

e-asTTle writing: The technical report

Contents

| | |
|---|-----------|
| Foreword to the technical report | 1 |
| Introduction | 1 |
| An overview of e-asTTle writing | 1 |
| The components of e-asTTle writing | 2 |
| Part 1: The development of the e-asTTle writing components | 3 |
| 1. The prompts | 3 |
| 1.1. Rationale | 3 |
| 1.1.1. Design process | 6 |
| 2. Piloting | 7 |
| 3. The rubric | 8 |
| 4. Annotated exemplars | 9 |
| 5. The trial marking process | 10 |
| 6. The teachers' manual | 12 |
| 6.1. Structure of the manual | 12 |
| 6.1.1. Features of the manual | 13 |
| Part 2: Statistical and psychometric development | 14 |
| 7. Overview | 14 |
| 7.1. The reference sample | 15 |
| 7.1.1. Trial design | 18 |
| 7.2. Marking design | 21 |
| 7.3. Multifacet Rasch modelling | 22 |
| 7.3.1. An overview of the MFRM | 24 |
| 7.3.2. The measurement scale | 26 |
| 7.4. Fit of the multifacet Rasch model and data | 26 |
| 7.5. Reliability | 38 |
| 7.5.1. Consistent use of the rubric | 38 |
| 7.5.2. Internal consistency | 39 |
| 7.5.3. Test-retest reliability | 40 |
| 7.5.4. Reliability of model parameters | 41 |
| 7.6. Writing scores in the sample | 41 |
| 7.7. Normative information | 45 |
| 7.7.1. The statistical model | 45 |
| 7.7.2. Modelled writing scores | 47 |
| 7.7.3. Extrapolation from previously collected data | 50 |
| 7.8. Standard setting for e-asTTle writing | 50 |
| 7.8.1. Introduction | 50 |
| 7.8.2. The literacy learning progressions | 50 |
| 7.8.3. Standard-setting process | 51 |

| | | |
|-------------------|--|-----------|
| 7.8.4. | The judges | 51 |
| 7.8.5. | The response booklet | 52 |
| 7.8.6. | Organisation | 52 |
| 7.8.7. | Observations from the first part | 53 |
| 7.8.8. | The second part | 53 |
| 7.8.9. | Assigning e-asTTle curriculum level descriptors | 54 |
| 7.8.10. | Curriculum expectations | 56 |
| Part 3: | Scoring and reporting systems for the e-asTTle writing tool | 59 |
| 8. | Introduction | 59 |
| 8.1. | Rubric scores | 59 |
| 8.2. | Scale scores | 59 |
| 8.2.1. | The overall scale score | 59 |
| 8.2.2. | Calculating overall scale scores | 60 |
| 8.2.3. | Calculating scale scores for elements | 61 |
| 8.2.4. | Conversion of logit scores to scale scores | 61 |
| 8.2.5. | Calculating curriculum level scores | 61 |
| 8.3. | Normative information | 62 |
| 8.4. | Changes to e-asTTle reports | 63 |
| 8.4.1. | The Individual Learning Pathways Report | 63 |
| 8.4.2. | The Curriculum Levels Report | 63 |
| References | | 64 |

Tables

| | | |
|----------|---|----|
| Table 1 | Response rates | 16 |
| Table 2 | Achieved sample by decile | 16 |
| Table 3 | Students in the sample by year level and gender | 17 |
| Table 4 | Students in the sample by year level and ethnicity | 17 |
| Table 5 | Students in the sample by year level and decile | 18 |
| Table 6 | Number of students attempting writing prompts | 20 |
| Table 7 | Decile and participation | 21 |
| Table 8 | Multiple marking | 22 |
| Table 9 | Prompt statistics | 27 |
| Table 10 | Element statistics | 27 |
| Table 11 | Marker statistics | 28 |
| Table 12 | Threshold statistics | 29 |
| Table 13 | Agreement rate amongst markers | 38 |
| Table 14 | e-asTTle writing sample statistics by year level and gender | 43 |
| Table 15 | e-asTTle writing sample statistics by year level and ethnicity | 44 |
| Table 16 | e-asTTle writing sample statistics by decile group | 45 |
| Table 17 | e-asTTle writing model of standard deviation values | 47 |
| Table 18 | e-asTTle writing model statistics for Quarter 3, by year level and gender | 48 |

| | | |
|----------|---|----|
| Table 19 | e-asTTle writing model statistics for quarter 3, by year level and ethnicity | 49 |
| Table 20 | The LLP levels and their link to curriculum levels | 51 |
| Table 21 | Minimum scores for each curriculum level | 54 |
| Table 22 | Relationship between e-asTTle curriculum level cut-points and the LLP level from standard setting | 54 |
| Table 23 | e-asTTle writing curriculum ranges | 55 |
| Table 24 | Expected score by quarter | 57 |
| Table 25 | Scoring categories by element | 59 |
| Table 26 | The relationship between prompt parameters and modelled category thresholds for a single element | 60 |
| Table 27 | Example of prompt parameters (in logits) | 60 |
| Table 28 | Scale score to curriculum level conversion | 62 |
| Table 29 | Changes to the Individual Learning Pathways Report | 63 |
| Table 30 | Changes to the Curriculum Levels Report | 63 |
| Table 31 | Prompt numbers and names | 65 |
| Table 32 | Jumping Girl | 66 |
| Table 33 | Girl | 66 |
| Table 34 | Adult and child | 66 |
| Table 35 | Whānau and family time | 67 |
| Table 36 | Time with friends | 67 |
| Table 37 | What I did well | 67 |
| Table 38 | Caring for planet earth | 68 |
| Table 39 | The life cycle of Monarch butterflies | 68 |
| Table 40 | A frog life cycle | 68 |
| Table 41 | Music is more important than sport | 69 |
| Table 42 | It is wrong to fight | 69 |
| Table 43 | The referee is always right | 69 |
| Table 44 | A special place in the community | 70 |
| Table 45 | Good friends | 70 |
| Table 46 | A community facility | 70 |
| Table 47 | The market | 71 |
| Table 48 | Stick insect | 71 |
| Table 49 | The day things started disappearing | 71 |
| Table 50 | The bush | 72 |
| Table 51 | I heard a whisper but no one was there | 72 |

Figures

| | | |
|----------|--|----|
| Figure 1 | Development of the prompts, rubric and annotated exemplars | 5 |
| Figure 2 | Scale locations for the difference facets used in developing a model for e-asTTle writing — the “Facets Ruler” | 25 |

| | | |
|-----------|--|-------------------------------------|
| Figure 3 | Graphical representations of Element 1: Ideas | 31 |
| Figure 4 | Graphical representations of Element 2: Structure and language | 32 |
| Figure 5 | Graphical representations of Element 3: Organisation | 33 |
| Figure 6 | Graphical representations of Element 4: Vocabulary | 34 |
| Figure 7 | Graphical representations of Element 5: Sentence structure | 35 |
| Figure 8 | Graphical representations of Element 6: Punctuation | 36 |
| Figure 9 | Graphical representations of Element 7: Spelling | 37 |
| Figure 10 | Test-retest data | Error! Bookmark not defined. |
| Figure 11 | e-asTTle writing sample score distributions by year level | 42 |
| Figure 12 | Graphical representation of the relationship between the aWs scale, e-asTTle curriculum level descriptors and the LLP levels | 55 |
| Figure 13 | The Quarter 4 distribution of scale scores by year level compared to curriculum level expectations | 56 |
| Figure 14 | Scale score expectations by year level | 58 |

Appendices

| | | |
|-------------|---------------------------------|----|
| Appendix B: | Prompt numbers, names and types | 65 |
| Appendix C: | Prompt parameters | 66 |

Foreword to the technical report

Introduction

This technical report provides a detailed description of key aspects of the development of the e-asTTle writing assessment tool. It is designed to provide readers with an overview of the development process and provide insights into decisions taken along the way and the rationale for these.

A range of readers, including teachers, professional development providers and technical experts, will be interested in parts, or all, of the technical report.

The technical report should be read alongside the e-asTTle writing teachers' manual (Ministry of Education & NZCER, 2012). Although the material covered in both documents overlaps, the technical report does not replace the teachers' manual; rather, it expands the information and provides more detail.

The technical report is divided into three parts. The first part deals with the development of the writing prompts, rubric, and annotated exemplars. The second part deals with the statistical and psychometric development of the tool, including the use of multifaceted Rasch modelling to construct a measurement scale, the construction of normative information, and the standard-setting exercise used to link the scale to curriculum levels.

The last part of the report deals with aspects of the development related to the e-asTTle application itself, including the scoring mechanisms and changes to reporting.

An overview of e-asTTle writing

e-asTTle writing (revised) is an online assessment tool designed to assess students' progress in writing from Years 1–10. It represents a complete revision of the original e-asTTle writing assessment tool.

e-asTTle writing assesses students' ability to independently write continuous text across a variety of communicative purposes (describe, explain, recount, narrate, persuade). It assesses generic writing competence rather than writing specific to any learning area, and so does not assess content knowledge.

An e-asTTle writing assessment involves students writing for up to 40 minutes to compose a response to a set prompt (formerly known as a task). Teachers then use a rubric, supported by notes and exemplars, to score the writing against seven different elements of writing. The online e-asTTle application is able to convert the rubric scores to scores on an e-asTTle writing scale and subsequently to curriculum levels, and then to produce a range of reporting at the individual and group level.

As well as helping teachers monitor their students' progress, the results from e-asTTle writing will help teachers to make informed decisions about the kinds of teaching materials, methods and programmes most suitable for their students. It also provides teachers with a means of measuring progress in writing over time and against the national expectations.

The components of e-asTTle writing

The e-asTTle tool is made up of the following components:

- 20 writing assessment prompts
- a marking rubric
- structure and language notes (to assist use of the marking rubric)
- 76 annotated exemplars
- a glossary and a list of definitions.

A teachers' manual has also been developed that describes the tool and provides information and advice about its use.

Part 1: The development of the e-asTTle writing components

1. The prompts

1.1. Rationale

The development of the prompts took into account recent theorising about genre—that is, about the work texts do in society (Bazerman, 2010, 2004; Bawarshi & Reiff, 2010; Bazerman, Bonini, & Figueiredo, 2009; Chapman, 1999; Dean, 2008). Genres are part of society: they shape regularised communicative practices that bind together organisations, institutions, and activity systems. In short, it is shared understandings of genre that help us and those we communicate with to be on the same page.

Five regular ways of writing are assessed by the e-asTTle writing tool: to describe, explain, persuade, narrate, and recount. In order to align the language of the tool with curriculum documents, the word “genre” is not used. Instead, the tool uses the word “purpose”. In essence, these five purposes were chosen because they are useful across subject areas (Schleppegrell, 2004). But specifically, describing was included, even though it is not generally considered to be a distinct genre, because it is often an important component within a larger piece of text. Narrating was included for a similar reason—the ability to tell a story is useful even when writing an informational text. Recounting was included because the mastery of sequence is a fundamental skill, and explaining and persuading were included because they are of particular use in science and the social sciences. Reporting was not included because advice from the e-asTTle reference group indicated that many students responding to reporting tasks in the previous version of the tool had not had the level of factual knowledge required.

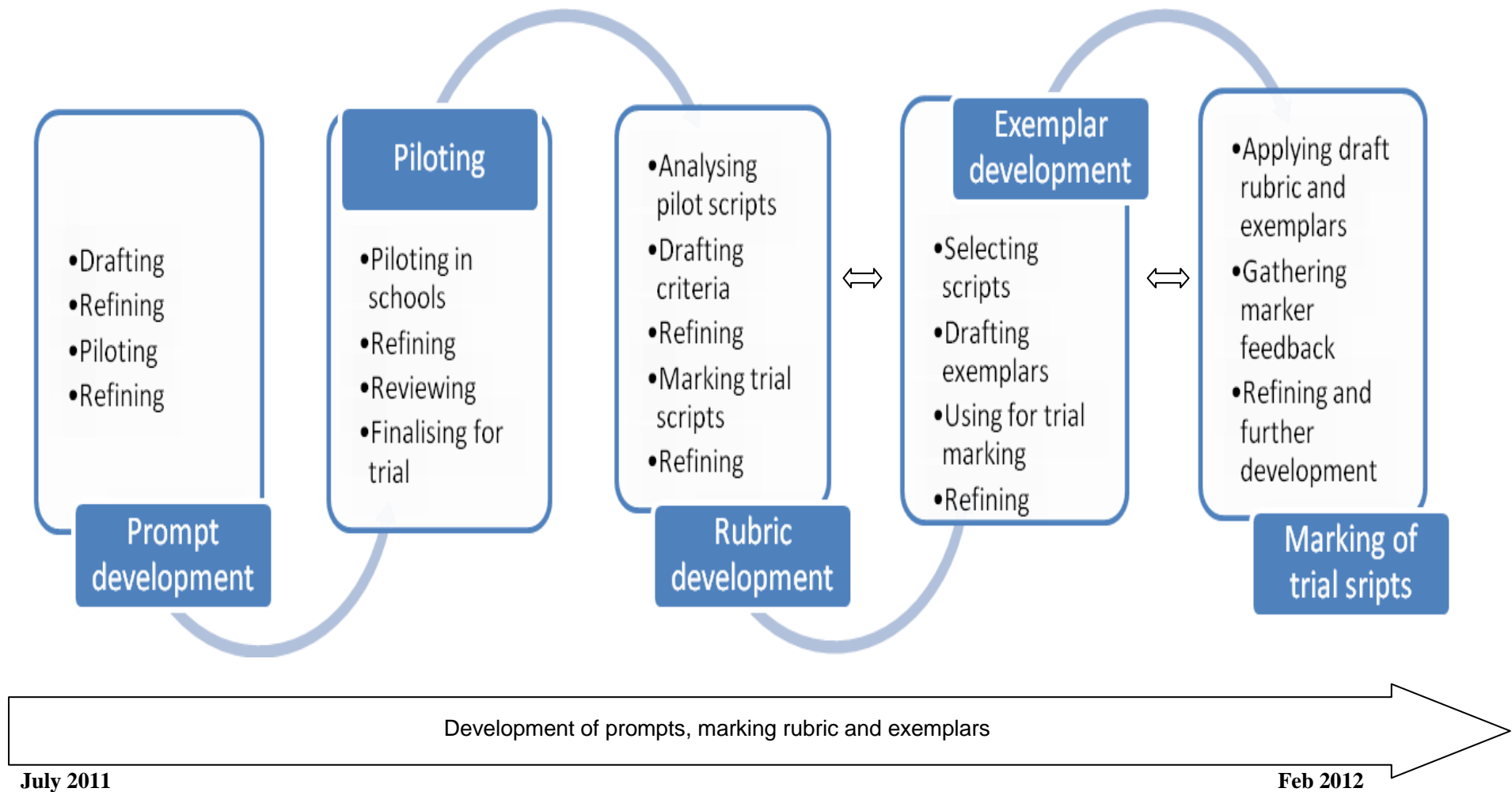
Although the tool assesses the five purposes separately in order to provide clear results, student use of multiple purposes can be recognised by the marking rubric. For example, if the specified purpose is to describe and the student also explains, only the descriptive features are scored within the structure and language element but the explanatory features can be scored within the ideas element. This is appropriate since explanation is a form of elaboration and elaboration is a focus of the ideas element. When multiple purposes are used, the teacher first identifies which purpose/s other than the specified are used, and then which element/s might be engaged to recognise their use.

e-asTTle writing prompts generate written responses that can be assessed in terms of general writing competence. Although it is acknowledged that the most authentic writing is generated within contexts rich with subject-area content knowledge (Moje, 2008; Purcell-Gates, Duke, & Martineau, 2007; Cervetti & Pearson, 2012), e-asTTle writing is not suited to assessing content

knowledge. Results are much clearer when the assessment focus is precise, and attempting to assess content knowledge as well as general writing competence would compromise precision. Accordingly, the purpose of the e-asTTle writing prompts is to generate written responses that can be assessed in terms of general writing competence only.

The process of developing the prompts extended over several months. Figure 1 below shows the overall process of developing the prompts, rubric and annotated exemplars. However, in practice, the stages of development were not clearly sequential. For example, the rubric development stage began with initial thoughts about approaches, before prompts were developed, and extended beyond the marking of trial scripts, as refinements were made in response to feedback. Similarly, early versions of annotated exemplars were drafted alongside the rubric, and refined during and following trial marking.

Figure 1 Development of the prompts, rubric and annotated exemplars



1.1.1. Design process

The prompts have been designed to generate continuous text on topics that are accessible to students, and which provide opportunities for individual interpretation. The topic outlined on the prompt is therefore intended as a springboard for writing, rather than as a tightly defined focus. The word “prompt” was chosen over “task” in order to emphasise that the tool prompts students to write, encouraging them to draw on their individual and cultural knowledge, rather than prescribes what is written.

The prompts were designed by members of the e-asTTle redevelopment team. They were reviewed internally by the wider NZCER literacy and assessment teams and were externally reviewed by researchers from the Australian Council for Educational Research (ACER) and the e-asTTle reference group.

Initially, two sets of prompts were developed. The larger set assessed writing of continuous text, and it was prompts from this set that later went on to be trialled and published. The smaller set assessed writing of several shorter pieces of text that together constituted text similar to that generated by the continuous text prompts. It was decided to pilot these because, in a previous project, it was found that some older students wrote very little. Breaking up the prompt into manageable sections with the students responding with several pieces of shorter text rather than one continuous one was seen as a possible way of scaffolding these students to write more. With sufficient text written, meaningful scores could be given.

It was decided early on in the piloting process that the second set of prompts was not necessary. The decision was made for two reasons:

1. The continuous text prompts were generating writing of sufficient length from a diverse range of students.
2. The shorter text prompts appeared to constrain all but the very weakest writers.

The prompts all have the same format. The opening section states the topic and purpose. The next section gives students reminders about their text as a whole—its structure, for example. The final section gives reminders at the sentence and word level—word choice and editing, for example. The recount instructions are written in slightly simplified language because of the likelihood that they will be used by younger students. Teachers are given instructions about the need to make sure all students understand what is expected of them, and they are encouraged to read the prompt to students if necessary. This approach was preferred over student instructions written in language so simple that their meaning was limited.

All prompts are essentially new but a small number have the same topics as tasks from the previous version of e-asTTle writing. These topics were retained because it was judged that they would continue to engage students.

2. Piloting

Once a pool of prompts across each of the five purposes had been developed, they were piloted in schools. Thirty prompts were piloted across a range of low-, middle- and high-decile schools and at year levels 1–10. Schools were approached and asked if they were willing to take part in the pilot and, if so, for a researcher to visit the school and work with a class-sized group of students. Permission was also sought for the resulting student work to be used as exemplar material. Each group of students was given a mixed range of prompts to pilot.

The pilot phase was designed to provide feedback on the suitability of the prompts; for example, whether students had difficulty accessing the topic or task, whether the instructions were clear, and whether the prompts elicited a range of writing performance. Feedback was gathered through a combination of methods: researcher observation, feedback from teachers and students, and analysis of the writing produced in response to each prompt. The feedback was used to answer the following questions:

- Did the instructions on the prompt work well for all students?
- Which prompts for each purpose worked best?
- Which prompts worked well for all students from Year 1 to 10?
- Which prompts were better suited to particular age ranges?
- What was the minimum level of support necessary to elicit appropriate writing under test conditions?
- Did the prompts elicit a representative range of student writing?

Feedback from the pilot phase informed both the decision about which prompts would go forward for national trialling and the review and refinement of these prompts before trialling began. The choice and refinement of prompts was an iterative process, informed by the need for a number of prompts within each purpose. For example, prompts were reviewed to make sure that wording was consistent and that there was a range of prompts within each purpose that would work for students from Years 1 to 10. As a result of the pilot phase, 25 prompts were selected to go forward for national trialling¹. The feedback was also used to help inform the development of a set of concise, clear instructions for teacher administrators of the national trial of the prompts.

A further purpose of the pilot process was to provide a bank of student writing samples to inform the development of the assessment criteria within the marking rubric and provide exemplar material to assist marking of the national trial scripts. During the pilot, it became clear that examples of writing at the upper ranges of skill development within each element of writing were

¹ NZCER was contracted to deliver 15–20 prompts but 25 prompts were trialled to ensure the most suitable were identified. After the trial, five prompts across the range of purposes were identified as working less well. Four were consequently removed from the pool before the marking process began, with one left in to act as a buffer. Although all 21 prompts came through the marking process well, the buffer prompt was still the least well performing prompt of its purpose. It was, therefore, not uploaded, taking the total number of uploaded prompts to 20. The data relating to this prompt has, nonetheless, been reported on in Parts 2 and 3.

relatively scarce. Further piloting with Year 10 students and at high-decile schools was needed to source these scripts. Given the tight time frame for the pilot process, further scripts at these levels were also sourced from the national trial.

3. The rubric

Studies into the large-scale assessment of writing (Humphry & Heldsinger, 2009; Andrich, 2005) have made the significant conclusion that framework documents such as literacy learning progressions and national standards are useful maps of curriculum and learning progression, and provide teachers with useful “big picture” descriptions of growth in specified learning areas across the years of schooling, but they do not constitute a measurement scale and the numbers assigned to levels or bands should not be treated as if they have measurement properties.

This conclusion was based on data derived from marking rubrics in which the marking criteria and categories were structured to directly match the framework levels and substrands. The data derived from the rubrics were of limited use because assessments made in terms of the levels of a framework classified student performance too crudely to be of much use to teachers in either planning for teaching or reporting performance.

These research findings informed the way NZCER approached the redevelopment of e-asTTle writing. In this approach, the status of the curriculum framework is not contested. It remains the conceptual framework for assessment development and it provides the framework for reporting the results of the assessment. What is different is that the marking rubric is developed to match the tasks and to capture the categories of performance that can be observed in student writing.

Based on this approach, the principles guiding the development of the revised e-asTTle marking rubric were as follows. The rubric would:

- link closely to the language and descriptions of writing in curriculum documentation; for example, the New Zealand Curriculum, the literacy learning progressions, and the national standards
- specifically match the new writing prompts
- be developed from careful observation of student writing in response to these prompts
- categorise different elements of writing (e.g., ideas, sentence structure, organisation, and spelling) across the range of performance observable in students’ writing
- use student writing to inform the selection and description of each scoring category
- capture fine-grained differences that can be observed in student writing, allowing precise marking
- allow teachers to discriminate reliably between student performances across and between year levels 1–10
- provide detailed and reliable information on specific aspects of writing-to-communicate.

