

e-asTTle Reading and the Year 8 National Standard

This material is provisional and under review. Overall Teacher Judgments should be supported by a range of evidence, as different assessment methods look at different aspects of student work.

The purpose of this work is to inform overall teacher judgement of student performance against the ‘by the end of year 8’ national standard in reading¹. In particular it will help teachers who use the e-asTTle² reading assessment tool as part of the evidence informing their overall teacher judgement. Those using this material to help them make teacher judgements using e-asTTle should bear in mind the important notes on page 4.

The work

A team of 12 experienced teachers and literacy professional development facilitators worked through sets of anonymised e-asTTle records of student (called ‘scripts’ in what follows). Each script consisted of a print-out of test questions (including associated reading texts) and an individual student’s responses to those questions, with a marking guide and scoring key for the test in question.

Because in e-asTTle tests are customised by teachers from a large bank of pre-calibrated items, there were up to 10 different test versions to be considered during one script scrutiny exercise. This meant that results could be generalised to e-asTTle as a whole, and were not confined to the outcomes from any particular customised test. All scripts and scoring algorithms used were taken from the recalibrated version of e-asTTle reading (released in April 2010).

Initially, the ‘by the end of year 8’ standard was introduced. This was followed by a description of how data collected using the e-asTTle assessment tool could reveal features of the standard. During this process the group cross-referenced the indicators used to assess the reading with the reading standard to identify what characteristics of student reading were assessed using e-asTTle. Then, after a discussion of the definitions of the standard’s four reporting bands (‘well below’, ‘below’, ‘at’ and ‘above’³) and a practice attempt at rating a script, the experts made independent judgements on a pack of six scripts each. Altogether 30 scripts were rated against the standard, with each being rated by two separate judges.

The teachers and other judges made their decisions independently, so that a range of judgements for a given level of performance was captured. This means that for any one piece of evidence describing student performance (such as the e-asTTle assessment), only the likelihood of that piece of evidence being judged as ‘well below’, ‘below’, ‘at’ or ‘above’ the relevant national standard can be provided.

¹ See <http://nzcurriculum.tki.org.nz/National-Standards/Reading-and-writing-standards/The-standards/End-of-year-8>

² See http://e-asttle.tki.org.nz/about_e_asttle

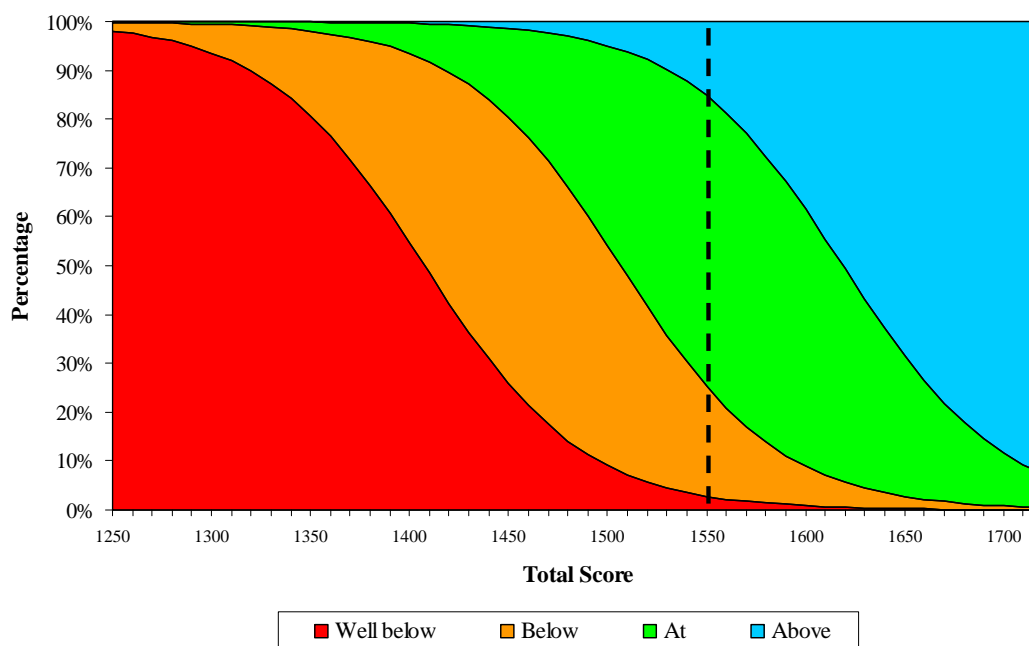
³ See <http://assessment.tki.org.nz/Effective-use-of-evidence/Overall-teacher-judgement-OTJ/A-student-s-achievement>

