemed to have been marked differently).

## Change in achievement over time

Between 2015 and 2019, there was no significant difference in average scale score at Year 4 or Year 8 overall ( $\mathrm{p}<.05$ ). However, there were some statistically significant differences among subgroups:

- Year 8 boys scored lower in 2019 than in 2015, on average
- Year 4 and Year 8 New Zealand European/Pākehā students scored lower in 2019 than in 2015, on average
- Year 4 and Year 8 students in high decile schools scored lower in 2019 than in 2015, on average,

More detailed findings are included in the NMSSA Report 22 ... English 2019- Key Findings.

## Linking error

The 2015 and 2019 English listening scales are considered equivalent. However, it is reasonable to assume some effect, on the uncertainty around estimates, of using 2015 item parameters with 2019 respondents. In order to approximate a measure of this uncertainty, the pairwise differences between the 2015 difficulties and the difficulties obtained from the new calibration of 2019 data were used. The linking error was calculated as 1.07 scale score units at Year 4, and 1.48 scale score units at Year 8.

The linking error was calculated as: $\sqrt{\sum_{i=1}^{L}\left(\delta_{i}-\delta_{i}^{\prime}\right)^{2} * \frac{1}{L(L-1)}}$, where $L$ is the number of link items.

## Standard error for differences between means

The formula used for calculating the confidence interval around an observed difference was:

$$
1.96 * \sqrt{\text { se }_{\text {pooled }}^{2}+\text { linking error }}{ }^{2}
$$

## Alignment to the New Zealand Curriculum

A curriculum alignment exercise in 2015 generated curriculum level cut-scores on the KALE scale. These cut-scores have been applied directly to the 2019 LELA scale. These cut-scores are presented in Table A6.4.

Table A6.4 Curriculum level cut-scores on the 2019 Listening in the English Learning Area scale

| Curriculum level | Cut-score (LELA units) |
| :---: | :---: |
| Level 2 | 72 |
| Level 3 | 87 |
| Level 4 | 102 |

## Appendix 7: <br> NMSSA Assessment Framework for English 2019

## Contents:

Introduction ..... 40
English in The New Zealand Curriculum ..... 40

1. Assessing achievement in English ..... 40
Focus on 'literary' texts in the English learning area ..... 41
2. Constructs for the 2019 assessments ..... 41
3. What does progress in the English learning area look like? ..... 41
4. Curriculum coverage in the assessments of the English learning area ..... 44
Making meaning ..... 44
Creating meaning ..... 47
References ..... 48
Tables:
Table A7.1 The constructs for the making-meaning and creating-meaning strands for assessing the English learning area ..... 41
Table A7.2 Curriculum Level 2 and Level 4 achievement objectives and indicators in the English learning area ..... 42
Table A7.3 Number of items by construct and text type for the RELA assessment at Year 4 ..... 44
Table A7.4 Number of items by construct and text type for the RELA assessment at Year 8 ..... 44
Table A7.5 Number of items by construct and text type for the LELA assessment at Year 4 ..... 45
Table A7.6 Number of items by construct and text type for the LELA assessment at Year 8 ..... 45
Table A7.7 Number of items by construct and text type for the VELA assessment at Year 4 ..... 46
Table A7.8 Number of items by construct and text type for the VELA assessment at Year 8 ..... 46
Table A7.9 Number of items by construct and moving image for the VELA assessment at Year 4 and Year 8 ..... 46
Table A7.10 Coverage of constructs by the elements for the WELA ..... 47
Table A7.11 Number of items by construct for the SELA assessment at Year 4 and Year 8 ..... 47
Table A7.12 Number of items by construct for the PELA assessment at Year 4 and Year 8 ..... 48

## Introduction

This appendix describes the assessment approach that the National Monitoring Study of Student Achievement (NMSSA) took to assess the English learning area in 2019. It describes how the English learning area is set out in The New Zealand Curriculum (2007) (NZC) and outlines the conceptual framework that guided the development of the assessments for the six modes of English: reading, listening, viewing, writing, speaking and presenting.