The initial draft of the rubric was developed from close analysis of student writing gathered during the pilot phase of the project. As student scripts produced in response to each new writing

prompt were received, they were read carefully and analysed to identify common themes or traits. From this first analysis, researchers made tentative decisions about possible elements and scoring ranges for the rubric. These decisions were informed not only by students' writing, but also by knowledge of general writing research, experience in developing similar writing tools, advice from reference group members, and guidance provided by curriculum documents.

Once initial decisions about scoring elements and ranges were made, researchers worked systematically, prompt by prompt, to rank the scripts from weakest to strongest and to enter brief descriptions of the writing features they observed to the rubric. Over time, patterns of features emerged. For example, it was found that the sentences in the lowest-ranked scripts consisted of fragments and phrases, and often had missing words, and that the next stage consisted of short simple, compound, and basic complex sentences. From this analysis of writing for each prompt and purpose, the elements of the scoring rubric were populated with descriptions of features observed in student writing. During these initial stages, combinations of writing elements and criteria were trialled and refined. Also at this stage, it was found that the structure and language element was proving the most difficult to score. Consequently, additional notes (the "Structure and Language Notes") were developed for each purpose to provide more information for markers.

The draft rubric was subsequently used to mark approximately 5000 scripts produced from the trial. This marking process made significant contributions to the refinement of the rubric, through feedback from the 26 trained markers and from the teachers involved in the trial. For example, most of the "Structure and Language Notes" for each purpose were substantially changed, to assist marking of this element of writing. The trial also produced scripts that became important additions to the pool of annotated exemplars, and the rubric was subsequently annotated with links to exemplars that illustrate each category score. The final version of the rubric has deliberately been kept clear and clean, to enable precise, reliable, consistent and robust marking.

4. Annotated exemplars

The rubric does not stand alone, but is supported by a range of meticulously annotated exemplars. These are typical examples of students' writing in response to e-asTTle writing prompts. Each exemplar shows how a student's writing has been marked using the rubric, with annotations explaining the marking decisions. Using the exemplars during the marking process enables teachers to apply the rubric consistently. The rubric and annotated exemplars illustrate a developmental pathway that will allow teachers to support students and plan future teaching programmes.

The annotated exemplars were drafted in tandem with the rubric, so that sets of exemplars would be available during the training process for markers of the trial scripts. Those sets included a booklet of draft exemplars drawn from writing across the range of purposes and prompts that enabled markers to become familiar with the various elements of writing and categories of performance described by the rubric. Exemplar scripts for this booklet were selected to illustrate the range and progression of development within each element of writing. In addition, scripts

were included that illustrated problematic aspects of marking (e.g., difficult-to-read handwriting or writing that appeared to be “off topic”).

A draft exemplar booklet was also prepared for the group of prompts within each purpose: to explain, describe, narrate, recount and persuade. These purpose-specific exemplars enable markers to see how each element and category of performance relates to writing developed for that purpose. As part of the marking of the trial scripts, each exemplar was discussed and refinements made in response to feedback.

Following the trial marking process, the exemplars continued to be refined alongside the rubric, to ensure consistency of language and approach. Care was taken to provide illustrations from the script to support the scoring. Further exemplars were written to ensure that there were at least three exemplars for each writing prompt, covering a range of skill levels, and at least two exemplars for every rubric category. All exemplars were moderated by at least two members of the writing team.

5. The trial marking process

Following the pilot phase of the project, the prompts were refined and sent out to schools for trialling. While the trialling was being completed, arrangements were put in place for the marking of trial scripts. A robust marking design was completed, which would involve the recruitment of 26 experienced teacher markers. An education recruitment company (Education Personnel Ltd) was approached to help source possible markers. The final team of markers was selected on the basis of each marker’s experience in teaching and assessing writing, as well as their previous involvement in similar projects.

A suitable venue for the marking was booked, and preparations were made to train markers in applying the rubric. Marker training took place over two days, immediately before the marking of the trial scripts. The training was led by a member of the e-asTTle writing development team, who had extensive experience of delivering similar training through her role with the Australian Council for Educational Research (ACER). The decision to use her expertise to develop and lead the training allowed other members of the team to concentrate on developing exemplars and other essential support material, and on refining the rubric. During the marker training, these team members were on hand to support markers, respond to feedback, and help the group to develop consistent understanding of the marking process. The training was also attended by 16 observers—the vast majority of these were professional learning and development facilitators.

The first day of training covered the following aspects:

- introduction and background to the revision of e-asTTle writing
- overview of the process to this point
- unpacking the rubric—focusing on each of the seven elements of writing in turn
- an overview of the marking process and psychometric scaling

- the role of the lead markers as co-ordinators of the group in general and link marking with other groups
- applying the rubric to sample scripts.

Before the training, a range of draft exemplars had been developed and moderated by members of the development team. Scripts from these exemplars were assembled into booklets for the markers (see “Annotated exemplars” above). These booklets were later refined and made available on the tool as the “generic” exemplar booklet (i.e., scripts from across the range of prompts) and sets of exemplars for the prompts within each writing purpose (i.e., scripts from within a specific purpose). However, at this stage, the booklets were given to markers without scores for each element. Instead, space was provided alongside each script for trainee markers to enter scores. The generic booklets were used by markers on the first, more general, training day. The purpose-specific booklets were used on the more focused second day.

During the first day of training, the whole group worked with the draft rubric and the generic exemplar booklet. The trainer began by providing background information about the various parts of the rubric (the elements, the skill focus, the definition, the descriptors and so on). Trainees then worked through each element in turn (apart from “structure and language”, which was considered in depth on day two, being purpose-specific). For each element, trainees explored the category descriptors, notes, and pathway of skill development. They then applied the rubric to selected scripts in the generic exemplar booklet, discussed the script and possible scores as a large group, and compared their scores with those given on the annotations. This process enabled the building of common understanding about the focus of each element and the range of development within the element. As they worked through the exemplars, trainees were given copies of the appropriate annotations to add to their exemplar booklets and refer to during subsequent marking.

On the second day, trainee markers were given sets of “Structure and Language Notes” and divided into groups, each focusing on a separate writing purpose (i.e., describe, explain, recount, persuade, and narrate). A member of the e-asTTle writing development team was also assigned to work with each group, to act as facilitator and assist with purpose-specific training. One member of each group was assigned as a “lead marker”. Lead markers had been selected before the training on the basis of their prior experience with similar marking projects, and on the recommendation of Education Personnel Ltd. The lead marker role involved general co-ordination of the group, and link marking with other groups.

Each group explored the “Structure and Language Notes” relevant to the purpose it would be marking. The group then practised marking sample scripts against the structure and language element, with guidance from a member of the e-asTTle writing development team. They then progressed to marking scripts from their purpose-specific exemplar booklets against the full range of elements in the rubric. Some scripts were marked individually, and some were marked as a group. Scores were discussed and compared with the annotations for each exemplar, to clarify understanding of marking criteria.

Later on the second day, markers began working independently to score the trial scripts. Marking continued for several days. At the beginning of each day, each group of markers was given a script to discuss, score, and moderate as a group, to reorient themselves and to ensure consistent application of rubric scores. Scripts for this purpose were selected in advance, from the pilot scripts.

The markers worked individually to score scripts, but were based in groups according to the purpose they were marking. This allowed markers to discuss scripts with each other as necessary. Breakout rooms were available, to accommodate those who preferred to work more quietly on their own.

Members of the development team were available throughout the marking process, to answer questions or assist with problem scripts. They also consulted with the psychometric team on a regular basis, to identify any problematic marking; for example, marking that was too harsh, lenient or inconsistent, or that displayed unusual patterns (such as consistently scoring scripts at the same category across all elements of the rubric—all R4s or all R3s and so on). Feedback was provided to individual markers or to the whole group of markers as appropriate. Issues discussed included the rate of marking (some markers worked much more slowly than others) and clarifications about how to determine whether writing was “on topic” for the ideas element.

Throughout the marking process, markers were invited to provide feedback on their experience of using the rubric to score scripts. Each group was given an exercise book in which to record feedback or concerns during the marking process. In addition, each marker completed an individual feedback sheet at the end of the marking process. This feedback was collected, analysed and subsequently used to make refinements to the rubric and exemplars. A summary of the feedback was also provided to the reference group.

6. The teachers’ manual

6.1. Structure of the manual

It was decided that the manual should follow a format similar to that of the Progressive Achievement Tests (PATs) because it was a format teachers were familiar with, and one which had generally been received favourably. The one area that was known to be of concern to teachers was the length. Because of this, a commitment was made to limit the manual to about 40 pages.

Because more general e-asTTle manuals already existed, it was decided that this manual would not attempt to replace any information; instead, its purpose would be to extend existing information by focusing on the revised writing tool.

Part A of the manual describes the tool as a whole—the thinking behind the tool, how it links to the curriculum, what it does and doesn’t assess, its components, administration and marking, and interpreting test scores. Part B describes the development of the e-asTTle writing measurement scale and the compilation of reference data.

6.1.1. Features of the manual

The manual was designed to enable markers to understand what the tool measures, and to understand how to administer, mark, and interpret scores proficiently. It includes the following features:

- It gives practical advice. For example, in section 3.2.3, users are advised that “To score the students’ completed writing, you will need ...”. However, it also outlines the thinking behind the tool’s development, when relevant to its optimum use. For example, “The term ‘prompt’ emphasizes the role of ‘prompting’ rather than prescribing writing. This emphasis encourages students to draw on their individual and cultural knowledge to interpret the writing topic.”
- The passive voice has generally been used, except in sections where declaratives were necessary for instructions or advice. The benefits of the more personal active voice were discussed; however, the passive voice was considered more appropriate, given the manual is an official document. The glossary and list of definitions and the rubric are included as appendices. This was done because it was envisaged that many teachers would print a hard copy of the manual, and these two components are essential to the assessment process, regardless of which prompt is being marked.
- Screen shots, tables, and figures (some with annotations) have been included to give teachers additional support. The diagrams on pages 7 and 8 support teachers to appreciate the relationship of the tool to the wider curriculum. This relationship needs to be understood if the tool is to be used effectively.
- Plain language has been used wherever possible. Where technical language is necessary, terms that may not be widely understood have been explained.
- Particular emphasis has been given to explaining the marking process and the tools used to mark student writing. This is because this process is crucial to accurate results and to using the tool formatively, but is often given much less attention by users than interpreting test scores.
- Some misconceptions about the use of the tool have been clarified. For example, in section 2.3.2, an explanation is given of why results from teacher-developed prompts should not be entered into the e-asTTle application.

Part 2: Statistical and psychometric development

7. Overview

The development of the e-asTTle writing tool involved:

- a sample of 4755 students from 160 schools
- 21 writing prompts representing 5 writing purposes:
 - describe
 - recount
 - explain
 - persuade
 - narrate
- 26 markers
- 7 assessed elements of writing:
 - ideas (with 6 rubric categories)
 - structure and language (with 6 rubric categories)
 - organisation (with 7 rubric categories)
 - vocabulary (with 6 rubric categories)
 - sentence structure (with 6 rubric categories)
 - punctuation (with 7 rubric categories)
 - spelling (with 6 rubric categories)

Part 2 of the e-asTTle technical manual describes the statistical and psychometric aspects of the development of the e-asTTle writing tool. The sections are broadly organised according to their order of implementation during the tool development process.

The selection of the sample was one of the early pieces of quantitative work undertaken. The aim was to select a sample representative of New Zealand's Year 1 to Year 10 student population. The constraints on the sample as well as the response rates of schools and the demographic distributions of students in the sample are provided in section 7.1. Section 7.1.1 describes how writing prompts were assigned to sample schools according to a design which ensured that students wrote to age-appropriate prompts and that responses to each prompt could be linked to responses to all other prompts.

A marking design was generated to ensure that the contribution of individual markers to the marking process could be compared to all other markers. The details of this design are described in section 7.2.

Following the completion of the marking process, the data was fitted to a multifacet Rasch model. This ultimately enabled the generation of a scale that measures writing proficiency—the

mechanism at the core of the e-asTTle tool. Together with an overview of multifacet Rasch models, this process is described in section 7.3. Section 7.4 describes the fit of the data to this multifacet Rasch model as well as providing the parameters specifying the model.

For the reporting facility of the e-asTTle tool to function properly, a substantial amount of normative information needed to be generated during development. The scope of the current work meant that some of this information was unable to be directly derived from the collected data described above. Instead, a statistical model based on the collected data was used to produce robust estimates of some of the required norms. The model is described in section 7.7.1. The remaining norms were extrapolated from data collected in the previous version of e-asTTle writing. This extrapolation is described in section 7.7.3.

The final section of part 2 of the technical manual describes the standard-setting exercise used to link the e-asTTle writing scale with the descriptions of performance outlined in the literacy learning progressions and define the curriculum level reporting.

7.1. The reference sample

The e-asTTle writing reference sample was designed to:

- be representative of New Zealand’s Year 1 to Year 10 student population
- minimise school burden by selecting 30 students from, at most, two year-levels per school.

The sample was drawn as a stratified two-stage random sample of students within schools. The two-stage nature of the sample minimised both cost and school burden across New Zealand. Given the target audience of the e-asTTle writing tool, only English-medium schools were included in the sample frame. In addition, special schools and very small schools were removed from the sample frame. A special school is defined as one that supports the education of students with behavioural, sensory, cognitive or physical needs requiring extra assistance. A very small school is defined as one with less than 15 students in the target year level.

The sampling of students within schools was the responsibility of the schools themselves. Schools were asked to systematically select students at the prescribed year level from their rolls. NZCER provided a methodology for schools to follow and offered to carry out the sampling for schools if it was required.

The sampling frame was stratified by year level, school decile and school roll. Together these variables tend to act as proxies for other factors (such as ethnicity and socioeconomic status) that are important from the perspective of educational statistics. The years were grouped as Years 1 and 2, Years 2 and 3, Years 4 and 5, Years 5 and 6, Years 7 and 8 and finally Years 9 and 10. The three decile groups consisted of deciles 1, 2 and 3, deciles 4, 5, 6 and 7 and deciles 8, 9 and 10. The school-size groups were determined by roll size at the relevant year levels and were nominated as small (16–45 students), medium (46–68 students) or large (more than 68 students). Within each decile group by size-group stratum, schools were selected randomly to form the sample.

Table 1 displays the overall response rates for schools and those rates broken down by decile grouping. Schools that were decile 1, 2 or 3 were more reluctant to participate in the development of the e-asTTle writing tool.

Table 1 **Response rates**

| | | Counts | | | Proportion of all schools approached | | |
|-----------------|--------------|--------------|---------------------|------------------|--------------------------------------|-------------------------|-------------|
| | | Participated | Did not participate | Total approached | Participated (%) | Did not participate (%) | Total (%) |
| Deciles 1,2,3 | Original | 22 | 26 | 48 | 27 | 32 | 59 |
| | Replacement | 23 | 10 | 33 | 28 | 12 | 41 |
| | Total | 45 | 36 | 81 | 56 | 44 | 100 |
| Deciles 4,5,6,7 | Original | 43 | 21 | 64 | 50 | 24 | 74 |
| | Replacement | 19 | 3 | 22 | 22 | 3 | 26 |
| | Total | 62 | 24 | 86 | 72 | 28 | 100 |
| Deciles 8,9,10 | Original | 33 | 21 | 54 | 42 | 27 | 69 |
| | Replacement | 20 | 4 | 24 | 26 | 5 | 31 |
| | Total | 53 | 25 | 78 | 68 | 32 | 100 |
| All deciles | Original | 98 | 68 | 166 | 40 | 28 | 68 |
| | Replacement | 62 | 17 | 79 | 25 | 7 | 32 |
| | Total | 160 | 85 | 245 | 65 | 35 | 1.00 |

Percentages that do not add up to 100 are caused by rounding errors

Table 2 describes the achieved sample of schools and students by school decile. Schools that declined to participate were replaced by schools with similar demographics.

Table 2 **Achieved sample by decile**

| Decile | Participating schools | Participating students |
|--------------|-----------------------|------------------------|
| 1 | 10 | 277 |
| 2 | 16 | 469 |
| 3 | 19 | 556 |
| 4 | 9 | 260 |
| 5 | 15 | 450 |
| 6 | 15 | 466 |
| 7 | 23 | 687 |
| 8 | 10 | 306 |
| 9 | 19 | 568 |
| 10 | 24 | 716 |
| Total | 160 | 4755 |

Table 3, Table 4 and Table 5 show the student-level demographics of the achieved e-asTTle writing sample. Table 3 shows the sample reported by the year level and gender of the students. There are slightly more boys than girls overall, although this is not true at all year levels.

Table 3 **Students in the sample by year level and gender**

| Year level | Girls | Boys | Missing gender | Total |
|--------------------|-------------|-------------|----------------|-------------|
| 1 | 219 | 229 | 2 | 450 |
| 2 | 259 | 213 | 1 | 473 |
| 3 | 230 | 246 | 1 | 477 |
| 4 | 217 | 258 | 1 | 476 |
| 5 | 249 | 233 | 0 | 482 |
| 6 | 216 | 214 | 0 | 430 |
| 7 | 251 | 235 | 0 | 486 |
| 8 | 230 | 254 | 0 | 484 |
| 9 | 239 | 260 | 1 | 500 |
| 10 | 229 | 263 | 0 | 492 |
| Missing year level | 0 | 2 | 3 | 5 |
| Total | 2339 | 2407 | 9 | 4755 |

Table 4 shows the sample reported by the year level and ethnicity of the students. Note that the students could identify with more than one ethnic group and therefore there were more identifications than there were students. Students were more likely to identify with more than one ethnic group in higher year levels.

Table 4 **Students in the sample by year level and ethnicity**

| Year level | NZ European | Māori | Pasifika | Asian | Other | Total identifications | Total students |
|--------------|-------------|-------------|------------|------------|------------|-----------------------|----------------|
| 1 | 251 | 115 | 32 | 24 | 28 | 450 | 450 |
| 2 | 257 | 130 | 47 | 20 | 29 | 483 | 473 |
| 3 | 280 | 115 | 39 | 8 | 34 | 476 | 477 |
| 4 | 277 | 91 | 61 | 21 | 64 | 514 | 476 |
| 5 | 287 | 82 | 61 | 23 | 55 | 508 | 482 |
| 6 | 292 | 55 | 38 | 15 | 51 | 451 | 430 |
| 7 | 306 | 101 | 43 | 27 | 46 | 523 | 486 |
| 8 | 321 | 84 | 43 | 21 | 51 | 520 | 484 |
| 9 | 342 | 103 | 37 | 41 | 48 | 571 | 500 |
| 10 | 295 | 126 | 52 | 37 | 52 | 562 | 492 |
| Total | 4694 | 1468 | 693 | 398 | 734 | 5058 | 4750 |

Five students were missing year-level data and 17 were missing ethnicity data. The missing ethnicity data accounts for fewer total identifications in Year 3 than total students.

Table 5 summarises the sample by the year level of the students and the decile group of the schools they attend. There are relatively fewer students from lower decile schools overall and for most individual year levels, reflecting the number of students attending these schools at a national level.

Table 5 **Students in the sample by year level and decile**

| Year level | Deciles 1–3 | Deciles 4–7 | Deciles 8–10 |
|--------------------|-------------|-------------|--------------|
| 1 | 153 | 158 | 139 |
| 2 | 177 | 155 | 141 |
| 3 | 182 | 160 | 135 |
| 4 | 137 | 159 | 180 |
| 5 | 146 | 155 | 181 |
| 6 | 107 | 139 | 184 |
| 7 | 105 | 201 | 180 |
| 8 | 107 | 197 | 180 |
| 9 | 87 | 278 | 135 |
| 10 | 98 | 260 | 134 |
| Missing year level | 3 | 1 | 1 |
| Total | 1302 | 1863 | 1590 |

7.1.1. Trial design

For the writing scores of any two students, *A* and *B*, to be able to be compared, the students needed to be “linked”. This means that either students *A* and *B* have written to the same prompt, or that student *A* has written to the same prompt as a student already linked to student *B*.

The e-asTTle tool includes 21 prompts. Therefore, in order to link all students, some students had to write to two prompts. For any such pair of prompts (both attempted by the same group of students), some of the students wrote to one of the prompts first, while the remaining students wrote to the other prompt first. This enabled any possible order effect to be detected and ameliorated. To minimise student burden, schools in which students were requested to write to two prompts were, as much as possible, selected with higher deciles.

Table 6 shows the number of students who wrote to a given pair of prompts and the order in which they attempted those prompts. For example, the first entry in the second row (with column labelled 1), indicates that 29 students attempted prompt 2 and later attempted prompt 1. The first entry in the first row (with column labelled 2), indicates that 30 students attempted prompt 1 and later attempted prompt 2. Adding these two numbers together shows that 59 students attempted

prompts 1 and 2. Cells where the “first prompt” entry is the same as the “second prompt” entry are blank.

The column labelled “one prompt only” indicates that 96 students were requested to attempt only prompt 1, that 144 students were requested to attempt only prompt 2 and so on. The column labelled “total responses” indicates the total number of students who attempted each prompt. Table 31 provides the correspondence between prompt numbers and names.

Table 6 Number of students attempting writing prompts

| First prompt | Second prompt | | | | | | | | | | | | | | | | | | | | | One prompt only | Total responses |
|--------------|---------------|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----------------|-----------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | | |
| 1 | | 30 | | | | | | 25 | | | | | | | 30 | | | 30 | 30 | | | 96 | 382 |
| 2 | 29 | | | | | | | | | | | | | | | | | | | 25 | 30 | 144 | 374 |
| 3 | | | | | | | 30 | | | | | | | | | 30 | | | | | 30 | 152 | 327 |
| 4 | | | | | 68 | | | | | 48 | | | | | 30 | | | | | | | 128 | 362 |
| 5 | | | 28 | | | | | | | | | | | 26 | 28 | | | | | | | 93 | 303 |
| 6 | | | | | | | | | | | | 29 | | | | | | | 30 | | | 215 | 448 |
| 7 | | | 27 | | | | | 60 | | | | | | | | | | | | | | 203 | 403 |
| 8 | 28 | | | | | | 23 | | | | | | | | | 30 | | | 24 | | 28 | 76 | 414 |
| 9 | | | | | | 58 | | | | | | | | | | | | | | | 27 | 92 | 343 |
| 10 | | | | 8 | | | | | | | | | | | | | | 27 | 30 | | 30 | 113 | 403 |
| 11 | | | | | | 30 | | | | | | 29 | | | 60 | 22 | 23 | 58 | | | | 61 | 313 |
| 12 | | 59 | | | | 58 | | | | | | | | | 29 | | | | 30 | | | 90 | 352 |
| 13 | 30 | 31 | | | | 28 | | | | | | | | | | | 29 | | | 30 | | 55 | 345 |
| 14 | | | | | 30 | | | | | 60 | | | | | | | | | 27 | | | 148 | 374 |
| 15 | 54 | | | 8 | | | | | | | | | | | | | 27 | | | | | 130 | 408 |
| 16 | | 26 | | | | | | 29 | 20 | | | 30 | | | | | | | | | 24 | 7 | 282 |
| 17 | | | 30 | | | | | 89 | | 30 | 30 | 30 | 29 | | | | | | | 37 | | 3 | 356 |
| 18 | | | | | | | 30 | | | 28 | | | | 28 | | | | | | | | 198 | 391 |
| 19 | | | | | | | | 31 | 56 | | | | | | | | | | | | | 8 | 297 |
| 20 | | | | | 30 | | 30 | 31 | | | | 56 | | 55 | | | | | | | | 11 | 305 |
| 21 | | | | | | | | | 30 | 29 | | | 25 | | | 26 | | | | | | 13 | 292 |

Table 7 shows that, overall, it was more likely for students at a high decile school to be writing to two prompts than for those at a low decile school.

