PAT: Reading Test 1 Comprehension and Vocabulary and the 'by the end of year 4' 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 4' national standard in reading[1]. In particular it will help teachers who use the PAT: Reading (Comprehension and Vocabulary) Test 1 assessment tool as part of the evidence informing their overall teacher judgement.

The work

A team of experienced teachers, literacy professional development facilitators and an assessment expert from NZCER, worked through sets of anonymised PAT scripts which record student performance during year 4.

Initially, the 'by the end of year 4' standard was introduced. This was followed by a description of how data collected using the PAT: Reading Test 1 assessment tool could reveal features of the standard such as the ability to read, respond to, and think critically about texts. Then, after a discussion of the definitions of the standard's four reporting bands ('well below', 'below', 'at' and 'above'[2] –) 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. It means that for any one piece of evidence describing student performance, (such as the PAT: Reading Test 1 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.

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 for the STAR assessment and the Observation Survey. It intends to follow this with similar work for other assessment tools.

Results

Comprehension

Student achievement in PAT: Reading Test 1 is measured by comprehension and vocabulary scale scores. The current results describe how each of these scales align

---

[1] See: http://nzcurriculum.tki.org.nz/National-Standards/Reading-and-writing-standards/The-standards/End-of-year-4

[2] See http://assessment.tki.org.nz/Effective-use-of-evidence/Overall-teacher-judgement-OTJ/A-student-s-achievement

with the 'by the end of year 4' national standard in reading. The 'area graph' below shows the percentage of scripts in each of the four reporting bands against the comprehension scale score. It was produced using a statistical modelling technique applied to the collected data.

**Percentages at Different Reporting Bands Depending on Comprehension Scale Score (Area Graph)**
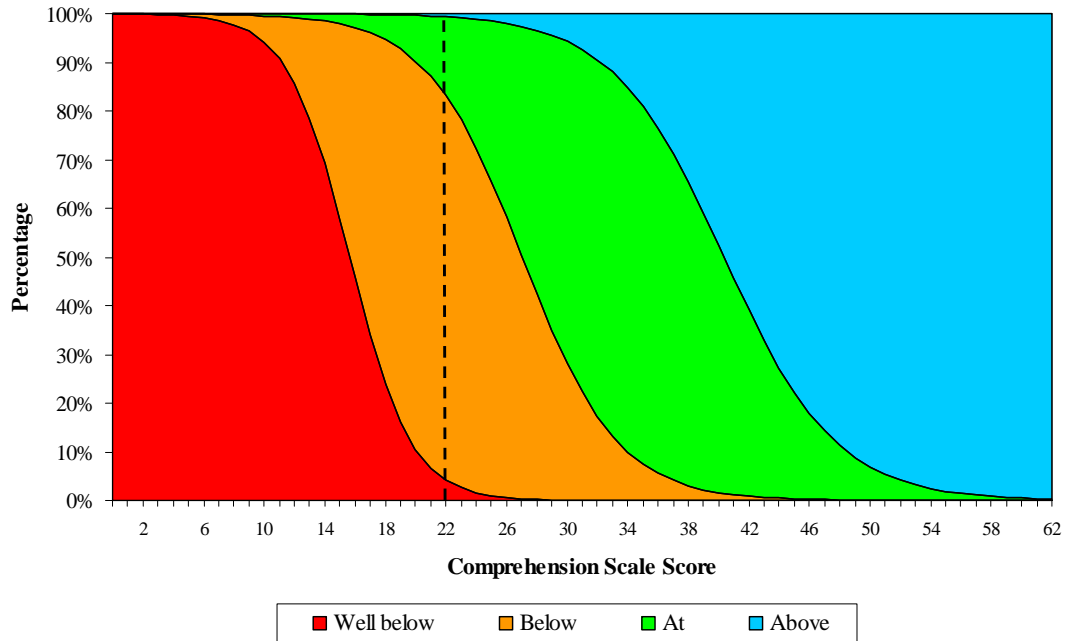
To read the graph, consider a vertical line above a comprehension scale score of, for example, 22 (illustrated). Looking at the areas with which the line intersects, we can conclude the following.

- Only a tiny part of that line is in the blue area, reflecting that a script with a comprehension scale score of 22 had almost no likelihood of being judged as *above* the 'by the end of year 4' national standard in reading.
- A small, but somewhat larger, part of that line is in the green area, reflecting that a script with a comprehension scale score of 22 had a small likelihood of being judged as *at* the 'by the end of year 4' national standard in reading.
- A large part of that line is in the orange area, reflecting that a script with a comprehension scale score of 22 had a large likelihood of being judged as *below* the 'by the end of year 4' national standard in reading.
- A very small part of that line is in the red area, reflecting that a script with a comprehension scale score of 22 had a very small likelihood of being judged as *well below* the 'by the end of year 4' national standard in reading.