## English in The New Zealand Curriculum

According to the NZC:
English is the study, use, and enjoyment of the English language and its literature, communicated orally, visually, and in writing for a range of purposes and audiences in a variety of text forms ... [it] gives students access to the understanding, knowledge, and skills needed 'to participate fully in the social, cultural, political, and economic life of New Zealand and the wider world'. (p. 18)

More specifically the NZC describes how success in English enables students to:

- become effective oral, written, and visual communicators who are able to think critically and in-depth
- make critically informed choices about their use of language in different contexts
- deconstruct and critically analyse texts
- appreciate and enjoy texts in a variety of forms
- build a sense of identity, of New Zealand's bicultural heritage, and of the world
- develop the key competencies
- achieve success across the curriculum.

Over the last decade or so, the Ministry of Education has developed a number of resources designed to support literacy across the curriculum in the primary school, such as: The Learning Progressions Frameworks; The PaCT reading framework; The PaCT writing framework; The Literacy Learning Progressions (Ministry of Education, 2010); Effective Literacy Practice in Years 1 to 4 (Ministry of Education, 2003); and Effective Literacy Practice in Years 5 to 8 (Ministry of Education, 2006).

## 1. Assessing achievement in English

The English learning area is structured around two strands. One of these strands - making meaning - focuses on the interpretation of text through reading, listening and viewing. The other strand - creating meaning - focuses on the production of text through writing, speaking and presenting. This approach is consistent with the way in which the English learning area is organised in the NZC. In addition, the NZC highlights the importance in the English learning area of: text purposes and audiences; ideas; language features; and structure.

The 2019 NMSSA assessment programme was based around an assessment for each mode of English. For five modes, NMSSA scales were constructed, which provided continuity with the four scales developed for writing, reading, listening and viewing in 2012, 2014 and 2015, respectively. The scales constructed included:

- Writing in the English Learning Area [WELA]
- Reading in the English Learning Area [RELA]
- Listening in the English Learning Area [LELA]
- Viewing in the English Learning Area [VELA]
- Speaking in the English Learning Area [SELA].

No scale was constructed for the presenting assessment.

## Focus on 'literary' texts in the English learning area

Previous NMSSA assessments in the English learning area - writing, reading, listening and viewing - focused on what might broadly be described as literary texts. The texts used in reading and listening included fiction (such as novels, short stories, plays, poems) and creative non-fiction, i.e. formal and discursive texts (such as essays, memoirs, biographies). Listening also used spoken poetry and scripted conversations. The texts used in viewing included: picture books and graphic novels. Writing also focused on writing for 'English purposes' - for example, character description, narration of a story, persuade, recount of a personal event.

Given that, as signalled in the English essence statement, literary texts are central to the English learning area, we continued to use literary texts (broadly defined) for the 2019 assessment in the English learning area. This focus informed the development of the new tasks in viewing (moving images), presenting and speaking.

## 2. Constructs for the 2019 assessments

The meaning-making strand and the creating-meaning strand are each represented by three constructs (Table A7.1).

Table A7.1 The constructs for the making-meaning and creating-meaning strands for assessing the English learning area

| Making meaning Reading Listening Viewing |  | Creating meaning <br> Writing Speaking Presenting |  |
| :---: | :---: | :---: | :---: |
| Construct | Definition | Construct | Definition |
| Locate and recall* | Can identify the information, ideas and features of print, oral and visual texts. | Construct and convey ideas | Can convey ideas and information through print, oral and visual texts for a range of purposes and audiences. |
| Interpret* | Can interpret print, oral, and visual texts by integrating text features and ideas, considering the relationship between ideas and text features, and by making inferences. | Express ideas with detail and colour | Can integrate text features and ideas when creating written, oral and visual texts. <br> Can engage the reader, listener or viewer through use of communicative features specific to the mode. <br> Can employ imagery and allusion. |
| Critically analyse* | Can critically analyse print, oral and visual texts by questioning texts rather than taking them at face value. This involves considering the construction of texts; questions of inclusion, exclusion and representation; and the ways in which texts can position a reader. | Critically analyse | Can analyse their own processes and impact of presentations, questioning the features used, and evaluating their effectiveness. <br> Can make deliberate choices of text structure, register and tone and use specific oral, visual or written language features to position the reader, viewer or listener. |

* These constructs were adapted (in a minor way) from the constructs used in Cycle 1 English assessments


## 3. What does progress in the English learning area look like?

The achievement objectives of the NZC ${ }^{19}$ specify what students are expected to achieve at each year level. It is expected that Year 4 students achieve, on balance, at Level 2 or above, and Year 8 students achieve, on balance, at Level 4 or above. Table A7.2, on the following page, lists the achievement objectives and indicators for Level 2 and Level 4 of the NZC.