Table 7 **Decile and participation**

| Decile | Participating schools | 1 prompt (participating schools) | 2 prompts (participating schools) |
|--------------|-----------------------|----------------------------------|-----------------------------------|
| 1 | 10 | 8 | 2 |
| 2 | 16 | 8 | 8 |
| 3 | 19 | 12 | 7 |
| 4 | 9 | 2 | 7 |
| 5 | 15 | 5 | 10 |
| 6 | 15 | 5 | 10 |
| 7 | 23 | 10 | 13 |
| 8 | 10 | 0 | 10 |
| 9 | 19 | 8 | 11 |
| 10 | 24 | 7 | 17 |
| Total | 160 | 65 | 95 |

7.2. Marking design

The marking exercise used 26 markers to score student responses. For the harshness of any two markers, *A* and *B*, to be able to be compared, the markers need to be “linked” in a way analogous to the student linking described in section 7.1.1.

This means that either some of the written student responses marked by markers *A* and *B* are the same, or that marker *A* has marked some of the same written student responses as a marker already linked to marker *B*. This design also allows the investigation of consistency of practice amongst the markers using inter-rater reliability statistics (see section 7.5).

To facilitate the marking process, markers were grouped according to the purpose of the writing prompts. Each group of markers was assigned to no more than four prompts, with the exception of “link” markers who were assigned to two groups.

Table 8 shows an example of how multiple marking was used to ensure the appropriate linkage between various parts of the e-asTTle writing dataset. It depicts the linkage between and within prompt groups—in this case prompt group 1 (consisting of prompts 1, 2 and 3) and prompt group 2 (consisting of prompts 4, 17 and 18)—both “describe” prompt groups. Markers 26, 1, 2 and 3 (indicated in Table 8 respectively by M26, M01, M02 and M03) marked the prompts of group 1. Markers 3, 4, 5, 6, and 7 (indicated in Table 8 respectively by M03, M04, M05, M06 and M07) marked the prompts of group 2. Marker 3 ensured the link between the prompts of groups 1 and 2 because marker 3 marked prompts of both groups. Each cell in Table 8 indicates a set of 10 scripts. For example, markers 26 and 1 both marked the set of scripts labelled B01(1)—this established a link between them for Prompt 1. They also both marked the scripts labelled B05(1)

(as did markers 2 and 3). This strengthened their link for Prompt 1. Similarly markers 26 and 1 were linked by scripts B01(2) and B05(2) for Prompt 2.

Table 8 **Multiple marking**

| Prompt group 1—describe | | | | | | | | | | | |
|-------------------------|--------|--------|--------|----------|--------|--------|--------|----------|--------|--------|--------|
| Prompt 1 | | | | Prompt 2 | | | | Prompt 3 | | | |
| M26 | M01 | M02 | M03 | M26 | M01 | M02 | M03 | M26 | M01 | M02 | M03 |
| B01(1) | B01(1) | | | B01(2) | B01(2) | | | B01(3) | B01(3) | | |
| | B02(1) | B02(1) | | | B02(2) | B02(2) | | | B02(3) | B02(3) | |
| | | B03(1) | B03(1) | | | B03(2) | B03(2) | | | B03(3) | B03(3) |
| B04(1) | | | B04(1) | B04(2) | | | B04(2) | B04(3) | | | B04(3) |
| B05(1) | B05(1) | B05(1) | B05(1) | B05(2) | B05(2) | B05(2) | B05(2) | B05(3) | B05(3) | B05(3) | B05(3) |

| Prompt type 2—describe | | | | | | | | | | | | | | |
|------------------------|--------|--------|--------|--------|-----------|---------|---------|---------|---------|-----------|---------|---------|---------|---------|
| Prompt 4 | | | | | Prompt 17 | | | | | Prompt 18 | | | | |
| M03 | M04 | M05 | M06 | M07 | M03 | M04 | M05 | M06 | M07 | M03 | M04 | M05 | M06 | M07 |
| B01(4) | B01(4) | | | | B01(17) | B01(17) | | | | B01(18) | B01(18) | | | |
| | B02(4) | B02(4) | | | | B02(17) | B02(17) | | | | B02(18) | B02(18) | | |
| | | B03(4) | B03(4) | | | | B03(17) | B03(17) | | | | B03(18) | B03(18) | |
| | | | B04(4) | B04(4) | | | | B04(17) | B04(17) | | | | B04(18) | B04(18) |
| B05(4) | | | | B05(4) | B05(17) | | | | B05(17) | B05(18) | | | | B05(18) |
| B06(4) | B06(4) | B06(4) | B06(4) | B06(4) | B06(17) | B06(17) | B06(17) | B06(17) | B06(17) | B06(18) | B06(18) | B06(18) | B06(18) | B06(18) |

7.3. Multifacet Rasch modelling

The e-asTTle writing scale is based on an extension of the widely used Rasch measurement model (RMM). The RMM is a mathematical model that can be used to transform ordinal observations (such as rubric scores) into linear measures (Wright & Masters, 1982). A simple RMM with a “pass or fail” rubric assumes:

- *unidimensionality*—the ability of test takers and the difficulty of test items can be measured on a single scale
- *local independence*—the success of a test taker of given ability on any item is independent of their success on any other item, and the success of a test taker on an item of given difficulty is independent of the success of any other test taker
- *the logistic item response function*—the probability of success of a test taker of a given ability on a test item of given difficulty is a function of the difference between the ability and difficulty measures.

RMMs with more complicated rubrics make similar assumptions (unidimensionality, local independence, the logistic item response function). These are described in Wright and Masters (1982).

The multifacet Rasch model (MFRM) extends the RMM by taking into account additional “facets” besides student proficiency and item difficulty that might be associated with the measurement process. MFRMs are described in detail in Linacre (1994). In the context of the e-asTTle writing assessment, these facets include marker severity and the difficulty of the prompt.

To develop the e-asTTle writing scale, a multifacet Rasch model was constructed that included:

- student writing proficiency
- the difficulty of the prompts to which the students were writing
- the difficulty of the elements against which the students’ written responses were being judged (ideas, structure and language, organisation, vocabulary, sentence structure, punctuation, spelling)
- the harshness of the markers judging the students’ written responses
- the thresholds or barriers to being observed in a scoring category for an element, relative to the scoring category below.

The MFRM also makes assumptions similar to those made by the RMM (unidimensionality, local independence and the logistic item response function) that incorporate the additional facets. Statistical and graphical fit indicators were used to study the extent to which prompts, markers, students and marking rubrics fit the MFRM.

In order to construct the measurement scale, student responses to 21 writing prompts (see Table 31) were collected in a national trial involving 4755 students from Years 1 to 10. The students involved were selected using a random sampling methodology, which is described in section 7.1. Care was taken so that all markers and prompts could be linked across the students involved. This meant that many of the students completed two prompts and that many of the responses were double-marked. This linking is described in section 7.2.

The markers involved in the study were trained teachers, or held relevant postgraduate degrees. Each marker attended a two-day training course at the start of the marking exercise. Marking was done in teams and moderation meetings were carried out on a daily basis. Each marker provided each script with seven marks—one for each of the elements: ideas; structure and language; organisation; vocabulary; sentence structure; punctuation; spelling. The markers provided these marks based on their interpretation of the marking rubric described in the e-asTTle tool. Data were entered and carefully validated before the analysis was carried out using the MFRM software package “Facets” (Linacre, 2010).

As with all robust modelling processes, a number of models were investigated before settling on the model described below. These intermediate models investigated a number of different configurations of the relevant facets. For example, one model allowed for a different set of rubric thresholds for each pair consisting of a prompt type (see Table 31) and an element—a total of 49 sets of thresholds. The final model described below and in section 7.4 was deemed to have the best combination of fit statistics, conceptual validity and utility. The fit statistics are described in section 7.4.

7.3.1. An overview of the MFRM

The final model is:

$$\log\left(\frac{P_{nijkh}}{1 - P_{nijkh}}\right) = B_n - D_i - C_j - E_k - T_{kh}$$

Where;

- P_{nijkh} represents the probability of student n , writing to prompt i , being scored by marker j on element k being scored in category h rather than category $h - 1$.
- B_n represents the writing proficiency of student n
- D_i represents the difficulty of prompt i
- C_j represents the harshness of marker j
- E_k represents the difficulty of element k
- T_{kh} represents the difficulty of threshold h of element k .

The final model had high reliability indices. In particular, the student reliability index (analogous to Cronbach's alpha) was 0.96.

Figure 2 provides a graphical representation of the measurement scale constructed by the analysis process. The scale itself is presented on the left of the figure in e-asTTle writing scale units (aWs). The scale locations of students, prompts, markers, the elements of the rubric and the scale thresholds are displayed from left to right. As can be seen, these locations vary. Prompt 20, for instance, is located slightly higher on the scale than Prompt 27, indicating it was the more difficult of the two prompts. Similarly, some markers (indicated by asterisks) are higher on the scale than others, indicating they applied the rubric more harshly.

The "Facets ruler" of Figure 2 is intended to provide a broad picture of how the different facets combine. All facets except the student facet are constrained to have an average measure of 0. To interpret Figure 1, see Linacre (2010).

Figure 2 Scale locations for the difference facets used in developing a model for e-asTTle writing — the “Facets Ruler”

| aWs | student | prompt | marker | element | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------|---------|---|--------|-------------------|-----|-----|-----|-----|-----|-----|-----|
| 1904 | + | + | + | + | (6) | (7) | (6) | (6) | (7) | (6) | (6) |
| 1843 | + | + | + | + | | --- | | | | --- | |
| 1782 | + | + | + | + | --- | 6 | --- | --- | --- | 5 | --- |
| 1721 | + | + | + | + | 5 | | | | 6 | | |
| 1660 | + | + | + | + | --- | --- | 5 | 5 | --- | --- | 5 |
| 1599 | + | + | + | + | 4 | 5 | --- | --- | 5 | 4 | --- |
| 1538 | + | + | + | + | --- | --- | 4 | 4 | --- | --- | 4 |
| 1477 | + | 20 21 23 26 29 31 11 12 13 14 19 22 15 16 17 18 25 27 | + | 7 2 4 5 3 1 | --- | 4 | --- | --- | 4 | 3 | --- |
| 1416 | + | + | + | + | 3 | | 3 | 3 | --- | --- | 3 |
| 1355 | + | + | + | 6 | | 3 | --- | --- | 3 | --- | --- |
| 1294 | + | + | + | + | --- | --- | | | | | |
| 1233 | + | + | + | + | 2 | 2 | 2 | 2 | 2 | | 2 |
| 1172 | + | + | + | + | 2 | 2 | | | 2 | | |
| 1111 | + | + | + | + | --- | --- | --- | --- | --- | 2 | --- |
| 1050 | + | + | + | + | | | | | | | |
| 989 | + | + | + | + | | | | | | | |
| 928 | + | + | + | + | | | | | | --- | |
| 867 | + | + | + | + | | | | | | | |
| 806 | + | + | + | + | | | | | | | |
| 745 | + | + | + | + | | | | | | | |
| 684 | + | + | + | + | (1) | (1) | (1) | (1) | (1) | (1) | (1) |
| aWs | * = 39 | | * = 1 | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

element
1= Ideas
2= Organisation
3= Vocabulary
4= Sentence structure
5= Punctuation
6= Spelling
7= Structure and language

7.3.2. The measurement scale

The e-asTTle writing scores produced by the MFRM were measured in “logits”. To convert these logit units to e-asTTle writing units (*aWs*), a linear transformation was applied to the logit scores, which ensures that the mean of the transformed scores for Year 6 students in Quarter 3 is 1500 units and the standard deviation is 100 units. To convert these back to logits, the following formula can be applied:

$$L = (aWs - 1500) * \frac{1.64}{100} + 0.38$$

Where:

- *aWs* denotes the e-asTTle writing scale score.
- *L* denotes the e-asTTle writing score measured in logits.

The conversion formula is designed to maintain as much consistency as possible with the previous version of e-asTTle writing. All writing scores in this document are reported in e-asTTle writing units.

7.4. Fit of the multifacet Rasch model and data

The Facets software provides several fit indices for different aspects of the model. In addition, a range of graphical displays are available to study the fit of the data to the measurement model.

As part of model development, it was noted that the data informing Element 2 (structure and language) was not fitting the MFRM well for some prompts. The anomalous data was isolated and not used to define the thresholds for this element. Consequently only data from prompts 8, 11, 12, 13, 14, 15, 16, 19, 20 and 21 informed the development of the thresholds for Element 2. This did not directly or substantially affect the other elements.

Overall model fit was very good. The tables in this section provide the value of the parameter (“Measure”) and the standard error on this (“se”). They also provide two fit statistics whose descriptions can be found in Linacre, 2010: the infit mean-square and the outfit mean-square. The infit mean-square indices of the prompts—which are displayed in Table 9—ranged from 0.78 to 1.23, and those of the elements—which are displayed in Table 10—ranged from 0.82 to 1.20. Values between 0.5 to 1.5 are generally considered good enough for measurement (Linacre, 2010).

Table 9 **Prompt statistics**

| Prompt | Measure | se | Infit mean-square | Outfit mean-square |
|--------|---------|----|-------------------|--------------------|
| 1 | 1465 | 2 | 1.16 | 1.10 |
| 2 | 1481 | 2 | 1.06 | 1.03 |
| 3 | 1471 | 2 | 1.01 | 0.97 |
| 4 | 1470 | 2 | 1.04 | 1.00 |
| 5 | 1454 | 2 | 0.78 | 0.75 |
| 6 | 1441 | 2 | 0.91 | 0.84 |
| 7 | 1452 | 2 | 0.91 | 0.86 |
| 8 | 1456 | 2 | 0.96 | 0.97 |
| 9 | 1483 | 2 | 1.10 | 1.08 |
| 10 | 1510 | 2 | 1.07 | 1.05 |
| 11 | 1507 | 2 | 1.00 | 1.00 |
| 12 | 1480 | 2 | 0.83 | 0.84 |
| 13 | 1515 | 2 | 0.84 | 0.84 |
| 14 | 1479 | 2 | 1.00 | 0.99 |
| 15 | 1443 | 2 | 0.98 | 0.98 |
| 16 | 1500 | 2 | 0.97 | 0.98 |
| 17 | 1459 | 2 | 1.03 | 1.04 |
| 18 | 1481 | 2 | 0.9 | 0.92 |
| 19 | 1495 | 2 | 1.23 | 1.24 |
| 20 | 1480 | 2 | 1.21 | 1.22 |
| 21 | 1493 | 2 | 1.07 | 1.08 |

Table 10 **Element statistics**

| Element | Measure | se | Infit mean-square | Outfit mean-square |
|------------------------|---------|----|-------------------|--------------------|
| Ideas | 1460 | 1 | 1.00 | 0.99 |
| Structure and language | 1512 | 1 | 1.05 | 1.04 |
| Organisation | 1507 | 1 | 1.19 | 1.17 |
| Vocabulary | 1478 | 1 | 0.82 | 0.81 |
| Sentence structure | 1496 | 1 | 0.88 | 0.86 |
| Punctuation | 1522 | 1 | 1.20 | 1.20 |
| Spelling | 1363 | 1 | 0.87 | 0.87 |

As might be expected given human variability, infit mean-square indices for markers—which are displayed in Table 11—varied a little more, ranging from 0.75 to 1.46.

Table 11 **Marker statistics**

| Marker | Measure | se | Infit mean-square | Outfit mean-square |
|--------|---------|----|-------------------|--------------------|
| 1 | 1461 | 2 | 1.34 | 1.26 |
| 2 | 1476 | 2 | 0.85 | 0.79 |
| 3 | 1437 | 2 | 0.84 | 0.89 |
| 4 | 1552 | 3 | 1.46 | 1.4 |
| 5 | 1493 | 2 | 1.04 | 1 |
| 6 | 1494 | 2 | 0.87 | 0.87 |
| 7 | 1490 | 2 | 0.91 | 0.88 |
| 8 | 1454 | 2 | 1.04 | 1.04 |
| 9 | 1441 | 2 | 1.1 | 1.1 |
| 10 | 1457 | 2 | 1.1 | 1.08 |
| 11 | 1434 | 1 | 0.75 | 0.75 |
| 12 | 1438 | 2 | 0.98 | 0.97 |
| 13 | 1492 | 2 | 0.83 | 0.84 |
| 14 | 1470 | 2 | 1.25 | 1.25 |
| 15 | 1468 | 2 | 1.3 | 1.33 |
| 16 | 1476 | 2 | 1.28 | 1.29 |
| 17 | 1463 | 2 | 0.99 | 1.01 |
| 18 | 1493 | 2 | 1 | 0.98 |
| 19 | 1476 | 2 | 1.06 | 1.06 |
| 20 | 1441 | 2 | 1.02 | 1.04 |
| 21 | 1526 | 2 | 0.86 | 0.84 |
| 22 | 1496 | 2 | 0.93 | 0.93 |
| 23 | 1542 | 2 | 0.77 | 0.73 |
| 25 | 1431 | 2 | 1.06 | 1 |
| 26 | 1513 | 2 | 0.77 | 0.75 |
| 27 | 1483 | 5 | 1.18 | 1.13 |

A brief description of how the errors in these tables are incorporated into the e-asTTle tool is provided in section 8.2.2.

Scale thresholds for the different elements are provided in Table 12 together with the indices of fit provided by the Facets software (Linacre, 2010). The lower Rasch-Andrich thresholds provided in Table 12 indicate the score boundaries for each rubric category for each element. The probability and characteristic curves are displayed in Figure 3 through to Figure 9.

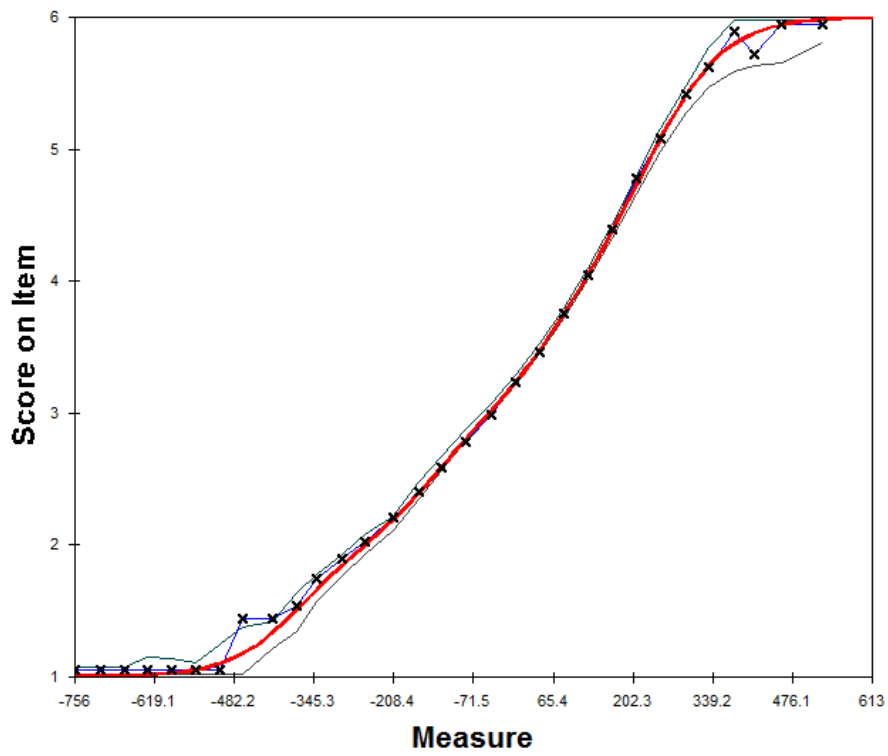
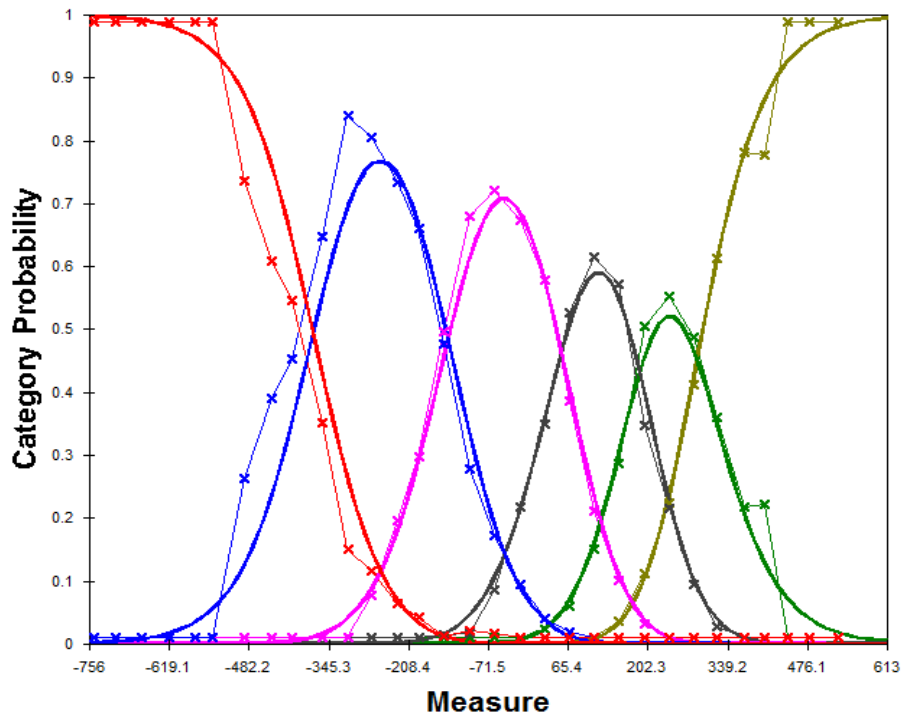
Table 12 **Threshold statistics**