A teacher whose student had a PAT: Reading Test 1 comprehension scale score of 22 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 4' 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 scale 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 Comprehension Scale Score
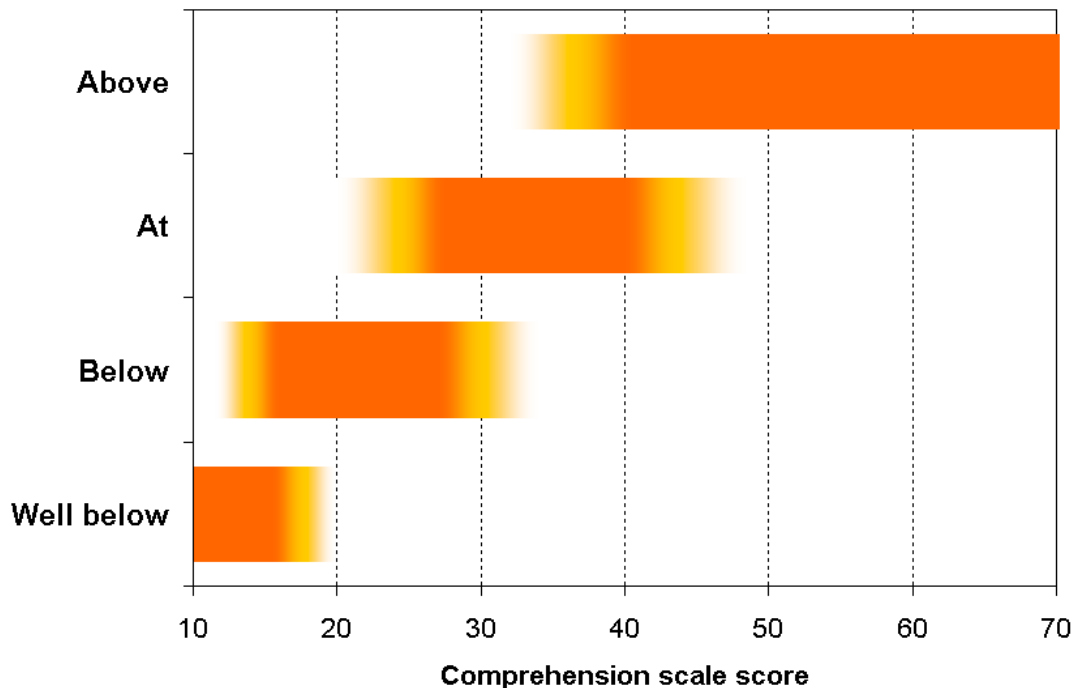
The following points summarise the main features of the above pair of graphs.

- 'Well below' judgements were somewhat likely[3] up to a comprehension scale score of about 19. They were clearly the most likely[4] judgements up to a comprehension scale score of about 16.
- 'Below' judgements were somewhat likely for comprehension scale scores ranging from about 13 to 32. They were clearly the most likely judgements for comprehension scale scores ranging from about 16 to 28.
- 'At' judgements were somewhat likely for comprehension scale scores ranging from about 24 to 47. They were clearly the most likely judgements for comprehension scale scores ranging from about 28 to 42.
- 'Above' judgements were somewhat likely for comprehension scale scores of about 36 and higher. They were clearly the most likely judgements for comprehension scale scores of about 42 and higher.

---

[3] i.e. with a probability of about 25% or more.
[4] i.e. with a probability greater than 50%.

Vocabulary

The area graph below shows the percentage of scripts in each of the four reporting bands against the vocabulary scale score. It was produced using a statistical modelling technique applied to the collected data.

Percentages at Different Reporting Bands Depending on Vocabulary Scale Score (Area Graph)
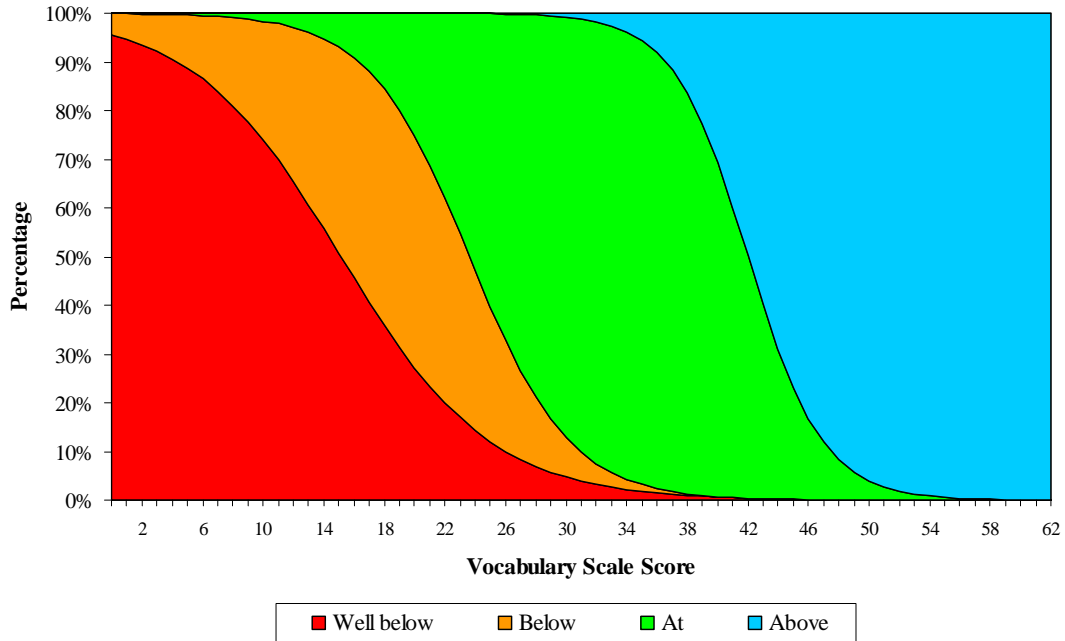
This graph may be read similarly to the comprehension area graph above. The graph below shows the same information, but now as a set of 'fuzzy bars' illustrating the overlap between the four reporting bands when considered against the scale score.