[^15]Table A7.2 Curriculum Level 2 and Level 4 achievement objectives and indicators in the English learning area

| Making Meaning: reading, listening, viewing |  |
| :---: | :---: |
| Level 2 | Level 4 |
| Processes and strategies <br> Select and use sources of information, processes, and strategies with some confidence to identify, form and express ideas. <br> - selects and reads texts for enjoyment and personal fulfilment <br> - recognises connections between oral, written and visual language <br> - selects and uses sources of information ... and prior knowledge with growing confidence to make sense of increasingly varied and complex texts <br> - selects and uses processing strategies and an increasing range of comprehension strategies with some understanding and confidence <br> - thinks critically about texts with some confidence <br> - monitors, self-evaluates and describes progress with some confidence <br> - uses an increasing knowledge of letter clusters, affixes, roots and compound words to confirm predictions. | Processes and strategies <br> Integrate sources of information, processes and strategies confidently to identify, form and express ideas. <br> - selects and reads texts for enjoyment and personal fulfilment <br> - recognises and understands the connections between oral, written and visual language <br> - integrates sources of information and prior knowledge confidently to make sense of increasingly varied and complex texts <br> - selects and uses appropriate processing and comprehension strategies with increasing understanding and confidence <br> - thinks critically about texts with increasing understanding and confidence <br> - monitors, self-evaluates, describes progress,and articulates learning with confidence. |
| Purposes and audiences <br> Show some understanding of how texts are shaped for different purposes and audiences. <br> - recognises how texts are constructed for different purposes, audiences and situations <br> - understands that texts are created from a particular point of view <br> - evaluates the reliability and usefulness of texts with some confidence. | Purposes and audiences <br> Show an increasing understanding of how texts are shaped for different purposes and audiences. <br> - recognises and understands how texts are constructed for a range of purposes, audiences and situations <br> - identifies particular points of view and recognises that texts can position a reader <br> - evaluates the reliability and usefulness of texts with increasing confidence. |

## Ideas

Show some understanding of ideas within, across and beyond texts.

- uses their personal experience and world and literacy knowledge to make meaning from texts
- makes meaning of increasingly complex texts by identifying main ideas
- makes and supports inferences from texts with some independence.


## Ideas

Show an increasing understanding of ideas within, across and beyond texts.

- makes meaning of increasingly complex texts by identifying and understanding main and subsidiary ideas and the links between them
- makes connections by thinking about underlying ideas within and between texts from a range of contexts
- recognises that there may be more than one reading available within a text
- makes and supports inferences from texts with increasing independence.


## Language features

Show some understanding of how language features are used for effect within and across texts.

- recognises that oral, written and visual language features can be used for effect
- uses a large and increasing bank of high-frequency, topicspecific and personal-content words to make meaning
- shows an increasing knowledge of the conventions of text
- recognises that authors have different voices and styles.


## Structure

Show some understanding of text structures.

- understands that the order and organisation of words, sentences, paragraphs and images contribute to text meaning
- recognises an increasing range of text forms and differences between them.


## Language features

Show an increasing understanding of how language features are used for effect within and across texts.

- identifies oral, written and visual features used and recognises and describes their effects
- uses an increasing vocabulary to make meaning
- shows an increasing knowledge of how a range of text conventions can be used appropriately and effectively
- knows that authors have different voices and styles and can identify and describe some of these differences


## Structure

Show an increasing understanding of text structures.