| Scale | Category | Number of students in each category | Percent (%) | Lower Rasch-Andrich threshold | se | Outfit mean-square |
|-------------------------------|----------|-------------------------------------|-------------|-------------------------------|----|--------------------|
| Ideas | R1 | 365 | 4 | | | 1.8 |
| | R2 | 1346 | 15 | -375 | 5 | 1.0 |
| | R3 | 3193 | 35 | -144 | 3 | 0.9 |
| | R4 | 2698 | 30 | 51 | 2 | 0.9 |
| | R5 | 1112 | 12 | 185 | 2 | 0.9 |
| | R6 | 330 | 4 | 283 | 4 | 1.1 |
| Structure and language | R1 | 111 | 3 | | | 1.9 |
| | R2 | 779 | 18 | -361 | 8 | 1.1 |
| | R3 | 1411 | 33 | -103 | 3 | 1 |
| | R4 | 1313 | 31 | 34 | 3 | 0.9 |
| | R5 | 504 | 12 | 162 | 3 | 0.9 |
| | R6 | 108 | 3 | 268 | 7 | 1.0 |
| Organisation | R1 | 442 | 5 | | | 1.1 |
| | R2 | 1390 | 15 | -392 | 5 | 0.9 |
| | R3 | 2174 | 24 | -171 | 3 | 0.9 |
| | R4 | 2183 | 24 | -34 | 2 | 1.3 |
| | R5 | 2042 | 23 | 51 | 2 | 1.5 |
| | R6 | 760 | 8 | 181 | 3 | 1.2 |
| | R7 | 53 | 1 | 365 | 9 | 1.1 |
| Vocabulary | R1 | 524 | 6 | | | 0.8 |
| | R2 | 1533 | 17 | -337 | 4 | 0.8 |
| | R3 | 3122 | 35 | -131 | 2 | 0.8 |
| | R4 | 2334 | 26 | 48 | 2 | 0.8 |
| | R5 | 1180 | 13 | 156 | 2 | 0.8 |
| | R6 | 351 | 04 | 263 | 4 | 0.9 |
| Sentence Structure | R1 | 588 | 7 | | | 1.1 |
| | R2 | 2029 | 22 | -342 | 4 | 0.9 |
| | R3 | 2755 | 30 | -103 | 2 | 0.8 |
| | R4 | 2219 | 25 | 36 | 2 | 0.8 |
| | R5 | 1183 | 13 | 140 | 2 | 0.8 |
| | R6 | 270 | 3 | 268 | 4 | 0.9 |
| Punctuation | R1 | 751 | 8 | | | 1.8 |
| | R2 | 892 | 10 | -310 | 4 | 1.2 |
| | R3 | 2300 | 25 | -214 | 3 | 1.2 |
| | R4 | 3098 | 34 | -62 | 2 | 1.3 |
| | R5 | 1510 | 17 | 91 | 2 | 1.1 |
| | R6 | 404 | 4 | 205 | 4 | 1.0 |
| | R7 | 89 | 1 | 290 | 7 | 1.2 |
| Spelling | R1 | 39 | 0 | | | 1.8 |
| | R2 | 1207 | 13 | -563 | 12 | 0.8 |
| | R3 | 2323 | 26 | -91 | 3 | 0.8 |
| | R4 | 2993 | 33 | 81 | 2 | 0.9 |
| | R5 | 1908 | 21 | 221 | 2 | 0.9 |
| | R6 | 574 | 6 | 351 | 3 | 0.9 |

In Figure 3 through to Figure 9, the first graph in each figure shows, for the element in question, the probability of a student's writing being judged in each rubric category given the student's writing ability measure (if rater harshness and prompt difficulty are held at their average values). The second graph in each figure shows the expected rubric score for the element in question given the student's writing ability measure (if rater harshness and prompt difficulty are held at their average values).

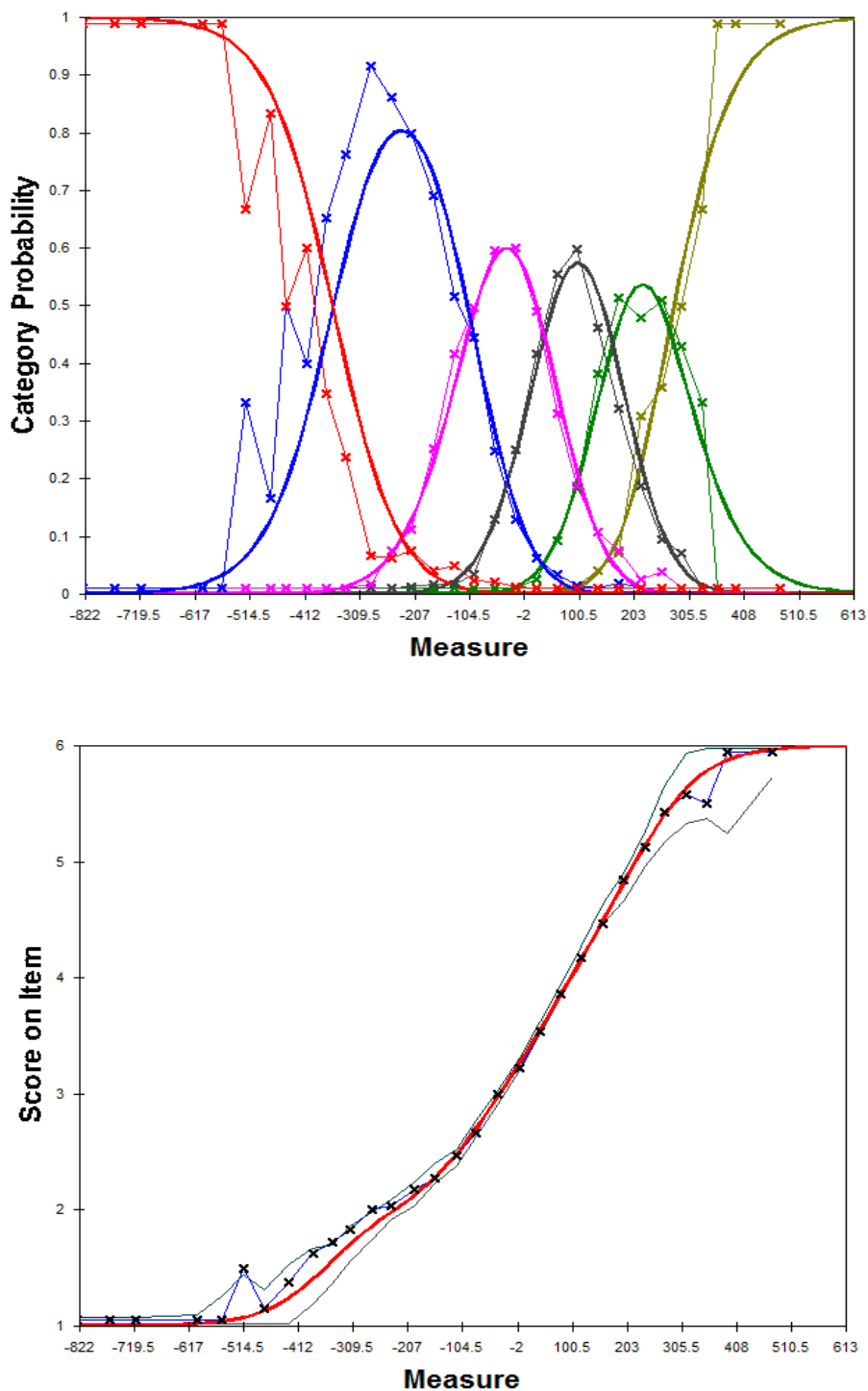
In both graphs, the smooth bold lines depict the model, the lines with "x"-markers show the data and, in the second graph of each figure, the thin bold lines depict the error in the model.

Figure 3 **Graphical representations of Element 1: Ideas**



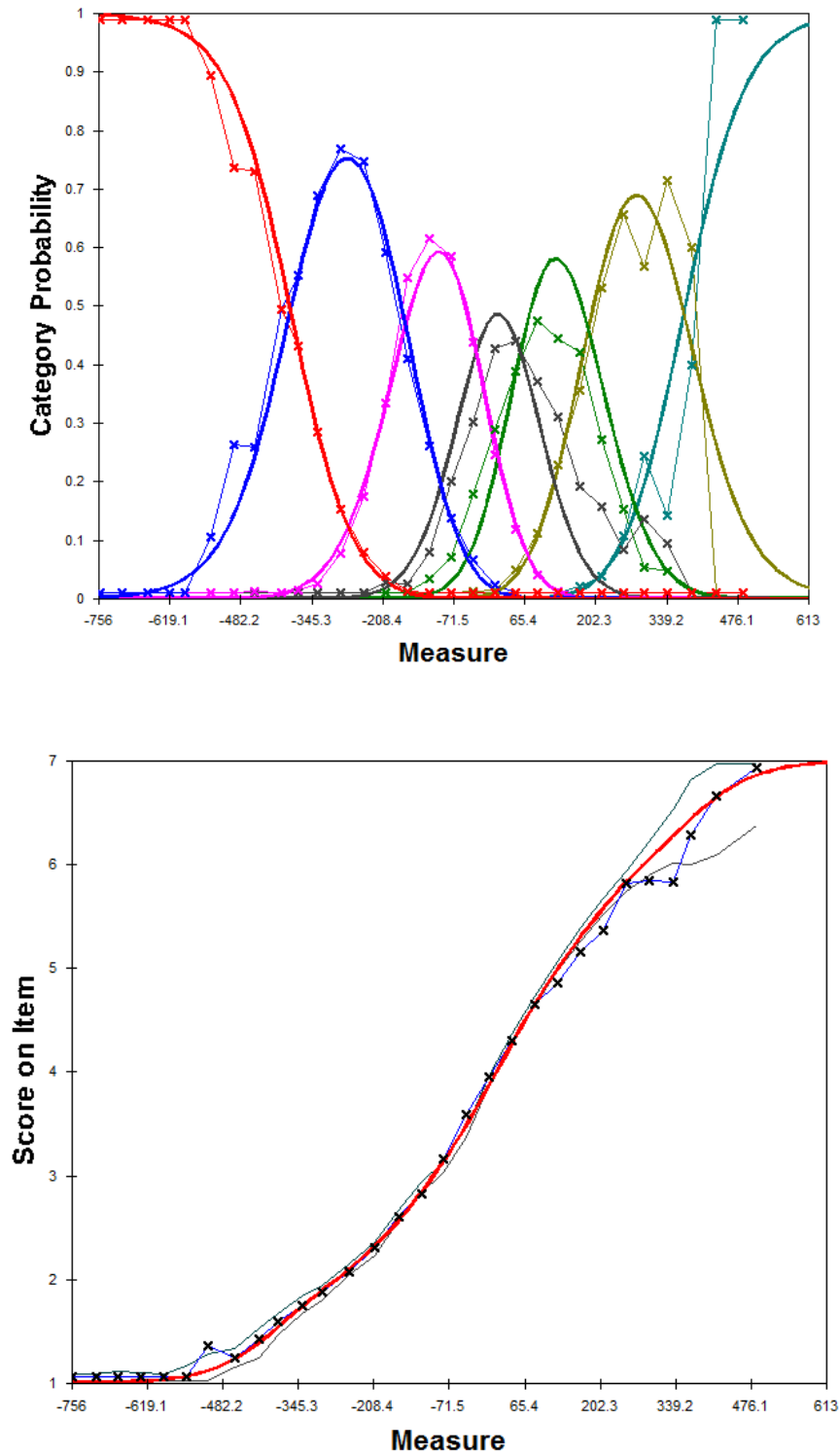
The two graphs show overall good fit for Element 1.

Figure 4 **Graphical representations of Element 2: Structure and language**



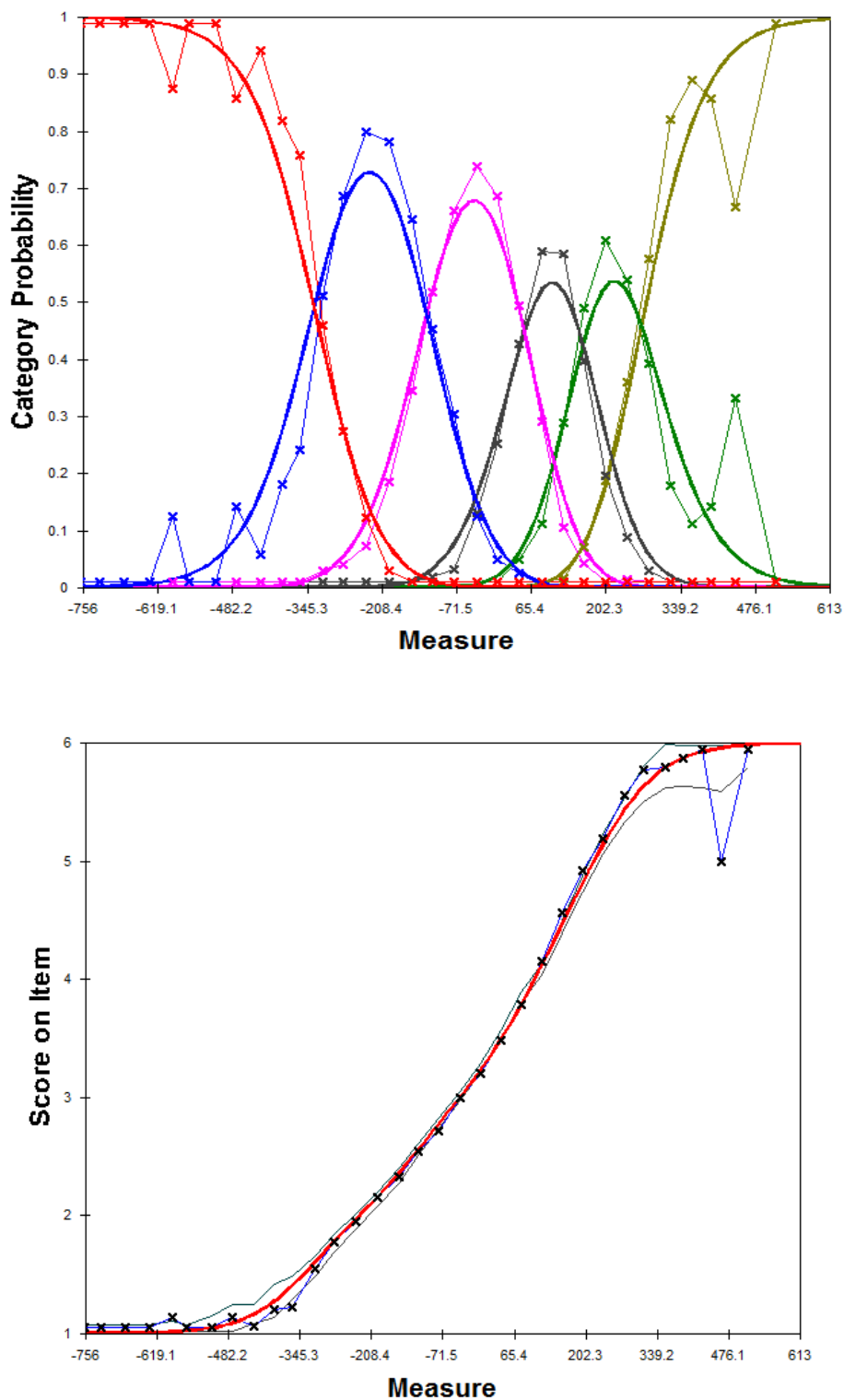
Element 2 is depicted with good fit here except for some slight anomalous behaviour around -500 aWs.

Figure 5 **Graphical representations of Element 3: Organisation**



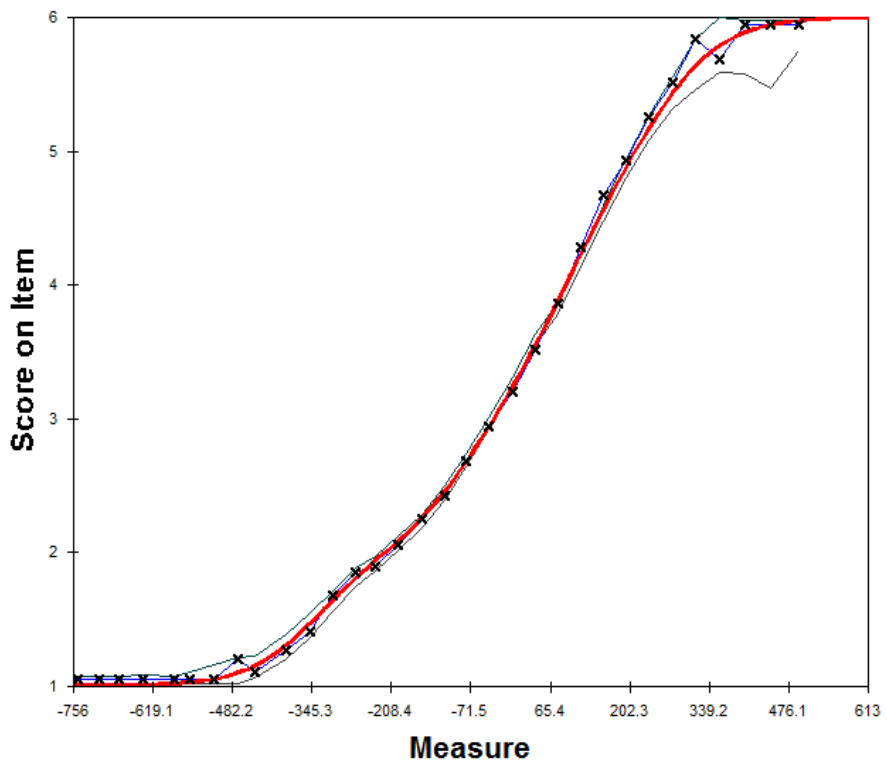
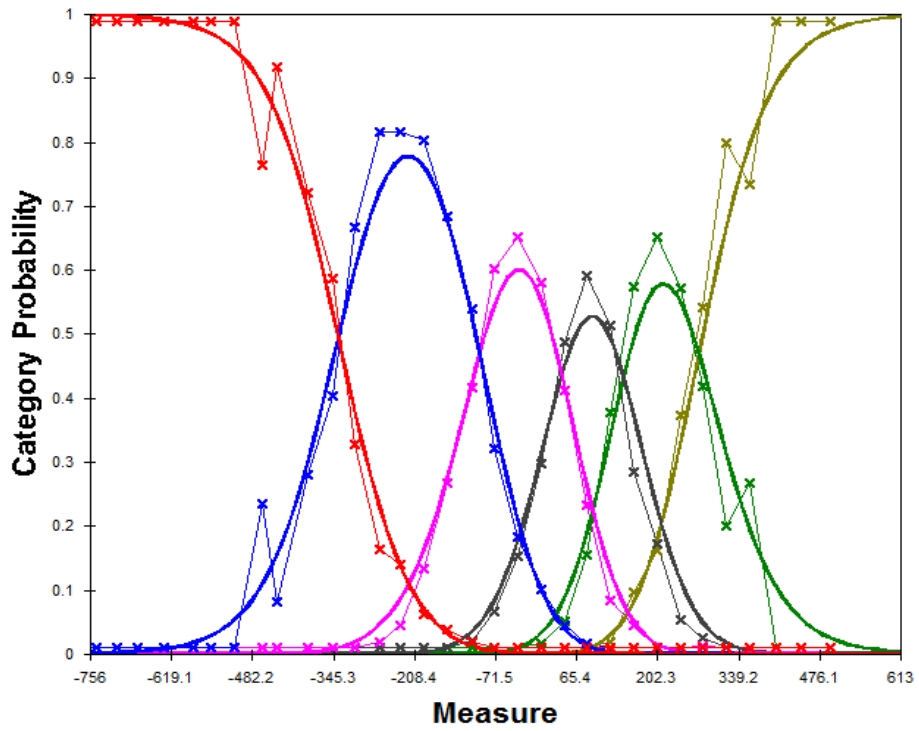
Element 3 shows overall good fit with a small exception around 300 aWs where sample students scored somewhat lower than might be expected.

Figure 6 **Graphical representations of Element 4: Vocabulary**



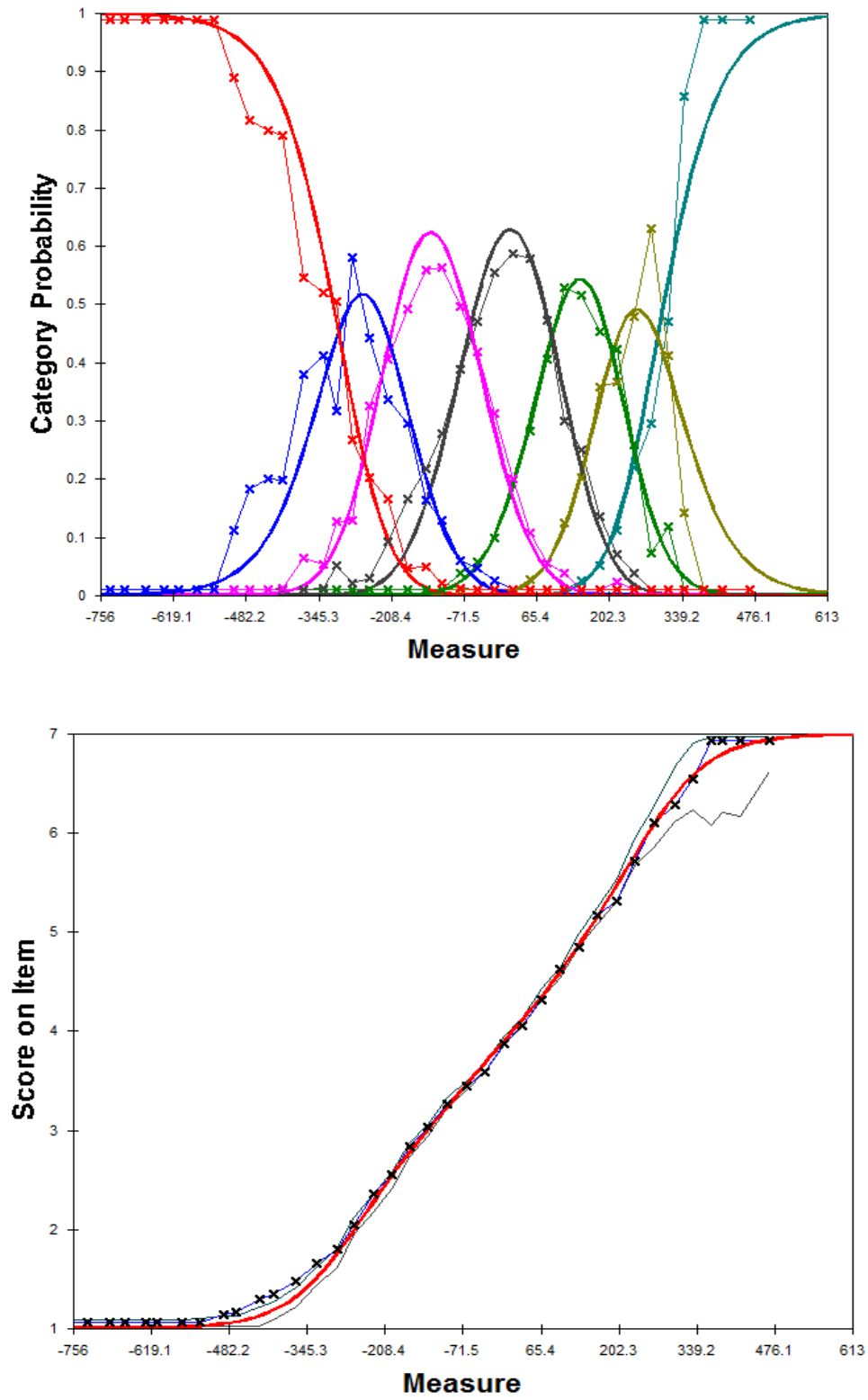
Element 4 displays good fit with some slight anomalous behaviour at around 400 aWs where sample students scored lower than might be expected.