The number of scripts used in this work, (30), was modest and the work has yet to be replicated on a larger scale. The Ministry has done similar work investigating the alignment of commonly-used assessment tools to the reading, writing and mathematics standards. It intends to ensure that this work is ongoing, with fresh information being published as it becomes available.

Results

Student achievement in e-asTTle is measured by a score on a scale derived from the marking of the student's performance. The results given here describe how this total scale score aligns with the 'by the end of year 8' national standard in reading. The 'area graph' below shows the percentage of scripts likely to be judged in each of the four reporting bands against the total e-asTTle score. It was produced using a statistical modelling technique applied to the collected data.

**Percentages at Different Reporting bands Depending on Total Score
(Area graph)**



To read the graph, consider a vertical line above a total score of, for example, 1550 (illustrated). Looking at the areas which the line intersects, we can conclude the following.

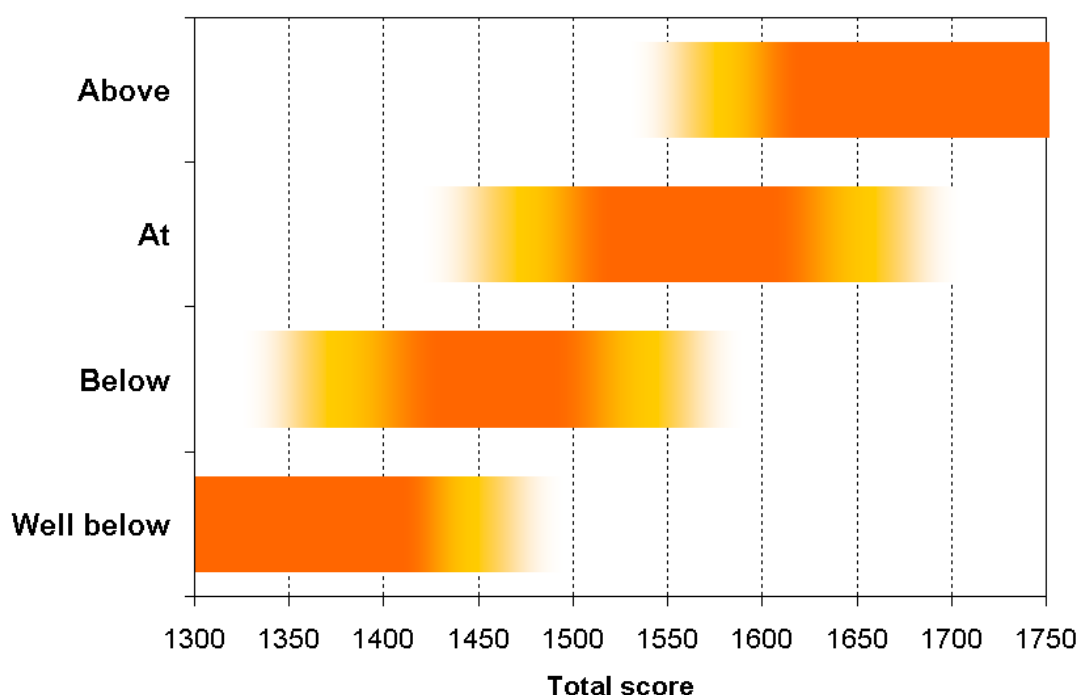
- Only a fairly small part of that line is in the blue area, reflecting that a script with a total score of 1550 had relatively little likelihood of being judged as *above* the 'by the end of year 8' national standard in reading.
- A large part of that line is in the green area, reflecting that a script with a total score of 1550 had a large likelihood of being judged as *at* the 'by the end of year 8' national standard in reading.

- A reasonably large part of that line is in the orange area, reflecting that a script with a total score of 1550 had a reasonably large likelihood of being judged as **below** the ‘by the end of year 8’ national standard in reading.
- A small part of that line is in the red area, reflecting that a script with a total score of 1550 had a small likelihood of being judged as **well below** the ‘by the end of year 8’ national standard in reading.