- understands that the order and organisation of words, sentences, paragraphs and images contribute to and affect meaning in a range of texts
- identifies an increasing range of text forms and recognises and describes their characteristics and conventions.

| Creating Meaning: writing, speaking, presenting |  |
| :---: | :---: |
| Level 2 | Level 4 |
| Processes and strategies <br> Select and use sources of information, processes, and strategies with some confidence to identify, form and express ideas. <br> - show some understanding of the connections between oral, written and visual language when creating texts <br> - creates texts by using meaning, structure, visual and grapho-phonic sources of information, and processing strategies with growing confidence <br> - seeks feedback and makes changes to texts to improve clarity and meaning <br> - is reflective about the production of texts: monitors, selfevaluates and describes progress with some confidence. | Processes and strategies <br> Integrate sources of information, processes and strategies confidently to identify, form and express ideas. <br> - uses an increasing understanding of the connections between oral, written and visual language when creating texts <br> - creates a range of texts by integrating sources of information and processing strategies with increasing confidence <br> - seeks feedback and makes changes to texts to improve clarity, meaning and effect <br> - is reflective about the production of own texts: monitors and self-evaluates progress, articulating learning with confidence. |
| Purposes and audiences <br> Show some understanding of how to shape texts for different purposes and audiences. <br> - constructs texts that demonstrate a growing awareness of audience and purpose through appropriate choice of content, language and text form <br> - expects the texts they create to be understood, responded to and appreciated by others <br> - develops and conveys personal voice where appropriate. | Purposes and audiences <br> Show an increasing understanding of how to shape texts for different purposes and audiences. <br> - constructs texts that show an awareness of purpose and audience through deliberate choice of content, language and text form <br> - conveys and sustains personal voice where appropriate. |
| Ideas <br> Select, form and express ideas on a range of topics. <br> - forms and expresses ideas and information with reasonable clarity, often drawing on personal experience and knowledge <br> - begins to add or delete details and comments, showing some selectivity in the process. | Ideas <br> Select, develop and communicate ideas on a range of topics. <br> - forms and communicates ideas and information clearly, drawing on a range of sources <br> - adds or changes details and comments to support ideas, showing thoughtful selection in the process <br> - ideas show increasing awareness of a range of dimensions or viewpoints. |
| Language features <br> Use language features appropriately, showing some understanding of their effects. <br> - uses oral, written and visual language features to create meaning and effect <br> - uses a large and increasing bank of high-frequency, topicspecific, and personal-content words to create meaning <br> - spells most high-frequency words correctly and shows growing knowledge of common spelling patterns <br> - uses a range of strategies to self-monitor and self-correct spelling <br> - writes legibly and with increasing fluency when creating texts <br> - gains increasing control of text conventions, including some grammatical conventions. | Language features <br> Use a range of language features appropriately, showing an increasing understanding of their effects. <br> - uses a range of oral, written and visual features to create meaning and effect and to sustain interest <br> - uses a range of vocabulary to communicate precise meaning <br> - demonstrates a good understanding of spelling patterns in written English, with few intrusive errors <br> - uses a wide range of strategies to self-monitor and selfcorrect spelling <br> - writes with increasing speed and endurance to suit the nature of the task and its purpose, without significant loss of legibility <br> - uses a range of text conventions, including grammatical conventions, appropriately, effectively and with increasing accuracy. |
| Structure <br> Organise texts, using a range of structures. <br> - uses knowledge of word and sentence order to communicate meaning when creating texts <br> - organises and sequences ideas and information with some confidence <br> - begins to use a variety of sentence structures, beginnings and lengths. | Structure <br> Organise texts, using a range of appropriate structures. <br> - achieves some coherence and wholeness when constructing texts <br> - organises and sequences ideas and information for a particular purpose or effect <br> - uses a variety of sentence structures, beginnings and lengths for effect. |

## 4. Curriculum coverage in the assessments of the English learning area

Using the conceptual framework for guidance, collections of tasks and items were developed around appropriate texts topics for the meaning-making strand, and appropriate topics for the creating-meaning strand. The number of tasks and items within each construct for each mode are summarised in Tables A7.3 to A7.12.

## Making meaning

## Reading in the English Learning Area (RELA)

For the RELA assessment there were five reading passages at Year 4 and nine reading passages at Year 8 across three text types. Select response and short response items were used and marked $0-1$. The numbers of items included in the group-administered task booklets and interview to assess each of the constructs and vocabulary are shown for Year 4 in Table A7.3 and for Year 8 in Table A7.4. At Year 4, a large proportion of the items emphasised the constructs of locate and recall, and interpret information. At Year 8, the items focused more heavily on the construct of interpret information.