Figure 7 **Graphical representations of Element 5: Sentence structure**



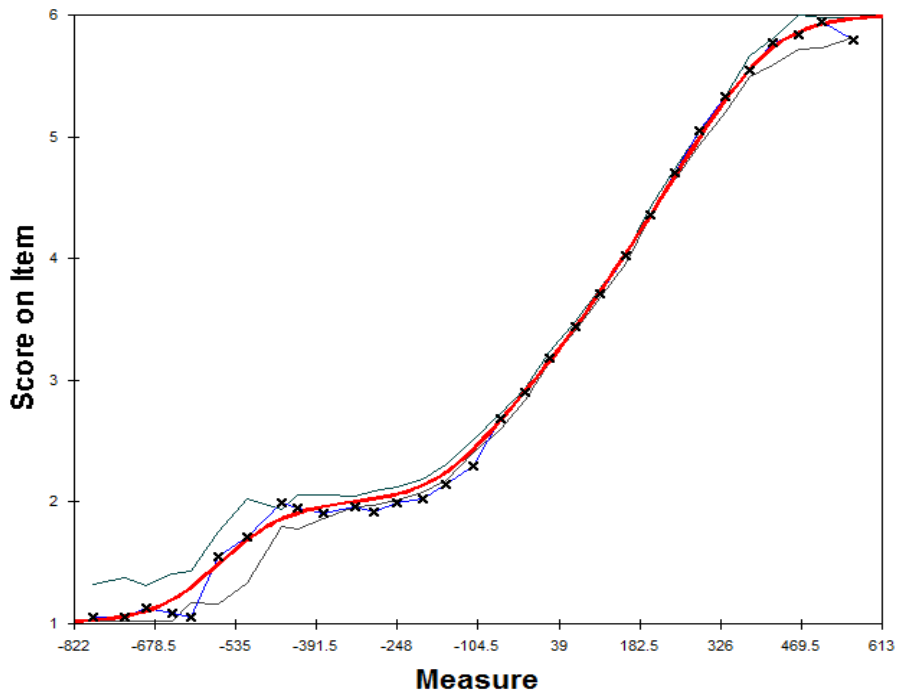
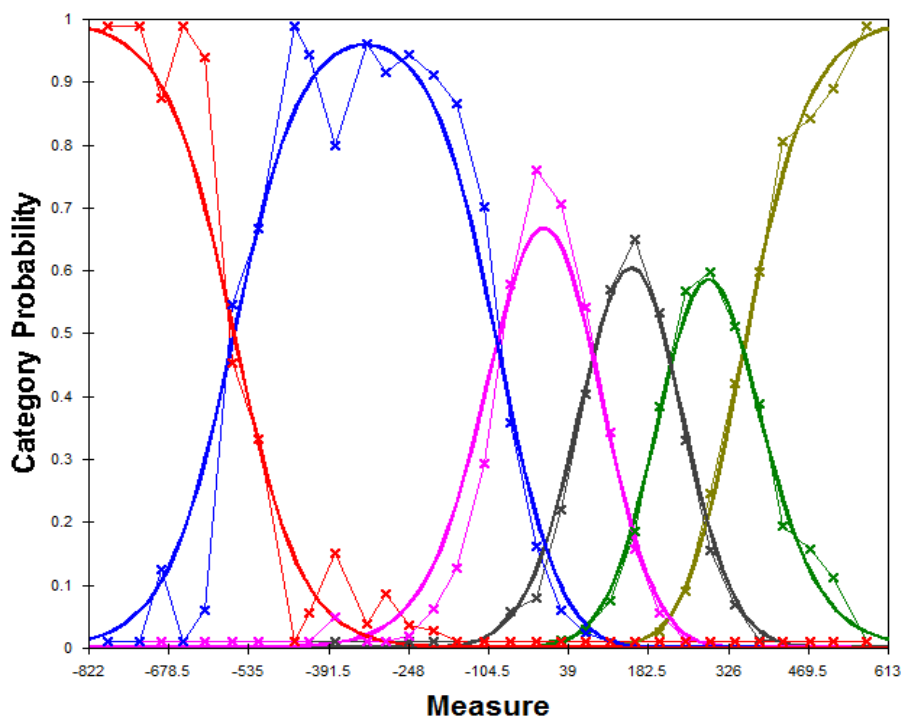
Element 5 shows good overall fit.

Figure 8 **Graphical representations of Element 6: Punctuation**



Similarly, Element 6 shows good overall fit.

Figure 9 **Graphical representations of Element 7: Spelling**



Element 7 shows good overall fit. Between around -400 aWs and -150 aWs, an increase in writing ability translates into very little gain in rubric score on Element 7.

7.5. Reliability

Reliability refers to the consistency of a measure. A highly reliable assessment administered on two separate occasions and under similar conditions will produce results which are consistent.

Reliability is often presented in the form of a reliability coefficient based on a statistical calculation. These coefficients normally range from 0 to +1 with 1 indicating a perfectly consistent measure (never attained in educational measurement).

Four aspects of reliability are considered in this section of the report using data collected during the national trial of e-asTTLe writing:

- the extent to which markers used the rubric in a consistent manner (inter-marker reliability)
- the internal consistency of the measure
- the extent to which repeated assessments using various prompts provided consistent measurements (analogous to parallel form reliability)
- the reliability of the model parameters.

7.5.1. Consistent use of the rubric

The trial involved 25 markers who were trained to use the rubric. The markers worked in groups with two markers in each group marking across groups. Each group were involved in daily moderation exercises and markers were able to discuss their rubric decisions with other markers. As part of the exercise a large number of scripts were scored by more than one marker so that the measurement model could adjust for any differences in the relative harshness/leniency of the markers.

Table 13 shows the agreement rate between markers for scripts which were double marked over all scores for each element of the rubric. Overall the markers agreed on an element score 48% of the time and were within one rubric scoring category 91% of the time.

Table 13 **Agreement rate amongst markers**

| Task | Exact agreement (%) | Agreement within 1 rubric level (%) |
|------|---------------------|-------------------------------------|
| 1 | 48 | 89 |
| 2 | 47 | 92 |
| 3 | 44 | 92 |
| 4 | 54 | 74 |
| 5 | 55 | 94 |
| 6 | 53 | 94 |
| 7 | 49 | 90 |
| 8 | 42 | 88 |

| | | |
|---------|----|----|
| 9 | 47 | 93 |
| 10 | 58 | 96 |
| 11 | 43 | 88 |
| 12 | 49 | 92 |
| 13 | 47 | 91 |
| 14 | 50 | 90 |
| 15 | 52 | 91 |
| 16 | 45 | 93 |
| 17 | 40 | 86 |
| 18 | 50 | 94 |
| 19 | 44 | 90 |
| 20 | 42 | 87 |
| 21 | 48 | 89 |
| Overall | 48 | 91 |

It is possible to create two sets of total rubric scores for the scripts that were marked twice (first mark and second mark). The correlation between these was 0.81.

Teachers who use the e-asTTle rubric are supported by an extensive bank of exemplars and encouraged to moderate with colleagues. This will support the consistent use of the rubric. However, some marker variation is inevitable. To recognise this, the tool includes an estimate of marker variation when it calculates the measurement error associated with a scale score (for more details see section 8.2.2).

7.5.2. Internal consistency

Internal consistency is generally based on correlations between the different parts of an assessment. In the case of e-asTTle Writing this involves considering the extent to which the scores on the different elements of the rubric associate with each other. Internal reliability is often reported using the Cronbach's Alpha statistic.

The table below provides the Cronbach's Alpha statistic for the e-asTTle writing trial data by year level.

| Year level | Cronbach's Alpha | Number of results informing the computation |
|------------|------------------|---|
| 1 | 0.892 | 204 |
| 2 | 0.878 | 235 |
| 3 | 0.921 | 189 |
| 4 | 0.867 | 624 |
| 5 | 0.874 | 770 |
| 6 | 0.892 | 897 |

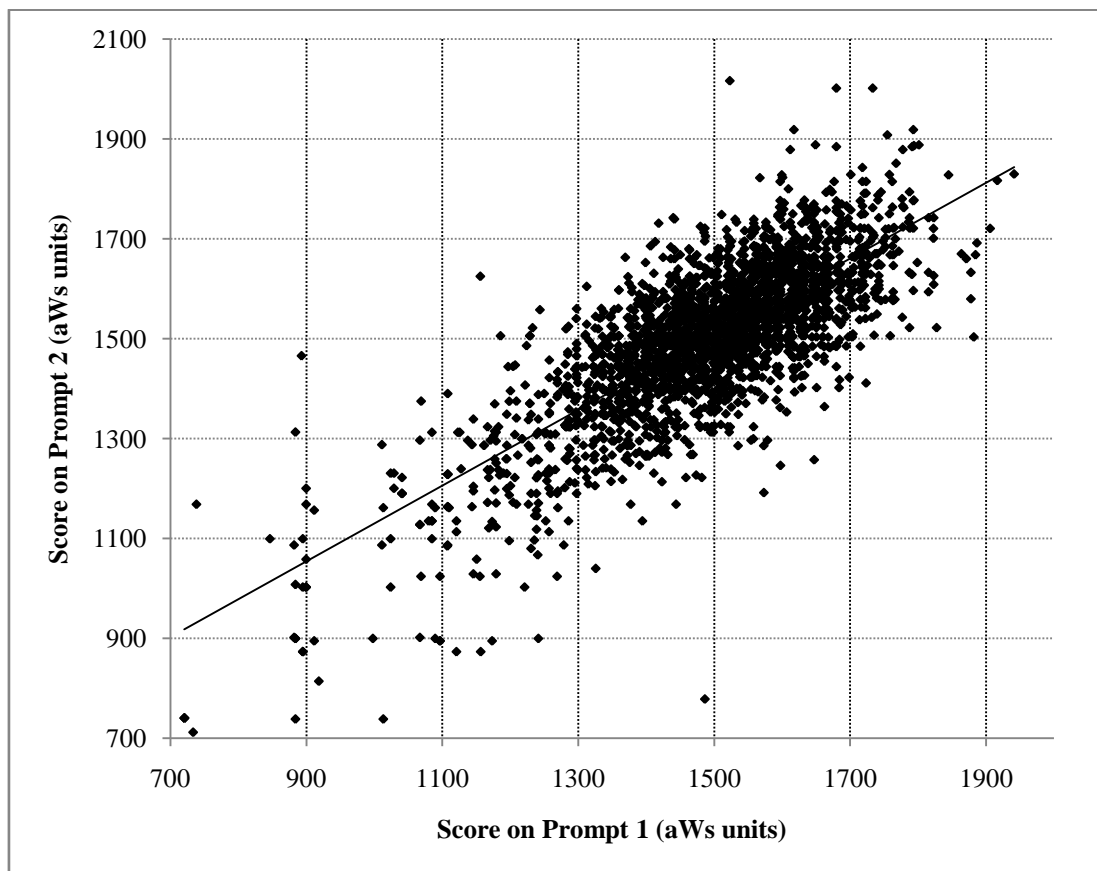
| | | |
|-----|-------|------|
| 7 | 0.901 | 1001 |
| 8 | 0.906 | 1049 |
| 9 | 0.905 | 837 |
| 10 | 0.907 | 827 |
| All | 0.934 | 6633 |

7.5.3. Test-retest reliability

The e-asTTle writing tool allows teachers to assess students using an e-asTTle writing prompt and then compare the resulting scale scores with scale scores from other e-asTTle writing assessments which may have used a different prompt. The multi-faceted Rasch model has been used to incorporate differences in the difficulty of the prompts in the scale score calculations.

During the trial a number of students were assessed twice using a different prompt each time. This was done in order to link all prompts to a common scale. **Error! Reference source not found.** plots the two sets of scale scores against each other. A line of best fit is also shown. The correlation between these two sets of scores was 0.77. Some students showed a substantial increase or decrease in score on the second occasion, which could be attributed to factors such as a practice effect, better use of test time, a personal reaction to the prompt, and/or a decline in test motivation. Since the tests are not designed to be administered in this way, it is considered that that the correlation reported represents a low estimate of the reliability of the test.

Figure 10 **First prompt vs second prompt scores**



7.5.4. Reliability of model parameters

The Facets software generates reliability indices for each of the facets included in the model (students, prompts, markers and elements). The student reliability index was 0.96, the prompt and marker reliabilities were both 0.99 and the element reliability was 1.00 (rounded to 2 decimal places). These indices are high, indicating that the model is able to precisely locate students, prompts, markers and rubric elements on the measurement scale.

7.6. Writing scores in the sample

This section describes the distribution of writing scores in the sample. For reasons described in section 7.7, it was not used directly to determine normative information. Instead, it was used to inform a statistical model which then determined the normative information. It is included here for comparison with the normative information as a check of the statistical model.

Figure 11, Table 14, Table 15 and Table 16 describe the distribution of writing scores in the sample by several relevant demographic characteristics. Figure 11 shows a box plot of the distribution of writing scale scores for each year level in the sample. The box plot displays—in ascending order—the fifth percentile, the lower quartile, the median, the upper quartile and the ninety-fifth percentile of the writing scores. The medians exhibit a typical curved growth pattern.

The variation in writing scores (as indicated by the interquartile range) is somewhat larger for Years 1, 2 and 3 than for the remaining year levels.

Figure 11 **e-asTTle writing sample score distributions by year level**

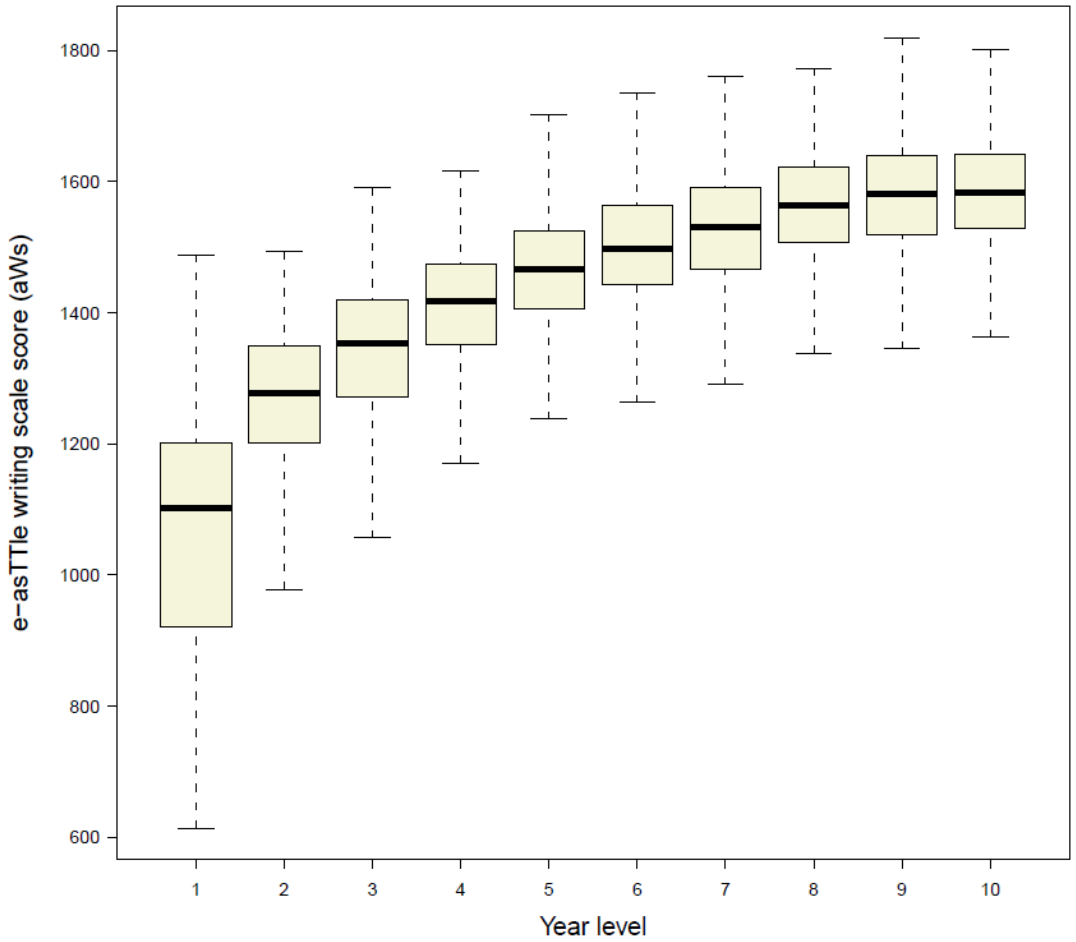


Table 14 shows, for each year level in the sample, the mean writing score and standard deviation in writing scores – together with their standard errors – for boys, girls and all students. The growth pattern and variation in scores is shown with means and standard deviations. At each year level, girls tended to have higher writing scores than boys.

Table 14

e-asTTle writing sample statistics by year level and gender

| Year level | Population | Mean | Mean se | Std deviation | Std error |
|-------------------|-------------------|-------------|----------------|----------------------|------------------|
| 1 | Boy | 1034 | 13 | 190 | 9 |
| | Girl | 1102 | 12 | 172 | 8 |
| | Total | 1068 | 9 | 184 | 6 |
| 2 | Boy | 1218 | 10 | 148 | 7 |
| | Girl | 1285 | 7 | 122 | 5 |
| | Total | 1255 | 6 | 138 | 4 |
| 3 | Boy | 1302 | 8 | 123 | 5 |
| | Girl | 1375 | 7 | 115 | 5 |
| | Total | 1337 | 5 | 125 | 4 |
| 4 | Boy | 1390 | 6 | 102 | 4 |
| | Girl | 1434 | 5 | 82 | 4 |
| | Total | 1410 | 4 | 96 | 3 |
| 5 | Boy | 1445 | 6 | 92 | 4 |
| | Girl | 1474 | 6 | 101 | 4 |
| | Total | 1460 | 4 | 98 | 3 |
| 6 | Boy | 1464 | 8 | 112 | 5 |
| | Girl | 1526 | 6 | 86 | 4 |
| | Total | 1495 | 5 | 104 | 4 |
| 7 | Boy | 1492 | 7 | 100 | 5 |
| | Girl | 1557 | 5 | 88 | 4 |
| | Total | 1526 | 4 | 99 | 3 |
| 8 | Boy | 1540 | 5 | 88 | 4 |
| | Girl | 1587 | 5 | 85 | 4 |
| | Total | 1562 | 4 | 90 | 3 |
| 9 | Boy | 1556 | 7 | 112 | 5 |
| | Girl | 1600 | 6 | 91 | 4 |
| | Total | 1576 | 5 | 105 | 3 |
| 10 | Boy | 1570 | 6 | 103 | 4 |
| | Girl | 1598 | 7 | 112 | 5 |
| | Total | 1583 | 5 | 108 | 4 |

Table 15 shows, for each year level in the sample, the mean writing score and standard deviation in writing scores—together with their standard errors—for New Zealand European, Māori, Pasifika, and Asian or Other students. The Asian and Other categories were merged to allow sufficiently robust statistics to be reported. For each year level, New Zealand European and Asian or Other students tended to have higher writing scores than Māori students, who in turn tended to have higher writing scores than Pasifika students.

Table 15 e-asTTle writing sample statistics by year level and ethnicity

| Year level | Ethnicity | Mean | Mean se | Std deviation | Std error |
|------------|----------------|------|---------|---------------|-----------|
| 1 | NZ European | 1093 | 11 | 177 | 8 |
| | Māori | 1018 | 18 | 189 | 13 |
| | Pasifika | 1004 | 31 | 176 | 23 |
| | Asian or Other | 1119 | 24 | 177 | 18 |
| 2 | NZ European | 1268 | 8 | 125 | 5 |
| | Māori | 1222 | 14 | 162 | 10 |
| | Pasifika | 1216 | 19 | 129 | 13 |
| | Asian or Other | 1303 | 18 | 124 | 13 |
| 3 | NZ European | 1355 | 7 | 113 | 5 |
| | Māori | 1293 | 12 | 129 | 9 |
| | Pasifika | 1277 | 25 | 156 | 18 |
| | Asian or Other | 1393 | 16 | 103 | 12 |
| 4 | NZ European | 1423 | 5 | 94 | 4 |
| | Māori | 1373 | 9 | 82 | 6 |
| | Pasifika | 1372 | 12 | 95 | 9 |
| | Asian or Other | 1430 | 12 | 105 | 8 |
| 5 | NZ European | 1473 | 5 | 87 | 4 |
| | Māori | 1417 | 10 | 89 | 7 |
| | Pasifika | 1425 | 15 | 113 | 10 |
| | Asian or Other | 1467 | 14 | 125 | 10 |
| 6 | NZ European | 1504 | 6 | 104 | 4 |
| | Māori | 1485 | 12 | 91 | 9 |
| | Pasifika | 1455 | 15 | 95 | 11 |
| | Asian or Other | 1493 | 13 | 111 | 10 |
| 7 | NZ European | 1536 | 5 | 95 | 4 |
| | Māori | 1488 | 9 | 87 | 6 |
| | Pasifika | 1482 | 13 | 83 | 9 |
| | Asian or Other | 1534 | 14 | 121 | 10 |
| 8 | NZ European | 1570 | 5 | 87 | 4 |
| | Māori | 1533 | 8 | 72 | 5 |
| | Pasifika | 1506 | 16 | 104 | 12 |
| | Asian or Other | 1578 | 12 | 98 | 8 |
| 9 | NZ European | 1586 | 5 | 92 | 4 |
| | Māori | 1556 | 9 | 92 | 7 |
| | Pasifika | 1513 | 28 | 170 | 20 |
| | Asian or Other | 1573 | 12 | 110 | 9 |
| 10 | NZ European | 1593 | 5 | 96 | 4 |
| | Māori | 1571 | 9 | 93 | 6 |
| | Pasifika | 1560 | 13 | 95 | 9 |
| | Asian or Other | 1590 | 16 | 150 | 12 |

Table 16 shows for each decile group in the sample, the mean writing score and standard deviation in writing scores—together with their standard errors—for all students from schools with that decile. Table 16 shows higher decile groups tending to perform somewhat better than lower decile groups.