Overlaps between Reporting Bands Depending on Vocabulary Scale Score
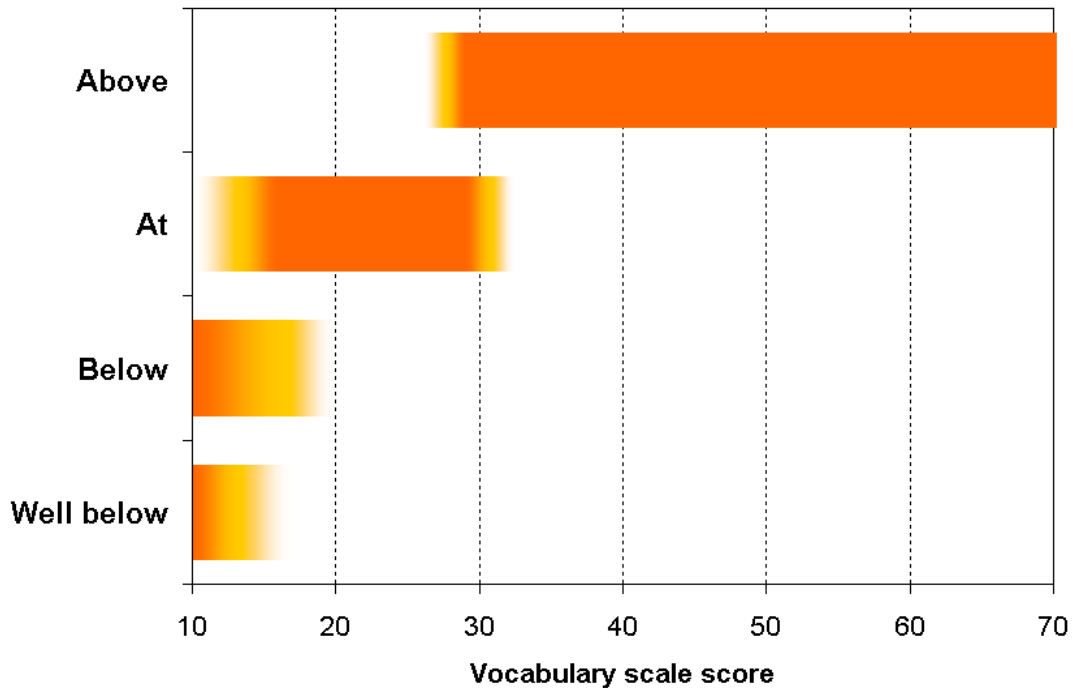
The following points summarise the main features of these two graphs for vocabulary.

- 'Well below' judgements were somewhat likely up to a vocabulary scale score of about 22. They were clearly the most likely judgements up to a vocabulary scale score of about 16.
- 'Below' judgements were somewhat likely for vocabulary scale scores ranging from about 9 to 27. They were clearly the most likely judgements for vocabulary scale scores ranging from about 16 to 23.
- 'At' judgements were somewhat likely for vocabulary scale scores ranging from about 19 to 46. They were clearly the most likely judgements for vocabulary scale scores ranging from about 23 to 42.
- 'Above' judgements were somewhat likely for vocabulary scale scores of about 39 and higher. They were clearly the most likely judgements for vocabulary scale scores of about 42 and higher.

Conclusion

This work will help teachers to use PAT: Reading Test 1 as part of the evidence informing the overall teacher judgement of student performance against the 'by the end of year 4' national standard in reading. It is modest in scale and is yet to be replicated in a larger setting.

For each point on the PAT: Reading Test 1 scales, this work provides the likelihood of a student with that PAT: Reading Test 1 scale score being judged as 'well below', 'below', 'at' and 'above' the 'by the end of year 4' 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.

One thing you may notice when you study these results is that often there is not a consistent relationship between a test's norms (e.g. stanines, average scale scores for a year level etc.) 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 (i.e. stanine 5) is at the required level. In other areas, there may be a general shortfall – only high-performing students (e.g. stanine 7+) 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 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, including members of the New Zealand Assessment Academy and AtoL directors.

Important things 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.