A teacher whose student had an e-asTTle total score of 1550 could use the above information, together with their knowledge of other aspects of that student’s reading performance, to make an overall teacher judgement against the ‘by the end of year 8’ national standard in reading.

The graph below shows the same information, but now as a set of ‘fuzzy bars’ illustrating the overlap between the four reporting bands when considered against the total score. In it, darker shades of orange correspond to higher likelihoods of a judgement against the standard falling into that reporting band.

Overlaps between Reporting Bands Depending on Total Score



The following points summarise the main features of the above graphs.

- ‘Well below’ judgements were somewhat likely⁴ up to a total score of about 1450. They were clearly the most likely⁵ judgements up to a total score of about 1410.
- ‘Below’ judgements were somewhat likely for total scores ranging from about 1370 to 1550. They were clearly the most likely judgements for total scores ranging from about 1430 to 1490.

⁴ i.e. with a probability of about 25% or more.

⁵ i.e. with a probability greater than 50%.

- ‘At’ judgements were somewhat likely for total scores ranging from about 1470 to 1660. They were clearly the most likely judgements for total scores ranging from about 1520 to 1610.
- ‘Above’ judgements were somewhat likely for total scores of about 1570 and higher. They were clearly the most likely judgements for total scores of about 1620 and higher.

Conclusion

This work will help teachers to use e-asTTle as part of the evidence informing the overall teacher judgement of student performance against the ‘by the end of year 8’ national standard in reading. It is modest in scale and is yet to be replicated in a larger setting.

For each point on the e-asTTle total score scale, this work provides the likelihood of a student with that score being judged as ‘well below’, ‘below’, ‘at’ and ‘above’ the ‘by the end of year 8’ national standard in reading. The fact that results are presented as likelihoods or probabilities reflects the fact that no one assessment tool will be enough to make a definite judgement against the reading standard. It highlights the need to use multiple pieces of evidence in informing overall teacher judgement of student performance against the national standards. All those using the material in this paper should pay attention to the important notes in the box below.

Important Notes

When using an e-asTTle reading test as one piece of evidence to inform a teacher judgement, it is worth considering more than just the total scale score (as presented here). Information from the Individual Learning Pathways report, as well as actual performance on different items, may help to refine a teacher’s judgement. Where customised tests are used, the difficulty levels for the items selected will need careful attention in order to make valid and reliable judgements in relation to the standards.

One thing you may notice when you study these results is that often there is not a consistent relationship between a test’s norms and the most likely national standard reporting category. This is not a cause for concern. Test norms are based on what the average student of a given age *can* do; the standard relates to what all students *should be able* to do, if they are on track for a successful educational outcome. In some areas of learning, the two coincide – the average student is at the required level. In other areas, there may be a general shortfall – only high-performing students are likely to reach the standard, with others needing to improve their achievement in order to do so. This is an important feature of national standards, and one of the ways in which they are intended to drive improved learning for all students.

Acknowledgements

We would like to thank all the people who have helped with and contributed to this work, especially staff at Auckland University who provided the scripts, the teachers

and other literacy experts who took part in the script scrutiny event, and those who commented on initial drafts of this material to improve it.

Key points to remember

Some key points to remember about this information:

- It is provisional, based on early analysis of existing data
- It is designed to help teachers make judgements against the standard on the basis of student performance on particular assessment tools
- It tries to capture the variability around the judgements made using any single tool, and shows the importance of pulling together different kinds of evidence to make an overall teacher judgement
- Results for more assessment tools are being processed and will be published as soon as possible
- Next year we will be publishing results based on more extensive data collection and analysis for a range of assessment tools
- Schools can use these results directly – they do not need to carry out their own ‘script scrutiny’ exercises
- When making judgements based on e-asTTle, teachers should use more than just the total score, and should make use of data presented on the Individual Learning Pathways reports
- Teacher judgements should be based on a wide range of assessment information, not just the outcomes from a single tool like e-asTTle
- The relationship between national norms on a test like e-asTTle and the national standards is not predetermined, and it may be that ‘average’ performance is not good enough to meet the standard.