Table 16 **e-asTTle writing sample statistics by decile group**

| Decile group | Mean | Mean se | Std deviation | Std error |
|--------------|------|---------|---------------|-----------|
| Deciles 1–3 | 1355 | 6 | 215 | 4 |
| Deciles 4–7 | 1452 | 4 | 175 | 3 |
| Deciles 8–10 | 1467 | 4 | 182 | 3 |

7.7. Normative information

This section describes the normative information used in the e-asTTle tool.

To function properly, the e-asTTle tool requires a substantial amount of normative information. For example, it requires summary statistics for the distribution of writing scores simultaneously broken down by year level, gender and ethnicity. The scope of the current work meant that some of the year level by gender by ethnicity cells were too small to produce summary statistics that were suitably robust.

The e-asTTle tool also requires all normative information for each quarter of the school year. This was also outside the scope of the current e-asTTle work. A statistical model based on the collected data was used to produce robust estimates of some of the required norms. The model is described in Section 7.7.1. The remaining norms were extrapolated from data collected in the previous version of e-asTTle. This extrapolation is described in Section 7.7.3.

7.7.1. The statistical model

The linear regression model of the e-asTTle sample data is described by:

$$\begin{aligned}
 aWs \approx & 1147 + 212 \\
 & \times \log_e(Years) - 72 \times Boy - 75 \times Māori - 58 \times Pasifika + 6 \times Other + 13 \\
 & \times \log_e(Years) \times Boy + 19 \times \log_e(Years) \times Māori + 63 \times Māori \times Pasifika
 \end{aligned}$$

Where:

- *aWs* denotes the e-asTTle writing score measured in e-asTTle writing units
- *Years* denotes year level but is to be interpreted as a continuous variable. Because the e-asTTle sample was collected in the third quarter of 2011, values of 2.5, 2.75, 3 and 3.25 for *Years*, for example, are interpreted as being the first, second, third and fourth quarters of year level 3 respectively. Similarly, values of 3.5, 3.75, 4 and 4.25 for *Years*, are interpreted as being the first, second, third and fourth quarters of year level 4 respectively, and so on.
- *Boy* denotes whether respondents were boys (*Boy* = 1) or girls (*Boy* = 0)

- *Māori* denotes whether respondents identified as Māori ($Māori = 1$) or not ($Māori = 0$)
- *Pasifika* denotes whether respondents identified as Pasifika ($Pasifika = 1$) or not ($Pasifika = 0$)
- *Other* denotes whether respondents identified as Asian or Other ($Other = 1$) or not ($Other = 0$).

All of the independent variables in the model are significant at the 0.01 level except for *Other*. The model explains around 67% of the variation in writing scores in the data and the overall fit to the data is good.

This model was used to produce the e-asTTle norms. This was done by substituting appropriate proportions for the independent variables. To estimate the mean e-asTTle writing score (measured in logits) of a subpopulation of interest, for example, Year 5 boys in Quarter 2, the independent variables and associated transforms take the following values:

| | | | |
|---|---|------|---|
| <i>Years</i> | = | 4.75 | The year level of subpopulation of interest (Quarter 2 of Year 5) |
| $\log_e(\textit{Years})$ | = | 1.56 | The value of $\log_e 4.75$ |
| <i>Boy</i> | = | 1 | The proportion of the subpopulation of interest that are boys |
| <i>Māori</i> | = | 0.21 | The proportion of the subpopulation of interest that identify as Māori—inferred from the proportion in the whole e-asTTle sample |
| <i>Pasifika</i> | = | 0.10 | The proportion of the subpopulation of interest that identify as Pasifika—inferred from the proportion in the whole e-asTTle sample |
| <i>Other</i> | = | 0.15 | The proportion of the subpopulation of interest that identify as Asian or Other—inferred from the proportion in the whole e-asTTle sample |
| $\log_e(\textit{Years}) * \textit{Boy}$ | = | 1.56 | The product of the two relevant values above |
| $\log_e(\textit{Years}) * \textit{Māori}$ | = | 0.48 | The product of 1.56 (see above) and the proportion of the subpopulation of interest that identify as Māori |
| $\textit{Māori} * \textit{Pasifika}$ | = | 0.01 | The proportion of the subpopulation of interest that identify as Māori and Pasifika—inferred from the proportion in the whole e-asTTle sample |

These values and the above model resulted in a logit value of the mean e-asTTle writing score of Year 5 boys in Quarter 2 of 1416.

The standard deviations for all subpopulations were modelled as follows:

Table 17 **e-asTTle writing model of standard deviation values**

| Apply to subpopulations | Standard deviation |
|--------------------------------|---------------------------|
| Not broken down by year level | 195 |
| Year level 1 | 184 |
| Year level 2 | 138 |
| Year level 3 | 125 |
| Year levels 4–10 | 100 |

For year levels 1, 2 and 3 these were simply the values from the e-asTTle writing sample. For year levels 4 through 10, the values from the e-asTTle writing sample were all very similar. The value for year levels 4 through 10 in Table 17 was the average of the standard deviations for each of these year levels.

7.7.2. Modelled writing scores

This subsection provides some norms from the e-asTTle model. Comparing the results of this section with those of the subsection “Writing scores in the sample” will also allow further insight into the fit of the e-asTTle model to the sample data. 0 and Table 19 are comparable with Table 14 and 0 Table 15 respectively. Table 16 has no analogue here as norms by decile are not explicitly required by the e-asTTle tool and subsequently decile was not included in the e-asTTle model.

0 shows similar patterns to Table 14.

Table 18 e-asTTle writing model statistics for Quarter 3, by year level and gender

| Year level | Population | Mean | Std deviation |
|------------|------------|------|---------------|
| 1 | Boy | 1055 | 184 |
| | Girl | 1127 | 184 |
| | Total | 1091 | 184 |
| 2 | Boy | 1216 | 138 |
| | Girl | 1278 | 138 |
| | Total | 1249 | 138 |
| 3 | Boy | 1309 | 125 |
| | Girl | 1367 | 125 |
| | Total | 1341 | 125 |
| 4 | Boy | 1376 | 100 |
| | Girl | 1429 | 100 |
| | Total | 1407 | 100 |
| 5 | Boy | 1427 | 100 |
| | Girl | 1478 | 100 |
| | Total | 1459 | 100 |
| 6 | Boy | 1470 | 100 |
| | Girl | 1518 | 100 |
| | Total | 1500 | 100 |
| 7 | Boy | 1506 | 100 |
| | Girl | 1551 | 100 |
| | Total | 1535 | 100 |
| 8 | Boy | 1536 | 100 |
| | Girl | 1581 | 100 |
| | Total | 1566 | 100 |
| 9 | Boy | 1564 | 100 |
| | Girl | 1606 | 100 |
| | Total | 1593 | 100 |
| 10 | Boy | 1588 | 100 |
| | Girl | 1629 | 100 |
| | Total | 1617 | 100 |

Table 19 shows similar patterns to Table 15.

Table 19 **e-asTTle writing model statistics for quarter 3, by year level and ethnicity**

| Year level | Ethnicity | Mean | Std deviation |
|------------|----------------|------|---------------|
| 1 | NZ European | 1111 | 184 |
| | Māori | 1036 | 184 |
| | Pasifika | 1051 | 184 |
| | Asian or Other | 1095 | 184 |
| 2 | NZ European | 1265 | 138 |
| | Māori | 1204 | 138 |
| | Pasifika | 1210 | 138 |
| | Asian or Other | 1254 | 138 |
| 3 | NZ European | 1355 | 125 |
| | Māori | 1302 | 125 |
| | Pasifika | 1302 | 125 |
| | Asian or Other | 1347 | 125 |
| 4 | NZ European | 1419 | 100 |
| | Māori | 1371 | 100 |
| | Pasifika | 1368 | 100 |
| | Asian or Other | 1413 | 100 |
| 5 | NZ European | 1469 | 100 |
| | Māori | 1426 | 100 |
| | Pasifika | 1419 | 100 |
| | Asian or Other | 1463 | 100 |
| 6 | NZ European | 1509 | 100 |
| | Māori | 1470 | 100 |
| | Pasifika | 1460 | 100 |
| | Asian or Other | 1505 | 100 |
| 7 | NZ European | 1544 | 100 |
| | Māori | 1507 | 100 |
| | Pasifika | 1496 | 100 |
| | Asian or Other | 1540 | 100 |
| 8 | NZ European | 1573 | 100 |
| | Māori | 1539 | 100 |
| | Pasifika | 1526 | 100 |
| | Asian or Other | 1571 | 100 |
| 9 | NZ European | 1600 | 100 |
| | Māori | 1567 | 100 |
| | Pasifika | 1553 | 100 |
| | Asian or Other | 1598 | 100 |
| 10 | NZ European | 1623 | 100 |
| | Māori | 1593 | 100 |
| | Pasifika | 1577 | 100 |
| | Asian or Other | 1622 | 100 |

7.7.3. Extrapolation from previously collected data

For the following variables used in e-asTTle reporting, the appropriate normative information was extrapolated from the data collected in the previous version of e-asTTle:

- location (urban/rural)
- school cluster (a geographical location)
- first language spoken at home.

This extrapolation was implemented by calculating the standardised effect of the mean writing score of any cell in the previous version of e-asTTle. That is, in the previous version of e-asTTle, the mean writing score for a cell described using the above variables was subtracted from the overall mean writing score and the result was divided by the overall standard deviation. This distance was the number of overall standard deviations in the current version of e-asTTle by which the specified cell's mean would differ from the overall mean.

7.8. Standard setting for e-asTTle writing

7.8.1. Introduction

NZCER carried out a standard-setting exercise to link scale scores for the revised e-asTTle writing with the levels of writing competency described by the literacy learning progressions (Ministry of Education, 2010). The results of the exercise were used to define the curriculum level performance bands (basic, advanced and proficient) used by the e-asTTle writing tool to report performance levels. The exercise was completed in two parts. Both parts followed the same format. The second part was initiated after judges experienced difficulty making levelling decisions using a response booklet organised by scale score. This section describes the exercise, explains the methodology used to construct the e-asTTle writing performance bands and provides the results.

7.8.2. The literacy learning progressions

The literacy learning progressions (LLP) provide descriptions of the writing competencies associated with seven different stages of schooling. Each stage is linked to a particular year level and curriculum level. The descriptions outline the processes students will be able to use, and the skills and knowledge they will be able to apply at each level. Table 20 shows how these descriptions are related to year-level expectations and curriculum levels.

Table 20 **The LLP levels and their link to curriculum levels**

| LLP level | Curriculum level |
|-------------------|------------------|
| After one year | Early Level 1 |
| After two years | Level 1 |
| After three years | Early Level 2 |
| End of Year 4 | Level 2 |
| End of Year 6 | Level 3 |
| End of Year 8 | Level 4 |
| End of Year 10 | Level 5 |

7.8.3. Standard-setting process

The standard-setting process was based on a modified bookmarking approach (Cizek & Bunch, 2007). The bookmark approach involves judges (subject matter experts) working their way through a book of test items that have been ordered from easiest to hardest according to their relative locations on an item response theory scale. Each judge independently places a bookmark at the first page where they consider a student who is minimally competent at the level under investigation could not be expected to correctly answer the item on the page at least two-thirds of the time. Working in groups, judges then share their bookmark placements and are given an opportunity to reconsider their placements. Several rounds of placements and discussions can take place before a final judgement is made. The judges' final placements are usually averaged to find a definitive cut-point on the scale.

Rather than items, the e-asTTle writing standard-setting process used actual student responses to the writing prompts (students' scripts). Each response had been double marked and the rubric scores converted to locations on an IRT scale. The responses were presented in a booklet (the response booklet) ordered according to scale score from the lowest scoring response to the highest scoring response. Focusing on one LLP level at a time, judges worked through the book to locate the script that they considered represented the minimally acceptable response for a student who was considered to be working at the LLP level in question. They then recorded the page number of the response. The average scale score for the bookmarked responses was then used as the scale score cut-point for the LLP level.

7.8.4. The judges

In total, eight judges were involved in both parts of the standard-setting exercise. In the first part, four of the judges were members of the e-asTTle writing reference group and three were members of the e-asTTle writing development team who had been involved in the development of prompts and the scoring rubric. The second part involved four judges from the e-asTTle writing development team. Three of the judges were involved in both parts.

7.8.5. The response booklet

The response booklet was made up of 76 scripts. Each of the scripts had been double marked by members of the e-asTTle writing development team and a consensus score arrived at. The scripts represented the range of e-asTTle prompts and were ordered according to their e-asTTle writing scale (aWs) score.

An aWs score represents a conversion of the overall rubric score to a location on an equal-interval measurement scale. The conversion process takes into account differences in the relative difficulty of the scoring criteria. It can also take into account differences in prompt difficulty. Prompt difficulty recognises that different writing prompts are more or less demanding, and that this affects students' performance levels. Prompt difficulty was determined as a parameter of the multifaceted Rasch model used to construct the scale.

For the first part of the exercise, the responses were ordered according to their e-asTTle scale score with prompt difficulty taken into account. Taking prompt difficulty into account when converting raw rubric scores to scale scores meant that scripts were not ordered in the way they would be if the ordering relied only on raw rubric scores. A rubric score for a more difficult prompt will convert to a higher scale score than a script with the same rubric score for an easier prompt. As a result, the script for the more difficult prompt will be on a later page in the response book.

An implication of taking prompt difficulty into account is that the judges have to consider the effect of prompt difficulty when deciding what a minimally acceptable response will look like. This will vary from prompt to prompt and adds complexity to the judgement when the response booklet includes a range of prompts.

The second part of the exercise did not include prompt difficulty in the scale score model (all prompts were assumed to be of equal difficulty). In this case, the rank order of prompts according to rubric score is preserved when the conversion to a scale score is made.

7.8.6. Organisation

Each part of the standard-setting exercise was divided into seven sessions with each session focussed on a different level of the literacy learning progressions.

Judges were seated in small groups. The seating arrangement was changed for each session to ensure that each judge worked with a range of others throughout the exercise.

The standard-setting process was introduced and discussed. This included a discussion of the concept of a "minimally acceptable script" and how the standard-setting process involved predicting the kinds of performance level that could be expected from a student working at an LLP level if they were asked to do an e-asTTle writing prompt.

A practice session was carried out to check participants had a strong understanding of the process.

Each standard-setting session started with a general discussion regarding the LLP level under consideration. Judges read the appropriate pages of the LLP document (Ministry of Education, 2010) and then discussed how the level differed from other levels and how this would be manifested in terms of performance on an e-asTTle writing prompt.

Each session was divided into three rounds. The first round was done individually. Each judge was asked to work their way through the book and place a bookmark when they reached a script that they believed was minimally acceptable for a person working at the LLP level.

The second round started with a group discussion of the bookmark decisions made by the individual group members. Each group member shared their decision and provided a rationale for where they had placed the bookmark. Group members were encouraged to consider other judges' perspectives and examine their own decision. At the end of the discussion, each group member placed his or her bookmarks for the second time.

The third round began with a discussion involving all judges. Each group had a chance to report back to the larger group. This was followed by more discussion at the group level. Each judge then made a final placement of the bookmark.

7.8.7. Observations from the first part

During the first part the judges appeared to have a strong understanding of the processes involved in the standard-setting exercise. However, a number of judges felt unhappy with the ordering of the scripts, particularly at the later levels. It was decided that asking judges to take into account the difficulty of a prompt was too difficult when comparing differences between scripts.

This led to a decision to rerun the process, this time using scale scores that did not take prompt difficulty into account.

7.8.8. The second part

The second part of the exercise went smoothly. The judges found the new ordering easier to work with and were able to reach agreement on bookmark placement at each level. Table 21 shows the aWs cut-points confirmed by the standard-setting exercise for each LLP level. The cut-points provided locate the minimum scale score for a student deemed to be at the corresponding LLP level.

Table 21 **Minimum scores for each curriculum level**

| Curriculum level | Cut-point (minimum aWs score in aWs units) |
|------------------|--|
| Early level 1 | 1091.723 |
| At level 1 | 1194.262 |
| Early level 2 | 1249.803 |
| At level 2 | 1344.408 |
| At level 3 | 1500.657 |
| At level 4 | 1565.354 |
| At level 5 | 1707.566 |

7.8.9. Assigning e-asTTle curriculum level descriptors

e-asTTle writing links the aWs scale with curriculum levels through the use of curriculum level descriptors. These split the range of scale scores into six curriculum levels, each of which is further demarcated into beginning (B), proficient (P) and advanced (A) sublevels.

After input from the reference group, a system was adopted where the cut-points generated for the curriculum expectations for the “at level 1”, “at level 2”, “at level 3”, “at level 4” and “at level 5” levels were equated with the midpoint of the 1P, 2P, 3P, 4P and 5P e-asTTle reporting ranges respectively. The midpoints between these proficient stages were then used to define the start and end of each e-asTTle curriculum level. Table 22 shows the calculation used to define the cut-points for the start of each e-asTTle curriculum range

Table 22 **Relationship between e-asTTle curriculum level cut-points and the LLP level from standard setting**

| e-asTTle writing level cut-point | Relationship to LLP cut-point |
|----------------------------------|--|
| 1B | Early level 1* |
| 2B | Average of “at level 1” and “at level 2” |
| 3B | Average of “at level 2” and “at level 3” |
| 4B | Average of “at level 3” and “at level 4” |
| 5B | Average of “at level 4” and “at level 5” |
| 6B | $5B + 2 \times (\text{at level 5} - 5B)$ |

* Once 1P and 1A had been defined, the 1B descriptor was set as the minimum curriculum level descriptor for all scale scores.

The B, P and A sublevels for each curriculum level were then defined by dividing each curriculum range into three equal parts. Table 23 shows the lower and upper limits scale scores for each curriculum level descriptor. Figure 12 presents this graphically and shows the relationship between the e-asTTle curriculum descriptors and the LLP levels.

Table 23 e-asTTle writing curriculum ranges

| | Mean score | Lower score (aWs units) | Upper score (aWs units) |
|---------------------|------------|-------------------------|-------------------------|
| 1 Basic | 1100.9 | -10,000 | 1150.9 |
| 1 Proficient | 1180.5 | 1150.9 | 1210.1 |
| 1 Advanced | 1239.7 | 1210.1 | 1269.3 |
| 2 Basic | 1294.85 | 1269.3 | 1320.4 |
| 2 Proficient | 1345.95 | 1320.4 | 1371.5 |
| 2 Advanced | 1397 | 1371.5 | 1422.5 |
| 3 Basic | 1440.95 | 1422.5 | 1459.4 |
| 3 Proficient | 1477.8 | 1459.4 | 1496.2 |
| 3 Advanced | 1514.6 | 1496.2 | 1533 |
| 4 Basic | 1550.25 | 1533 | 1567.5 |
| 4 Proficient | 1584.75 | 1567.5 | 1602 |
| 4 Advanced | 1619.25 | 1602 | 1636.5 |
| 5 Basic | 1660.2 | 1636.5 | 1683.9 |
| 5 Proficient | 1707.6 | 1683.9 | 1731.3 |
| 5 Advanced | 1755 | 1731.3 | 1778.7 |
| 6 Basic | 1802.4 | 1778.7 | 1826.1 |
| > 6 Basic | 1876.1 | 1826.1 | 10,000 |

Figure 12 Graphical representation of the relationship between the aWs scale, e-asTTle curriculum level descriptors and the LLP levels

