

# Technical Information

## New Group Reading Test<sup>®</sup> (NGRT) Digital Edition

### Trialling and standardisation

---

The sentence completion, reading comprehension and phonics questions for *New Group Reading Test (NGRT)* were developed by English subject experts.

These questions were initially developed for the paper version of *NGRT*. They were trialled in 2009 using 2,481 students in 73 schools. After the trials seven paper tests were developed for standardisation for students ages 5 to 16 years. The paper standardisation was conducted in the United Kingdom and took place in 2010 using a total of 11,716 students from 476 schools. A further four paper tests were developed and trialled in 2013 using 1,635 students in 34 schools. The standardisation for these new paper tests took place in 2014 using a total of 4,960 students from 187 schools. The reliability of every paper test is above 0.9 which is high. More details of the trial and standardisation of the paper tests are given in the *NGRT* paper Teacher Guides.

The digital version of *NGRT* uses the same questions as those in the paper version. Three parallel versions of the standardised tests (Forms A, B and C) were created, with questions of similar levels of difficulties in each form. The individual question statistics derived from the paper trials and standardisation were used to map all the reading questions across all the booklets onto a common difficulty scale, and estimates of an overall reading ability were derived for the paper versions of the tests. The standardised scores for the digital version are based on the reading ability scores. Item Response Theory was used to derive the common difficulty scale and model the adaptive method used to select questions in the digital version of the test.

### Paper and Digital Equating study

---

To verify the digital version, an equating study with students doing both the paper and the digital versions of the tests was carried out in 2012 to see how the scores compared. The results showed that the digital test scores for 6 year old students (Year 1 in England and Wales) were much lower than for the paper version. It may be that the students at this age were having difficulty with using the digital version of the test; for this reason we recommend that the digital version of *NGRT* is not used by Year 1 students; the youngest recommended year group is Year 2.

There were 1721 students in Years 2 to 10 who did both the paper and digital versions of the test. The adjusted correlation between paper and digital versions after correcting for measurement error was 0.91. Overall digital scores were slightly lower than paper versions; adjustments were made to the scores of the digital version of the test to account for this difference.

### Gender differences

---

Overall females perform better than males by an average of 3.1 SAS points. This gender difference is significant and females in general, on average, tend to perform better than males at all age groups in primary and secondary schools.

## EAL differences

---

Overall non-EAL students perform better than EAL students by an average of 3.6 SAS points. This difference is significant and the differences exists at all age groups. EAL on its own does not give an indication of a student's proficiency in the English language as other contextual factors such how long the student has been in the country or age of student at arrival in the country etc. need to be considered.

## Test reliability

---

The reliability of a test is a measure of the consistency of a student's test scores over repeated testing, assuming conditions remain the same – that is, there was no fatigue, learning effect or lack of motivation. Tests with poor reliability might result in very different scores for a student across two test administrations. The reliability of the test was estimated using the Cronbach's Alpha formula which produces values ranging from 0 to 1. Values above 0.80 are very good. The reliability of *NGRT* was very high at 0.90.

For interpreting the score of an individual student, the Standard Error of Measurement (SEM) is a more useful statistic than a reliability coefficient. It indicates how large, on average, the fluctuations in standard scores may be and indicates the 68% chance or confidence band. However, most tests show the 90% chance or confidence bands. The SEM for *NGRT* is 4.7 and for an average-performing student with a Standard Age Score (SAS) of 100, there is a 90% chance that the student's true SAS will be in the range +/- 8, i.e. from 92 to 108.

The three parallel versions of *NGRT* are used to monitor student reading ability over time. A study of around 59,000 students who took different versions of the *NGRT* tests on average 6 months apart found the correlation was 0.83. The correlation for students who took the tests on average one year apart was 0.82 based on around 44,000 students.

## Test validity

---

The good level of reliability for *NGRT* indicates that the *NGRT* questions are testing the same construct, namely reading. The types of questions used in *NGRT* are used in other tests of reading.