Table A7.3 Number of items by construct and text type for the RELA assessment at Year 4

|  |  | Constructs |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Text types | Vocabulary | Locate and recall | Interpret | Critically analyse | Totals |
| Poetry | 1 | 5 | $9(2)$ | $2(2)$ | $17(4)$ |
| Literary fiction | 1 | $11(1)$ | $8(1)$ | 1 | $21(2)$ |
| Poetry | 2 | $4(1)$ | $3(1)$ | $1(1)$ | $10(3)$ |
| Totals | 4 | $20(2)$ | $20(4)$ | $4(3)$ | $48(9)$ |

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

Table A7.4 Number of items by construct and text type for the RELA assessment at Year 8

|  |  | Constructs |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Text types | Vocabulary | Locate and recall | Interpret | Critically analyse | Totals |
| Poetry | 1 | 5 | $27(1)$ | $2(1)$ | $35(2)$ |
| Literary fiction | 2 | 4 | $17(1)$ | $3(2)$ | $26(3)$ |
| Poetry | 4 | 2 | 3 | $2(2)$ | $11(2)$ |
| Totals | 7 | 11 | $47(2)$ | $7(5)$ | $72(7)$ |

* Numbers in parentheses indicate items asked in one-to-one interviews with a subsample of Year 8 students


## Listening in the English Learning Area (LELA)

For the LELA assessment there were seven audio scripts of three text types at each year level. Tables A7.5 and A7.6 show that the majority of the items were categorised under the 'interpret' construct.

Table A7.5 Number of items by construct and text type for the LELA assessment at Year 4

|  |  | Constructs |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Text types | Locate and recall | Interpret | Critically analyse | Totals |
| Fiction - extracts, stories, poetry <br> (including performance poetry) | 3 | 30 | 2 | 35 |
| Literary non-fiction | 1 | 2 | 1 | 4 |
| Conversations (scripted, as near as <br> possible to authentic) | - | 5 | - | 5 |
| Totals | 4 | 37 | 3 | 44 |

Table A7.6 Number of items by construct and text type for the LELA assessment at Year 8

|  |  | Constructs |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Text types | Locate and recall | Interpret | Critically analyse | Totals |
| Fiction - extracts, stories, poetry <br> (including performance poetry) | 4 | 17 | 2 | 23 |
| Literary non-fiction 1 7 1 |  |  |  |  |
| Conversations (scripted, as near as <br> possible to authentic) | 1 | 5 | - | 9 |
| Totals | 6 | 29 | 3 | 6 |

## Viewing in the English Learning Area (VELA)

The VELA assessment comprised two components: a group-administered component and an in-depth (InD) component. For the VELA group-administered assessment there were five tasks presented in booklets and items were scored 0 or 1 . The items assessing each of the constructs for the booklets are shown for Year 4 in Table A7.7 and for Year 8 in Tables A7.8. The original constructs of 'Locate and recall' and 'Interpret' were collapsed into 'Locate and recall, Interpret' for marking purposes. Note that graphic novel extracts were included in the assessment of Year 8 students only.

Table A7.7 Number of items by construct and text type for the VELA assessment at Year 4

|  |  | Constructs |  |
| :--- | :---: | :---: | :---: |
| Text types | Locate and recall | Interpret | Critically analyse |
| Picture book illustrations with written <br> text | 12 | 2 | 14 |
| Picture book illustrations without written <br> text | 12 | 1 | 13 |
| Totals | 24 | 3 | 27 |

Table A7.8 Number of items by construct and text type for the VELA assessment at Year 8

|  |  | Constructs |  |
| :--- | :---: | :---: | :---: |
| Text types | Locate and recall | Interpret | Critically analyse |
| Picture book illustrations with written <br> text | 14 | 3 | 17 |
| Picture book illustrations without written <br> text | 14 | 2 | 16 |
| Extracts from graphic novels | 9 | 1 | 10 |
| Totals | 37 | 6 | 43 |

Five tasks of viewing moving images (short film clips) were presented on laptops and were scored on a scale from $0-2$ or $0-3$. Both Year 4 and Year 8 students undertook the same tasks. Some items assessed several constructs. Table A7.9 presents the constructs assessed through the items for each task.