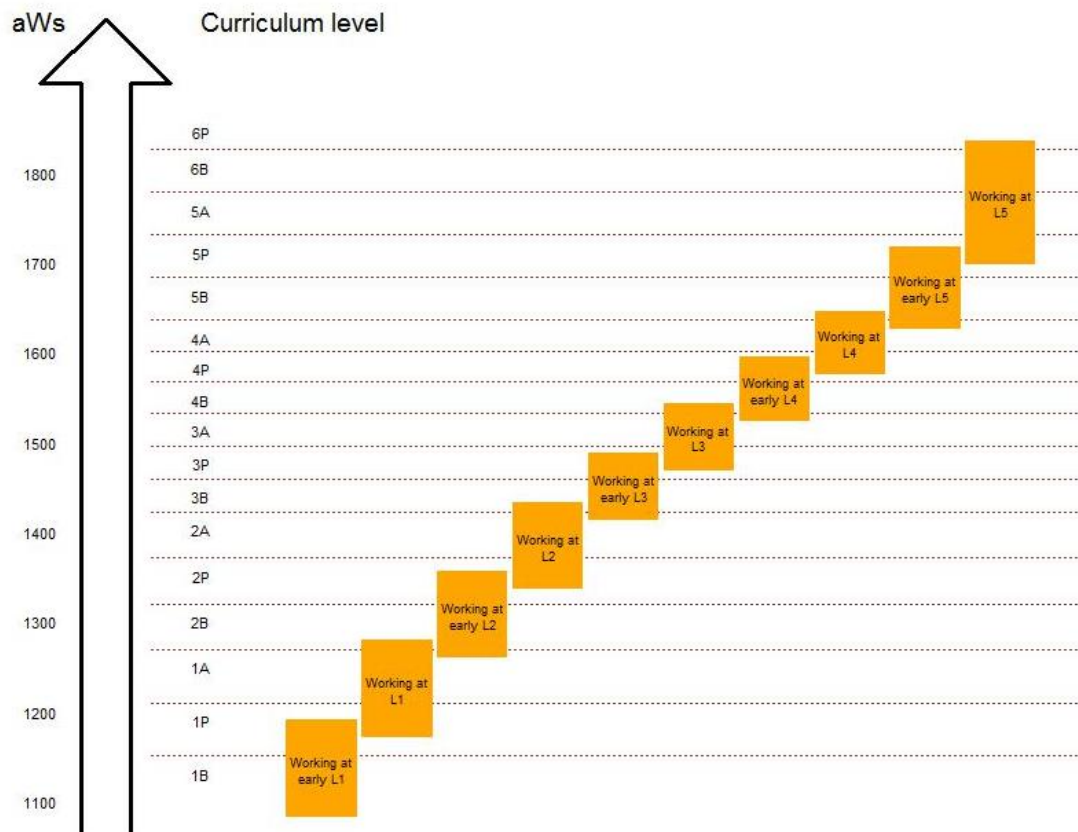
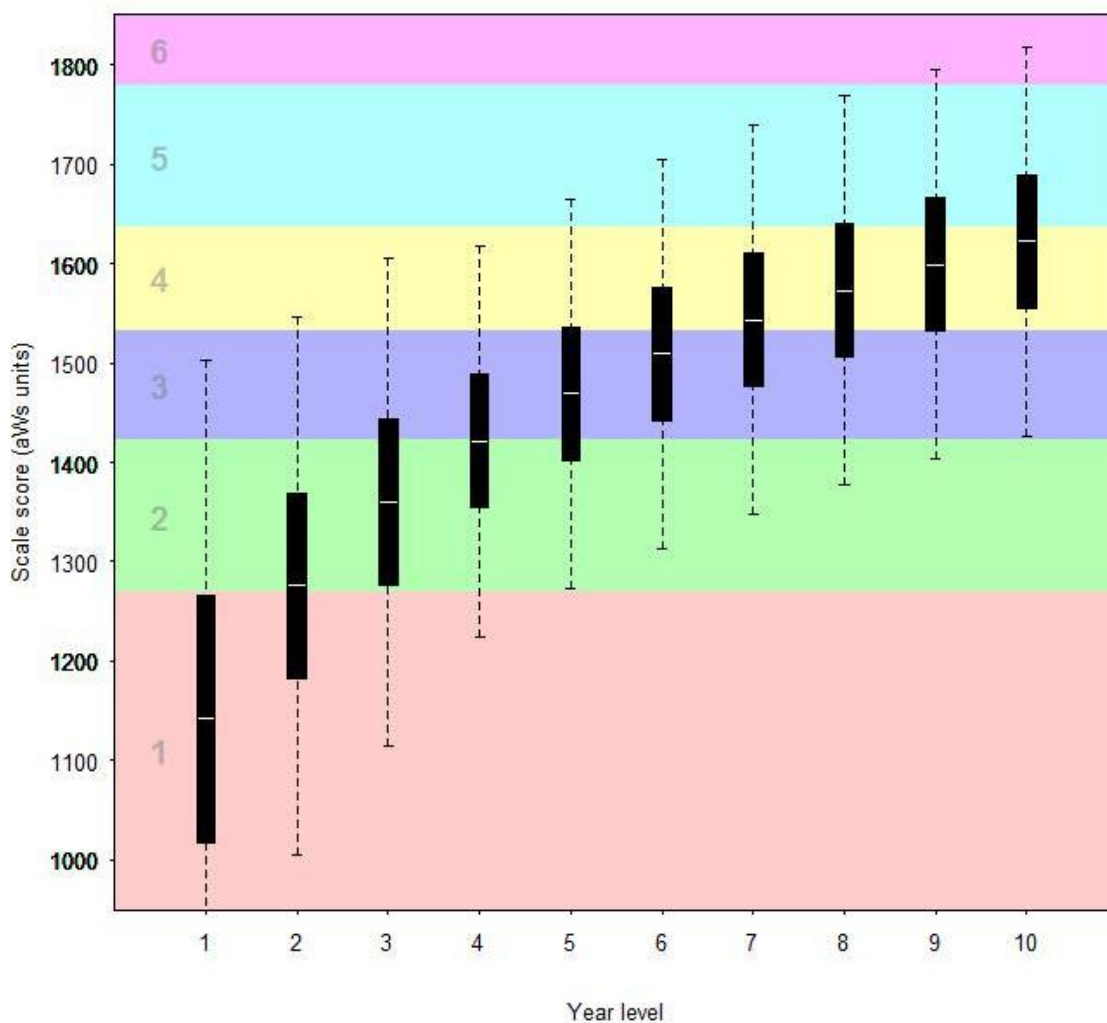


Figure 13 shows the relationship between the achievement of the reference sample at each year level (for Quarter 4) and the cut-points set by the standard-setting exercise. As can be seen, fewer students in the later year levels perform at or above the curriculum expectations set by the standard-setting exercise, compared to students in the earlier year levels.

Figure 13 **The Quarter 4 distribution of scale scores by year level compared to curriculum level expectations**



7.8.10. Curriculum expectations

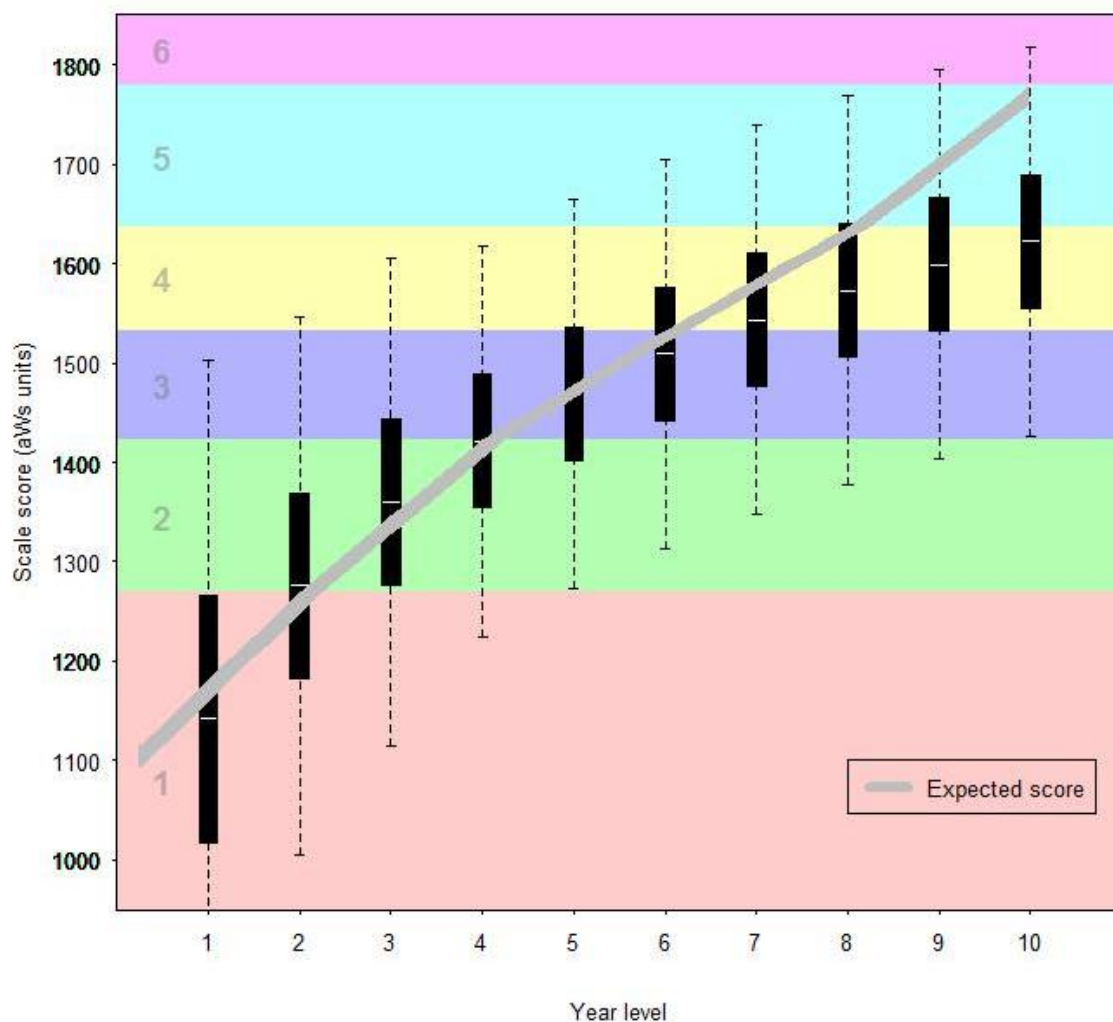
The e-asTTle tool includes a report that compares students' progress with a band of scores showing curriculum expectations by quarter for each year level. The band provides a range of scale scores that match the expected LLP curriculum level for each year level as defined by the standard-setting exercise. Table 24 shows the expected score ranges by quarter.

Table 24 **Expected score by quarter**

| Year | Quarter | Expected score | Lower bound | Upper bound |
|------|---------|----------------|-------------|-------------|
| 1 | 1 | 1102.8 | 1091.7 | 1113.9 |
| 1 | 2 | 1125 | 1113.9 | 1136.1 |
| 1 | 3 | 1147.2 | 1136.1 | 1158.3 |
| 1 | 4 | 1169.4 | 1158.3 | 1180.5 |
| 2 | 1 | 1191.6 | 1180.5 | 1202.7 |
| 2 | 2 | 1213.8 | 1202.7 | 1224.9 |
| 2 | 3 | 1236 | 1224.9 | 1247.1 |
| 2 | 4 | 1258.2 | 1247.1 | 1269.3 |
| 3 | 1 | 1278.875 | 1269.3 | 1288.45 |
| 3 | 2 | 1298.025 | 1288.45 | 1307.6 |
| 3 | 3 | 1317.175 | 1307.6 | 1326.75 |
| 3 | 4 | 1336.325 | 1326.75 | 1345.9 |
| 4 | 1 | 1355.475 | 1345.9 | 1365.05 |
| 4 | 2 | 1374.625 | 1365.05 | 1384.2 |
| 4 | 3 | 1393.775 | 1384.2 | 1403.35 |
| 4 | 4 | 1412.925 | 1403.35 | 1422.5 |
| 5 | 1 | 1429.40625 | 1422.5 | 1436.3125 |
| 5 | 2 | 1443.21875 | 1436.3125 | 1450.125 |
| 5 | 3 | 1457.03125 | 1450.125 | 1463.9375 |
| 5 | 4 | 1470.84375 | 1463.9375 | 1477.75 |
| 6 | 1 | 1484.65625 | 1477.75 | 1491.5625 |
| 6 | 2 | 1498.46875 | 1491.5625 | 1505.375 |
| 6 | 3 | 1512.28125 | 1505.375 | 1519.1875 |
| 6 | 4 | 1526.09375 | 1519.1875 | 1533 |
| 7 | 1 | 1539.46875 | 1533 | 1545.9375 |
| 7 | 2 | 1552.40625 | 1545.9375 | 1558.875 |
| 7 | 3 | 1565.34375 | 1558.875 | 1571.8125 |
| 7 | 4 | 1578.28125 | 1571.8125 | 1584.75 |
| 8 | 1 | 1591.21875 | 1584.75 | 1597.6875 |
| 8 | 2 | 1604.15625 | 1597.6875 | 1610.625 |
| 8 | 3 | 1617.09375 | 1610.625 | 1623.5625 |
| 8 | 4 | 1630.03125 | 1623.5625 | 1636.5 |
| 9 | 1 | 1645.3875 | 1636.5 | 1654.275 |
| 9 | 2 | 1663.1625 | 1654.275 | 1672.05 |
| 9 | 3 | 1680.9375 | 1672.05 | 1689.825 |
| 9 | 4 | 1698.7125 | 1689.825 | 1707.6 |
| 10 | 1 | 1716.4875 | 1707.6 | 1725.375 |
| 10 | 2 | 1734.2625 | 1725.375 | 1743.15 |
| 10 | 3 | 1752.0375 | 1743.15 | 1760.925 |
| 10 | 4 | 1769.8125 | 1760.925 | 1778.7 |

Figure 14 presents the scale score expectations by year level graphically. The box plots are used to show the distributions of scale score by year level for the national reference sample. The distributions shown relate to performance in Quarter 4 of each year.

Figure 14 **Scale score expectations by year level**



Part 3: Scoring and reporting systems for the e-asTTLe writing tool

8. Introduction

The e-asTTLe writing tool converts raw rubric scores entered by users into scale scores. Two levels of scale scores are calculated: a scale score related to the total rubric score (an overall scale score) and a scale score for each of the seven elements assessed by the rubric. In addition, scale scores are linked to curriculum levels. This section describes the processes used by the e-asTTLe writing tool to calculate and report scores, and describes changes made to the reporting functionality in the e-asTTLe application to accommodate the revised writing tool.

8.1. Rubric scores

e-asTTLe writing prompts are scored using the e-asTTLe writing marking rubric. The rubric is divided into seven elements. Each element involves either six or seven scoring categories. Table 25 shows the number of scoring categories associated with each element.

Table 25 **Scoring categories by element**

| Element | Score categories |
|------------------------|------------------|
| Ideas | 1–6 |
| Structure and language | 1–6 |
| Organisation | 1–7 |
| Vocabulary | 1–6 |
| Sentence structure | 1–6 |
| Punctuation | 1–7 |
| Spelling | 1–6 |

The total rubric score is the sum of the rubric scores for each element.

8.2. Scale scores

8.2.1. The overall scale score

The partial credit formulation of the Rasch model (PCM) is used to transform the total rubric score for a student to a score on the e-asTTLe writing scale (the aWs scale). The e-asTTLe tool incorporates the required mathematical functions to make this calculation. The function requires as input the rubric scores a student has achieved on each element and a set of prompt-specific parameters.

Prompt parameters

Each e-asTTLe writing prompt has a set of prompt-specific parameters. These are based on the category thresholds associated with the PCM formulation for that prompt. Parameter values are provided in logits (the unit used by the PCM). Table 26 shows the relationship between prompt parameters and category thresholds for an example element with six scoring categories. Here $\bar{\delta}_{1,2}$ stands for the threshold between categories 1 and 2. Elements with seven categories follow a similar pattern.

Table 26 **The relationship between prompt parameters and modelled category thresholds for a single element**

| Category | 1 | 2 | 3 | 4 | 5 | 6 |
|-----------|--------------------------|---|---|---|---|--------------------------|
| Parameter | $\bar{\delta}_{1,2} - 1$ | $(\bar{\delta}_{1,2} + \bar{\delta}_{2,3})/2$ | $(\bar{\delta}_{2,3} + \bar{\delta}_{3,4})/2$ | $(\bar{\delta}_{3,4} + \bar{\delta}_{4,5})/2$ | $(\bar{\delta}_{4,5} + \bar{\delta}_{5,6})/2$ | $\bar{\delta}_{5,6} + 1$ |

Each e-asTTLe writing prompt is associated with 44 parameters (one parameter per scoring category). Table 27 provides an example of the prompt parameters for an e-asTTLe writing prompt (“Dogs at the beach”).

Table 27 **Example of prompt parameters (in logits)**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|--------|-------|-------|------|------|------|------|
| Ideas | -7.62 | -4.72 | -1.23 | 1.46 | 3.36 | 5.17 | |
| Structure | -6.53 | -3.42 | -0.18 | 1.98 | 3.89 | 5.76 | |
| Organisation | -7.14 | -4.32 | -1.39 | 0.42 | 2.19 | 4.77 | 7.28 |
| Vocabulary | -6.71 | -4.02 | -0.86 | 1.49 | 3.25 | 5.13 | |
| Sentence | -6.49 | -3.53 | -0.44 | 1.56 | 3.45 | 5.50 | |
| Punctuation | -5.53 | -3.75 | -1.72 | 0.78 | 2.97 | 4.60 | 6.57 |
| Spelling | -12.28 | -7.42 | -2.14 | 0.42 | 2.63 | 4.69 | |

8.2.2. Calculating overall scale scores

Overall scale scores are calculated using a score conversion algorithm. The algorithm uses the unconditional maximum likelihood approach outlined in Wright and Masters (1982) to calculate the expected scale location for a response on the prompt with the given rubric scores. The e-asTTLe application passes the prompt parameters and overall rubric score to the algorithm function in order to calculate the scale score. Before calculating the scale score the prompt thresholds are converted to the original model thresholds for the prompt. For instance, $\bar{\delta}_{1,2} = P1 + 1$ and $\bar{\delta}_{2,3} = 2 \times P2 - \bar{\delta}_{1,2}$ and so on.

Treatment of maximum and minimum scores

It is not possible to calculate an overall scale score when the total rubric score is either the minimum or maximum possible rubric score (7 or 44 respectively). In this case an estimate of the

scale score is provided by passing the algorithm the minimum score + 0.3 in the case of a minimum score and maximum score - 0.3 in the case of a maximum score.

The calculation of the standard error of measurement

The standard error of measurement (SEM) is reported for each overall scale score. This is calculated at the same time as the total scale score. The SEM includes a component of error associated with the imprecision of the model and a component associated with an estimate of the standard error associated with markers. This later estimate was derived from the modelling of marker harshness undertaken when constructing the scale. It is set at 0.48 logits.

$$S.E.M. = \sqrt{(s.e.modelled\ score)^2 + (s.e.markers)^2}$$

The standard error is greatest for scale scores associated with total rubric scores close to the minimum or maximum possible rubric scores.

8.2.3. Calculating scale scores for elements

The scale score for an element is given by the corresponding prompt parameter for that category. For instance, for the prompt shown in Table 27, a rubric score of 4 for ideas would be given a scale score of 1.46 logits. This is then converted to aWs units. The scale scores for elements represent the most probable place on the scale for a student scoring in that category.

8.2.4. Conversion of logit scores to scale scores

All scale scores, including the standard errors, are reported using e-asTTLe writing scale units (aWs units). These are a transformation of the logit scores used for calculation and parameterisation.

The transformation used is:

$$\text{Score (aWs units)} = (\text{logit} - 0.379234874545207)/1.63840555986589 * 100 + 1500$$

8.2.5. Calculating curriculum level scores

Curriculum levels have been linked to the e-asTTLe writing scale. This allows a best-fit curriculum level to be reported for a given scale score. A look-up process is used to find the curriculum level associated with a given scale score. 0 shows the link between scale scores (in logits and aWs units) and the associated curriculum level descriptor.

Table 28 **Scale score to curriculum level conversion**

| | Mean score (aWs units) | Lower score (logit) | Upper score (logit) | Lower score (aWs units) | Upper score (aWs units) |
|---------------------|---------------------------|------------------------|------------------------|----------------------------|----------------------------|
| 1 Basic | 1100.9 | -10000 | -5.34 | -10,000 | 1150.9 |
| 1 Proficient | 1180.5 | -5.34 | -4.37 | 1150.9 | 1210.1 |
| 1 Advanced | 1239.7 | -4.37 | -3.4 | 1210.1 | 1269.3 |
| 2 Basic | 1294.85 | -3.4 | -2.563 | 1269.3 | 1320.4 |
| 2 Proficient | 1345.95 | -2.563 | -1.727 | 1320.4 | 1371.5 |
| 2 Advanced | 1397 | -1.727 | -0.89 | 1371.5 | 1422.5 |
| 3 Basic | 1440.95 | -0.89 | -0.287 | 1422.5 | 1459.4 |
| 3 Proficient | 1477.8 | -0.287 | 0.317 | 1459.4 | 1496.2 |
| 3 Advanced | 1514.6 | 0.317 | 0.92 | 1496.2 | 1533 |
| 4 Basic | 1550.25 | 0.92 | 1.485 | 1533 | 1567.5 |
| 4 Proficient | 1584.75 | 1.485 | 2.05 | 1567.5 | 1602 |
| 4 Advanced | 1619.25 | 2.05 | 2.615 | 1602 | 1636.5 |
| 5 Basic | 1660.2 | 2.615 | 3.392 | 1636.5 | 1683.9 |
| 5 Proficient | 1707.6 | 3.392 | 4.168 | 1683.9 | 1731.3 |
| 5 Advanced | 1755 | 4.168 | 4.945 | 1731.3 | 1778.7 |
| 6 Basic | 1802.4 | 4.945 | 5.722 | 1778.7 | 1826.1 |
| > 6 Basic | 1876.1 | 5.722 | 10,000 | 1826.1 | 10,000 |

8.3. Normative information

The e-asTTle writing application also reports normative information. This is displayed graphically. Most of these norms have been constructed from data collected during the development trials for the revised e-asTTle writing. Where the sample used in the trial was not deemed to be big enough to support subgroup norms, patterns in data from the existing norms have been used to provide estimated values. This was not possible at Years 1 to 3, where there was no existing data. When reference information is not available, the report will still be produced, but will not show any reference information.

The reference information is provided in quarter-year intervals. The e-asTTle application automatically adjusts the norming information to reflect the quarter the test was administered.

The normative information sourced from data collected as part of the development process for the new e-asTTle writing is for year level, year level by gender and year level by ethnicity. Reference information reported for “English at Home”, “Region” and “Schools Like Us” is based on achievement data from the original e-asTTle writing.

8.4. Changes to e-asTTLe reports

A number of changes were made to the reports available on the e-asTTLe application to accommodate the revised e-asTTLe writing tool. The main changes are described in the following sections.

8.4.1. The Individual Learning Pathways Report

The Individual Learning Pathways Report is designed to provide an overview of an individual student’s performance on a writing test. This report is available to both teachers and students. The Individual Learning Pathways Report used in the new version of e-asTTLe writing differs in several ways from the way it is presented in other learning areas and how it was used in the original e-asTTLe writing. These differences and the rationale for them are described in Table 29.

Table 29 **Changes to the Individual Learning Pathways Report**

| Change to the report | Rationale |
|---|--|
| The rubric scores for each element (curriculum function score) are reported. | This allows the reader to see exactly which rubric category was awarded for each element along with the maximum score available for the element. |
| The margin of error associated with the student’s scale score is shown as a plus or minus (\pm) range. The size of the error is also represented by the width of the circle in the graphic. | It is important to provide an indication of precision to report readers. This allows them to take imprecision into account when comparing scores. |
| Scale scores for the individual elements and their corresponding “dials” are not shown. | At the individual student level, these element scores have limited precision and are difficult to interpret. A curriculum level is still provided for each element to provide a broadly comparable indication of the level of performance. |
| Scores for shallow and deep features are not provided. | This reflects a belief that it is impossible to categorise any of the elements assessed by e-asTTLe as representing only shallow or deep features of writing. |

8.4.2. The Curriculum Levels Report

The Curriculum Levels Report aggregates the curriculum levels on each element across a group of learners. This report has been updated to show rubric scores as well as curriculum levels.

Table 30 describes the changes made to this report, and the rationale for them.