Table A7.9 Number of items by construct and moving image for the VELA assessment at Year 4 and Year 8

|  |  | Constructs |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Text types | Locate and recall | Interpret | Critically analyse | Totals |
| Mouse for Sale | 2 | 5 | 3 | 10 |
| After School | 2 | 5 | 5 | 12 |
| Jonasi's Underwater Find | 3 | 3 | 5 | 11 |
| Māui's Catch | 4 | 4 | 4 | 12 |
| Birds | 1 | 1 | 1 | 3 |
| Totals | 12 | 18 | 18 | 48 |

## Creating meaning

## Writing in the English Learning Area (WELA)

For the WELA assessment, five purposes for writing were chosen: to describe, explain, persuade, narrate and recount. Prompts were developed for each purpose. Students were expected to write to one prompt for up to 40 minutes. Students' writing was marked using the detailed e-asTTle ${ }^{20}$ rubric that evaluated seven elements of writing: ideas, organisation, punctuation, sentence structure, structure and language, spelling and vocabulary. The rubric was supported by examples/exemplars of student writing at different levels, with annotations explaining the marking decisions. The e-asTTle writing tasks were not designed a priori to match the NMSSA constructs for creating meaning. Therefore, the elements were mapped backwards to the three constructs; some elements overlapped on constructs. Table A7.10 shows that the writing task strongly represented the construct of 'constructing and conveying ideas'.

Table A7.10 Coverage of constructs by the elements for the WELA

|  | Constructs |  |
| :--- | :---: | :---: |
| Elements | Construct and convey Ideas | Express ideas with detail and colour |
| Ideas | V | V |
| Structure and language | V | V |
| Organisation | V |  |
| Punctuation | V |  |
| Sentence structure | V |  |
| Spelling | V | V |
| Vocabulary | V |  |

## Speaking in the English Learning Area (SELA)

The SELA assessment consisted of four InD tasks taken by students at Year 4 and Year 8. Three tasks were undertaken in groups of four students. Individual students chose a personal topic and a poem to present to the group. In pairs, students used puppets to create and present a conversation. The final task that required students to retell a story was undertaken in the one-to-one interview. Some items assessed more than one construct and all were marked on a scale of $0-2$ or $0-3$. Table A7.11 presents the constructs assessed through the items for each task. Each construct was equally represented across the set of tasks.

Table A7.11 Number of items by construct for the SELA assessment at Year 4 and Year 8

|  |  | Constructs |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Text types | Construct and <br> convey Ideas | Express ideas with <br> detail and colour | Critically <br> analyse | Totals |
| Talk Time (present a personal topic) | 6 | 4 | 3 | 13 |
| The Aliens Have Landed <br> (paired puppet conversation) | 6 | 5 | 5 | 16 |
| Pick a Poem (speak to text) | 1 | 4 | 4 | 9 |
| Birds (retell a story) | 5 | 4 | 5 | 14 |
| Totals | 18 | 17 | 17 | 52 |

## Presenting in the English Learning Area (PELA)

The PELA assessment consisted of one InD task at both year levels. The students identified a key message from a short film and created a poster to convey the message to students at their school. Students discussed their poster design decisions in one-to-one interviews. All responses were marked on a scale of $0-2$ or $0-3$. Table A7.12 presents the coverage of the constructs for the poster task.

Table A7.12 Number of items by construct for the PELA assessment at Year 4 and Year 8

|  | Constructs |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Text types | Construct and <br> convey Ideas | Express ideas with <br> detail and colour | Critically <br> analyse | Totals |
| Poster | 8 | 6 | 7 | 21 |

## References

EARU \& NZCER (2013). NMSSA English: Writing 2012. Wellington, New Zealand: Ministry of Education.
EARU \& NZCER (2015). NMSSA English: Reading: 2014 Key findings. Wellington, New Zealand: Ministry of Education.
EARU \& NZCER (2016). NMSSA English: Listening: 2015 - Key findings. Wellington, New Zealand: Ministry of Education.

EARU \& NZCER (2016). NMSSA English: Viewing: 2015 - Key findings. Wellington, New Zealand: Ministry of Education.
Ministry of Education (2003). Effective literacy practice in years 1 to 4. Wellington, New Zealand: Learning Media Ltd.
Ministry of Education.(2006). Effective literacy practice in years 5 to 8. Wellington, New Zealand: Learning Media Ltd.
Ministry of Education (2007). The New Zealand curriculum. Wellington, New Zealand: Learning Media Ltd.
Ministry of Education (2009). Reading and writing standards for years 1-8. Wellington, New Zealand: Learning Media Ltd.
Ministry of Education (2009). Learning through talk: Oral language in years 4 to 8. Wellington, New Zealand: Learning Media Ltd.

Ministry of Education (2010). The literacy learning progressions: Meeting the reading and writing demands of the curriculum. Wellington, New Zealand: Learning Media Ltd.

Ministry of Education (n.d.). The learning progressions framework. Retrieved March 162020 from https://curriculumprogresstools.education.govt.nz/lpf-tool/
Ministry of Education (n.d.). The Progress and Consistency Tool (PaCT). Retrieved March 162020 from https://curriculumprogresstools.education.govt.nz/pact/learn-about-pact/

Ministry of Education (n.d.). The PACT reading framework. Retrieved March 162020 from https://curriculumprogresstools.education.govt.nz/lpfs/understanding-the-reading-framework/
Ministry of Education (n.d.). The PACT writing framework. Retrieved March 162020 from https://curriculumprogresstools.education.govt.nz/lpfs/understanding-the-writing-framework/

MINISTRY OF EDUCATION
TE TĀHUHU O TE MĀTAURANGA


[^0]:    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.

[^1]:    4 School participation rate is defined as the number of schools that participated as a percentage of the number of schools invited, including replacements.

[^2]:    5 Student participation rate is defined as the number of students assessed as a percentage of the total number of participating students who were originally selected, the number of substitute students and the number of students who were withdrawn or excluded.

[^3]:    Note: Ministry of Education July 2019 school returns for Year 8 were used for the population percentages.

    * Ethnicity is based on the Ministry of Education's prioritised ethnicity.

[^4]:    ${ }^{6}$ In Cycle 1 of NMSSA, four language modes associated with the English learning area (writing, reading, speaking and listening) were assessed. Writing was assessed in 2012, reading in 2014, listening in 2015, and viewing in 2015. Speaking and presenting were not assessed in Cycle 1.
    7 In a cloze passage, a number of words have been systematically deleted and students are required to supply the words.

[^5]:    8 IRT is an approach to constructing and scoring assessments and surveys that measure mental competencies and attitudes. IRT seeks to establish a mathematical model to describe the relationship between people (in terms of their levels of ability or the strengths of their attitude) and the probability of observing a correct answer or a particular level of response to individual questions. IRT approaches provide flexible techniques for linking assessments made up of different questions to a common reporting scale. The common scale allows the performance of students to be compared regardless of which form of the assessment they were administered.

[^6]:    9 All estimates of means and standard errors are calculated using the full sample size rather than the effective sample size defined by the design effect calculations. See Appendix 4 (page 23).

[^7]:    * New Zealand European/Pākehā

[^8]:    ${ }^{10}$ See Variance Estimation in NMSSA, at https://nmssa.otago.ac.nz/reports/Variance_Estimation_NMSSA.pdf.

[^9]:    * New Zealand European/Pākehā

[^10]:    * New Zealand European/Pākehā

[^11]:    ${ }_{16}$ Taylor Series Linearisation method

[^12]:    Figure A5.1 Overview of NMSSA assessment development process

[^13]:    ${ }^{17}$ To assist the panel members to think about their judgements, feedback was given regarding what the judgements meant in terms of the actual percentage of students at each year level achieving at different score points on each item. This was provided graphically. Panel members were able to consider this information before submitting their final judgements.

[^14]:    ${ }^{18}$ NMSSA Report 2: English:Writing 2012

[^15]:    ${ }^{19} \mathrm{https}: / /$ nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum/English/Achievement-objectives