Table 30 **Changes to the Curriculum Levels Report**

| Change to the report | Rationale |
|--|--|
| Two sets of bar charts are now presented, one by curriculum level and one by rubric score. | The curriculum levels associated with the rubric scores are imprecise and spread across the scale. The rubric scores name the actual category achieved by the students on the element. |

References

- Andrich, D. (2005). *A report to the curriculum council of Western Australia regarding assessment for tertiary selection*. Retrieved March, 2011.
http://www.curriculum.wa.edu.au/internet/Publications/Reports/General_Reports
- Bawarshi, A., & Reiff, M. (2010). *Genre: An introduction to history, theory, research, and pedagogy*. West Lafayette, In: Parlor Press.
- Bazerman, C. (2010). Series editor's preface. In A. S. Bawarshi & M. J. Reiff, *Genre: An introduction to history, theory, research, and pedagogy* (pp. xi–xii). West Lafayette, In: Parlor Press.
- Bazerman, C. (2004). Speech acts, genres, and activity systems: How texts organise activity and people. In C. Bazerman & P. Prior (Eds.), *What writing does and how it does it: An introduction to analyzing texts and textual practices* (pp. 309–339). New York, NY: Lawrence Erlbaum.
- Bazerman, C., Bonini, A., & Figueiredo, D. (Eds.). (2009). *Genre in a changing world*. West Lafayette, In: Parlor Press.
- Cervetti, G. N., & Pearson, P.D. (2012). Commentary. Reading, writing, and thinking like a scientist. *Journal of Adolescent and Adult Literacy*, 55(7), 580–586.
- Chapman, M. L. (1999). Situated, social, active: Rewriting genre in the elementary classroom. *Written Communication*, 16, 469–90.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Dean, D. (2008). *Genre theory: Teaching, writing, and being*. Urbana, Ill: National Council of Teachers of English.
- Humphry, S., & Heldsinger, S. (2009). *Do rubrics help to inform and direct teaching practice?* Retrieved March, 2011 from
http://research.acer.edu.au/cgi/viewcontent.cgi?article=1051&context=research_conference
- Linacre, J. M. (1994). *Many-facet Rasch measurement*. Chicago: MESA Press.
- Linacre, J. M. (2010) *Facets Rasch measurement computer program, version 3.67.1*. Chicago: Winsteps.com
- Moje, E.B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent and Adult Literacy*, 52(2), 96–107.
- Ministry of Education. (2010). *The literacy learning progressions: Meeting the reading and writing demands of the curriculum*. Wellington: Learning Media.
- Ministry of Education & NZCER. (2012). *e-asTTle writing (revised)*. Wellington: Ministry of Education.
- Purcell-Gates, V., Duke, N.K., & Martineau, J.A. (2007). Learning to read and write genre-specific text: Roles in authentic experience and explicit teaching. *Reading Research Quarterly*, 42(1), 8–45.
- Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics approach*. Mahwah, NJ: Lawrence Erlbaum.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Appendix B: Prompt numbers, names and types

Table 31 describes the correspondence between prompt numbers, names and types.

Table 31 **Prompt numbers and names**

| Prompt number | Prompt name | Prompt type number | Prompt type descriptor |
|---------------|---|--------------------|------------------------|
| 1 | Dogs at the beach | 1 | Describe |
| 2 | Jumping girl | 1 | Describe |
| 3 | Girl | 1 | Describe |
| 4 | Adult and child | 2 | Describe |
| 5 | Whānau and family time | 7 | Recount |
| 6 | Time with friends | 7 | Recount |
| 7 | What I did well | 7 | Recount |
| 8 | Caring for planet Earth | 6 | Explain |
| 9 | The life cycle of the Monarch butterfly | 3 | Describe |
| 10 | A frog life cycle | 3 | Describe |
| 11 | Music is more important than sport | 4 | Persuade |
| 12 | It is wrong to fight | 4 | Persuade |
| 13 | The referee is always right | 4 | Persuade |
| 14 | A special place in the community | 6 | Explain |
| 15 | Good friends | 6 | Explain |
| 16 | A community facility | 6 | Explain |
| 17 | The market | 2 | Describe |
| 18 | Stick insect | 2 | Describe |
| 19 | The day things started disappearing | 5 | Narrate |
| 20 | The bush | 5 | Narrate |
| 21 | I heard a whisper but no one was there | 5 | Narrate |

Appendix C: Prompt parameters

The tables below provide the prompt parameters (in logits) for each e-asTTLe writing prompt. These parameters are used by the e-asTTLe scoring algorithm to calculate scale scores.

Table 32 **Jumping Girl**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.34312 | -4.44824 | -0.95893 | 1.735134 | 3.637277 | 5.439778 | |
| Structure | -6.25686 | -3.14549 | 0.089241 | 2.254885 | 4.165347 | 6.03352 | |
| Organisation | -6.86305 | -4.04619 | -1.11407 | 0.697911 | 2.46588 | 5.043889 | 7.55065 |
| Vocabulary | -6.43236 | -3.7434 | -0.58528 | 1.76534 | 3.525907 | 5.405009 | |
| Sentence | -6.21399 | -3.25654 | -0.16253 | 1.829367 | 3.727626 | 5.770552 | |
| Punctuation | -5.26057 | -3.4743 | -1.44594 | 1.052672 | 3.241957 | 4.871851 | 6.568973 |
| Spelling | -12.0105 | -7.14205 | -1.86751 | 0.689258 | 2.902007 | 4.964067 | |

Table 33 **Girl**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.50477 | -4.60989 | -1.12058 | 1.573488 | 3.475631 | 5.278132 | |
| Structure | -6.41851 | -3.30713 | -0.0724 | 2.093239 | 4.003701 | 5.871874 | |
| Organisation | -7.0247 | -4.20784 | -1.27572 | 0.536265 | 2.304234 | 4.882243 | 7.389004 |
| Vocabulary | -6.594 | -3.90504 | -0.74693 | 1.603694 | 3.364261 | 5.243363 | |
| Sentence | -6.37564 | -3.41819 | -0.32417 | 1.667721 | 3.56598 | 5.608906 | |
| Punctuation | -5.42222 | -3.63595 | -1.60758 | 0.891026 | 3.080311 | 4.710205 | 6.407327 |
| Spelling | -12.1722 | -7.30369 | -2.02915 | 0.527612 | 2.740361 | 4.802421 | |

Table 34 **Adult and child**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.53748 | -4.6426 | -1.15329 | 1.540773 | 3.442915 | 5.245417 | |
| Structure | -6.45122 | -3.33985 | -0.10512 | 2.060524 | 3.970986 | 5.839158 | |
| Organisation | -7.05742 | -4.24055 | -1.30843 | 0.50355 | 2.271518 | 4.849527 | 7.356288 |
| Vocabulary | -6.62672 | -3.93776 | -0.77964 | 1.570979 | 3.331546 | 5.210648 | |
| Sentence | -6.40835 | -3.4509 | -0.35689 | 1.635006 | 3.533264 | 5.576191 | |

| | | | | | | | |
|-------------|----------|----------|----------|----------|----------|----------|----------|
| Punctuation | -5.45494 | -3.66866 | -1.6403 | 0.858311 | 3.047596 | 4.677489 | 6.374611 |
| Spelling | -12.2049 | -7.33641 | -2.06187 | 0.494896 | 2.707645 | 4.769706 | |

Table 35 **Whānau and family time**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.79444 | -4.89956 | -1.41025 | 1.283818 | 3.18596 | 4.988461 | |
| Structure | -6.70818 | -3.59681 | -0.36208 | 1.803568 | 3.71403 | 5.582203 | |
| Organisation | -7.31437 | -4.49751 | -1.56539 | 0.246594 | 2.014563 | 4.592572 | 7.099333 |
| Vocabulary | -6.88367 | -4.19471 | -1.0366 | 1.314023 | 3.07459 | 4.953692 | |
| Sentence | -6.66531 | -3.70786 | -0.61385 | 1.37805 | 3.276309 | 5.319236 | |
| Punctuation | -5.71189 | -3.92562 | -1.89725 | 0.601356 | 2.790641 | 4.420534 | 6.117656 |
| Spelling | -12.4618 | -7.59336 | -2.31882 | 0.237941 | 2.45069 | 4.51275 | |

Table 36 **Time with friends**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -8.00018 | -5.1053 | -1.61599 | 1.078073 | 2.980215 | 4.782716 | |
| Structure | -6.91393 | -3.80255 | -0.56782 | 1.597823 | 3.508285 | 5.376458 | |
| Organisation | -7.52012 | -4.70325 | -1.77113 | 0.040849 | 1.808818 | 4.386827 | 6.893588 |
| Vocabulary | -7.08942 | -4.40046 | -1.24234 | 1.108278 | 2.868845 | 4.747947 | |
| Sentence | -6.87105 | -3.9136 | -0.81959 | 1.172305 | 3.070564 | 5.113491 | |
| Punctuation | -5.91764 | -4.13137 | -2.103 | 0.395611 | 2.584896 | 4.214789 | 5.911911 |
| Spelling | -12.6676 | -7.79911 | -2.52457 | 0.032196 | 2.244945 | 4.307005 | |

Table 37 **What I did well**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.82772 | -4.93284 | -1.44354 | 1.250531 | 3.152673 | 4.955175 | |
| Structure | -6.74147 | -3.63009 | -0.39536 | 1.770282 | 3.680744 | 5.548916 | |
| Organisation | -7.34766 | -4.5308 | -1.59867 | 0.213308 | 1.981276 | 4.559285 | 7.066046 |
| Vocabulary | -6.91696 | -4.228 | -1.06989 | 1.280736 | 3.041304 | 4.920406 | |
| Sentence | -6.6986 | -3.74115 | -0.64713 | 1.344764 | 3.243022 | 5.285949 | |
| Punctuation | -5.74518 | -3.95891 | -1.93054 | 0.568069 | 2.757354 | 4.387247 | 6.084369 |
| Spelling | -12.4951 | -7.62665 | -2.35211 | 0.204654 | 2.417403 | 4.479464 | |

Table 38 **Caring for planet earth**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.75733 | -4.86244 | -1.37314 | 1.32093 | 3.223072 | 5.025574 | |
| Structure | -6.67107 | -3.55969 | -0.32496 | 1.840681 | 3.751143 | 5.619315 | |
| Organisation | -7.27726 | -4.4604 | -1.52827 | 0.283707 | 2.051675 | 4.629684 | 7.136445 |
| Vocabulary | -6.84656 | -4.1576 | -0.99949 | 1.351136 | 3.111703 | 4.990805 | |
| Sentence | -6.6282 | -3.67075 | -0.57673 | 1.415163 | 3.313421 | 5.356348 | |
| Punctuation | -5.67478 | -3.88851 | -1.86014 | 0.638468 | 2.827753 | 4.457646 | 6.154768 |
| Spelling | -12.4247 | -7.55625 | -2.28171 | 0.275053 | 2.487802 | 4.549863 | |

Table 39 **The life cycle of Monarch butterflies**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.31369 | -4.41881 | -0.9295 | 1.764569 | 3.666711 | 5.469212 | |
| Structure | -6.22743 | -3.11605 | 0.118675 | 2.284319 | 4.194781 | 6.062954 | |
| Organisation | -6.83362 | -4.01676 | -1.08464 | 0.727345 | 2.495314 | 5.073323 | 7.580084 |
| Vocabulary | -6.40292 | -3.71396 | -0.55585 | 1.794774 | 3.555341 | 5.434443 | |
| Sentence | -6.18456 | -3.22711 | -0.13309 | 1.858801 | 3.75706 | 5.799986 | |
| Punctuation | -5.23114 | -3.44487 | -1.4165 | 1.082106 | 3.271391 | 4.901285 | 6.598407 |
| Spelling | -11.9811 | -7.11261 | -1.83807 | 0.718692 | 2.931441 | 4.993501 | |

Table 40 **A frog life cycle**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -6.88161 | -3.98673 | -0.49742 | 2.196643 | 4.098785 | 5.901286 | |
| Structure | -5.79536 | -2.68398 | 0.550749 | 2.716393 | 4.626855 | 6.495028 | |
| Organisation | -6.40155 | -3.58468 | -0.65256 | 1.159419 | 2.927388 | 5.505397 | 8.012158 |
| Vocabulary | -5.97085 | -3.28189 | -0.12377 | 2.226848 | 3.987415 | 5.866517 | |
| Sentence | -5.75248 | -2.79503 | 0.29898 | 2.290875 | 4.189134 | 6.23206 | |
| Punctuation | -4.79907 | -3.0128 | -0.98443 | 1.51418 | 3.703465 | 5.333359 | 7.030481 |
| Spelling | -11.549 | -6.68054 | -1.406 | 1.150766 | 3.363515 | 5.425575 | |

Table 41 **Music is more important than sport**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -6.9155 | -4.02062 | -0.53131 | 2.162753 | 4.064895 | 5.867396 | |
| Structure | -5.82925 | -2.71787 | 0.516859 | 2.682503 | 4.592965 | 6.461138 | |
| Organisation | -6.43544 | -3.61857 | -0.68645 | 1.125529 | 2.893498 | 5.471507 | 7.978268 |
| Vocabulary | -6.00474 | -3.31578 | -0.15766 | 2.192958 | 3.953525 | 5.832627 | |
| Sentence | -5.78637 | -2.82892 | 0.26509 | 2.256985 | 4.155244 | 6.198171 | |
| Punctuation | -4.83296 | -3.04669 | -1.01832 | 1.480291 | 3.669576 | 5.299469 | 6.996591 |
| Spelling | -11.5829 | -6.71443 | -1.43989 | 1.116876 | 3.329625 | 5.391685 | |

Table 42 **It is wrong to fight**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.36347 | -4.46858 | -0.97928 | 1.71479 | 3.616933 | 5.419434 | |
| Structure | -6.27721 | -3.16583 | 0.068897 | 2.234541 | 4.145003 | 6.013176 | |
| Organisation | -6.8834 | -4.06654 | -1.13441 | 0.677567 | 2.445536 | 5.023545 | 7.530306 |
| Vocabulary | -6.4527 | -3.76374 | -0.60563 | 1.744996 | 3.505563 | 5.384665 | |
| Sentence | -6.23434 | -3.27689 | -0.18287 | 1.809023 | 3.707282 | 5.750208 | |
| Punctuation | -5.28092 | -3.49465 | -1.46628 | 1.032328 | 3.221613 | 4.851507 | 6.548629 |
| Spelling | -12.0309 | -7.16239 | -1.88785 | 0.668914 | 2.881663 | 4.943723 | |

Table 43 **The referee is always right**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -6.78579 | -3.89091 | -0.4016 | 2.292466 | 4.194608 | 5.997109 | |
| Structure | -5.69953 | -2.58816 | 0.646572 | 2.812216 | 4.722678 | 6.590851 | |
| Organisation | -6.30572 | -3.48886 | -0.55674 | 1.255242 | 3.023211 | 5.60122 | 8.107981 |
| Vocabulary | -5.87503 | -3.18607 | -0.02795 | 2.322671 | 4.083238 | 5.96234 | |
| Sentence | -5.65666 | -2.69921 | 0.394803 | 2.386698 | 4.284957 | 6.327883 | |
| Punctuation | -4.70324 | -2.91697 | -0.88861 | 1.610003 | 3.799288 | 5.429182 | 7.126304 |
| Spelling | -11.4532 | -6.58472 | -1.31018 | 1.246589 | 3.459338 | 5.521398 | |

Table 44 **A special place in the community**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.38478 | -4.48989 | -1.00059 | 1.69348 | 3.595623 | 5.398124 | |
| Structure | -6.29852 | -3.18714 | 0.047587 | 2.213231 | 4.123693 | 5.991866 | |
| Organisation | -6.90471 | -4.08785 | -1.15572 | 0.656257 | 2.424226 | 5.002235 | 7.508996 |
| Vocabulary | -6.47401 | -3.78505 | -0.62694 | 1.723686 | 3.484253 | 5.363355 | |
| Sentence | -6.25565 | -3.2982 | -0.20418 | 1.787713 | 3.685972 | 5.728898 | |
| Punctuation | -5.30223 | -3.51596 | -1.48759 | 1.011018 | 3.200303 | 4.830197 | 6.527319 |
| Spelling | -12.0522 | -7.1837 | -1.90916 | 0.647604 | 2.860353 | 4.922413 | |

Table 45 **Good friends**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.97103 | -5.07615 | -1.58684 | 1.107225 | 3.009367 | 4.811869 | |
| Structure | -6.88477 | -3.7734 | -0.53867 | 1.626976 | 3.537438 | 5.40561 | |
| Organisation | -7.49096 | -4.6741 | -1.74198 | 0.070002 | 1.83797 | 4.415979 | 6.92274 |
| Vocabulary | -7.06027 | -4.37131 | -1.21319 | 1.137431 | 2.897998 | 4.7771 | |
| Sentence | -6.8419 | -3.88445 | -0.79044 | 1.201458 | 3.099717 | 5.142643 | |
| Punctuation | -5.88848 | -4.10221 | -2.07385 | 0.424763 | 2.614048 | 4.243942 | 5.941064 |
| Spelling | -12.6384 | -7.76996 | -2.49542 | 0.061348 | 2.274097 | 4.336158 | |

Table 46 **A community facility**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.04073 | -4.14585 | -0.65654 | 2.037523 | 3.939665 | 5.742167 | |
| Structure | -5.95447 | -2.8431 | 0.39163 | 2.557274 | 4.467736 | 6.335908 | |
| Organisation | -6.56067 | -3.7438 | -0.81168 | 1.0003 | 2.768268 | 5.346277 | 7.853038 |
| Vocabulary | -6.12997 | -3.44101 | -0.28289 | 2.067729 | 3.828296 | 5.707398 | |
| Sentence | -5.9116 | -2.95415 | 0.13986 | 2.131756 | 4.030015 | 6.072941 | |
| Punctuation | -4.95819 | -3.17191 | -1.14355 | 1.355061 | 3.544346 | 5.17424 | 6.871362 |
| Spelling | -11.7081 | -6.83966 | -1.56512 | 0.991646 | 3.204395 | 5.266456 | |

Table 47 **The market**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.71611 | -4.82123 | -1.33192 | 1.362143 | 3.264285 | 5.066786 | |
| Structure | -6.62986 | -3.51848 | -0.28375 | 1.881894 | 3.792356 | 5.660528 | |
| Organisation | -7.23605 | -4.41918 | -1.48706 | 0.324919 | 2.092888 | 4.670897 | 7.177658 |
| Vocabulary | -6.80535 | -4.11639 | -0.95827 | 1.392348 | 3.152915 | 5.032017 | |
| Sentence | -6.58698 | -3.62953 | -0.53552 | 1.456375 | 3.354634 | 5.397561 | |
| Punctuation | -5.63357 | -3.84729 | -1.81893 | 0.679681 | 2.868966 | 4.498859 | 6.195981 |
| Spelling | -12.3835 | -7.51504 | -2.2405 | 0.316266 | 2.529015 | 4.591075 | |

Table 48 **Stick insect**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.35809 | -4.46321 | -0.9739 | 1.720166 | 3.622308 | 5.424809 | |
| Structure | -6.27183 | -3.16046 | 0.074272 | 2.239916 | 4.150378 | 6.018551 | |
| Organisation | -6.87802 | -4.06116 | -1.12904 | 0.682942 | 2.450911 | 5.02892 | 7.535681 |
| Vocabulary | -6.44733 | -3.75837 | -0.60025 | 1.750371 | 3.510938 | 5.39004 | |
| Sentence | -6.22896 | -3.27151 | -0.1775 | 1.814398 | 3.712657 | 5.755583 | |
| Punctuation | -5.27554 | -3.48927 | -1.46091 | 1.037703 | 3.226988 | 4.856882 | 6.554004 |
| Spelling | -12.0255 | -7.15702 | -1.88248 | 0.674289 | 2.887038 | 4.949098 | -12.0255 |

Table 49 **The day things started disappearing**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.12908 | -4.2342 | -0.74489 | 1.949178 | 3.85132 | 5.653821 | |
| Structure | -6.04282 | -2.93145 | 0.303285 | 2.468929 | 4.379391 | 6.247563 | |
| Organisation | -6.64901 | -3.83215 | -0.90003 | 0.911954 | 2.679923 | 5.257932 | 7.764693 |
| Vocabulary | -6.21831 | -3.52935 | -0.37124 | 1.979383 | 3.73995 | 5.619052 | |
| Sentence | -5.99995 | -3.0425 | 0.051515 | 2.04341 | 3.941669 | 5.984596 | |
| Punctuation | -5.04653 | -3.26026 | -1.23189 | 1.266716 | 3.456001 | 5.085894 | 6.783016 |
| Spelling | -11.7965 | -6.928 | -1.65346 | 0.903301 | 3.11605 | 5.17811 | |

Table 50 **The bush**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.36444 | -4.46955 | -0.98025 | 1.71382 | 3.615963 | 5.418464 | |
| Structure | -6.27818 | -3.1668 | 0.067927 | 2.233571 | 4.144033 | 6.012206 | |
| Organisation | -6.88437 | -4.06751 | -1.13538 | 0.676597 | 2.444566 | 5.022575 | 7.529336 |
| Vocabulary | -6.45367 | -3.76471 | -0.6066 | 1.744026 | 3.504593 | 5.383695 | |
| Sentence | -6.23531 | -3.27786 | -0.18384 | 1.808053 | 3.706312 | 5.749238 | |
| Punctuation | -5.28189 | -3.49562 | -1.46725 | 1.031358 | 3.220643 | 4.850537 | 6.547659 |
| Spelling | -12.0318 | -7.16336 | -1.88882 | 0.667944 | 2.880693 | 4.942753 | |

Table 51 **I heard a whisper but no one was there**

| Element | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
|--------------|----------|----------|----------|----------|----------|----------|----------|
| Ideas | -7.15843 | -4.26354 | -0.77424 | 1.91983 | 3.821972 | 5.624474 | |
| Structure | -6.07217 | -2.96079 | 0.273937 | 2.439581 | 4.350043 | 6.218215 | |
| Organisation | -6.67836 | -3.8615 | -0.92937 | 0.882607 | 2.650575 | 5.228584 | 7.735345 |
| Vocabulary | -6.24766 | -3.5587 | -0.40059 | 1.950035 | 3.710603 | 5.589705 | |
| Sentence | -6.0293 | -3.07185 | 0.022167 | 2.014063 | 3.912321 | 5.955248 | |
| Punctuation | -5.07588 | -3.28961 | -1.26124 | 1.237368 | 3.426653 | 5.056546 | 6.753668 |
| Spelling | -11.8258 | -6.95735 | -1.68281 | 0.873953 | 3.086702 | 5.148763 | |