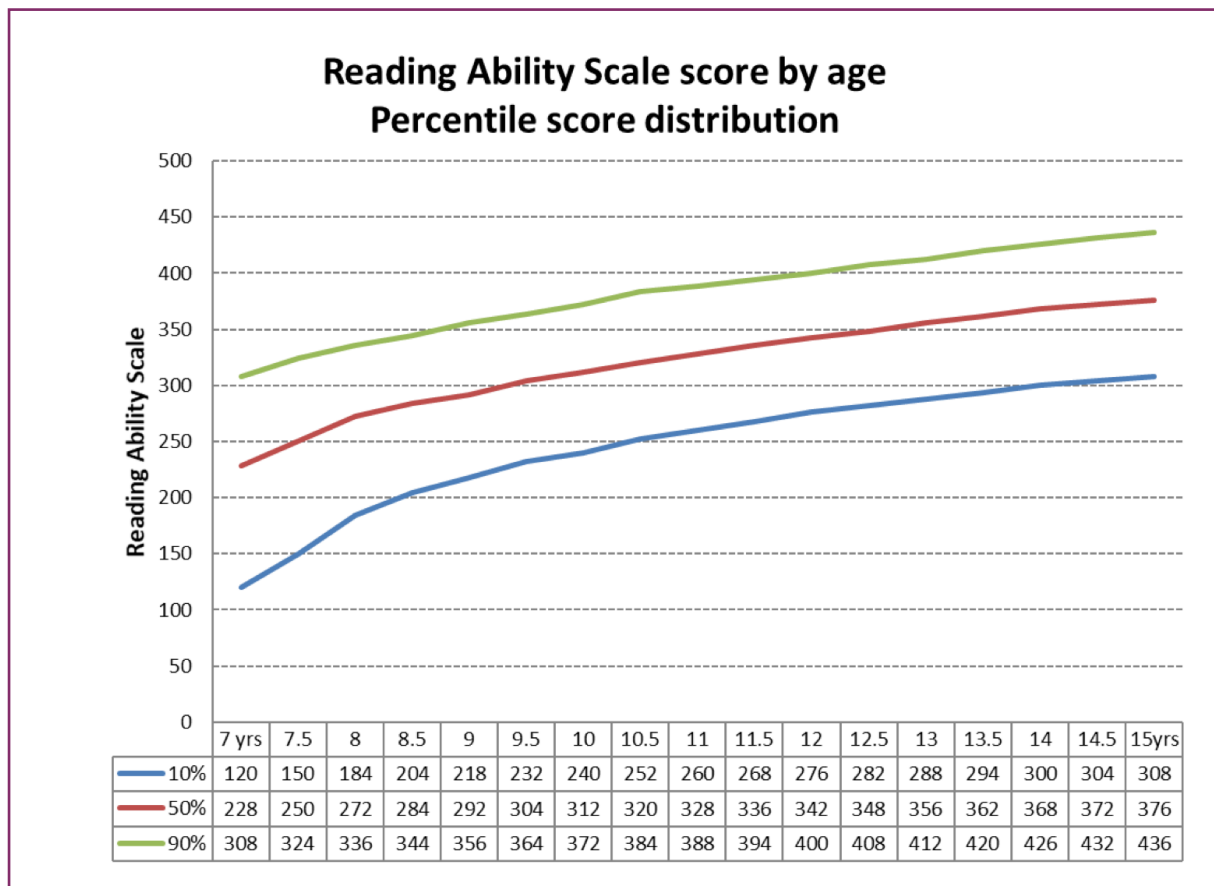
Information on Teacher Assessment (TA) levels in Reading for England was collected when the Equating study was conducted in 2012; the correlation between TA levels and the reading ability score was 0.81.

Results were collected from 7,275 Year 6 students who took *NGRT* and the Key Stage 2 SATS in England in 2016/17. The correlation between *NGRT* SAS scores and the KS2 Scaled scores in Reading was 0.75 and for the Grammar, Punctuation and Spelling it was 0.72.

## Reading Ability Scale

With an adaptive test, students within a class answer different questions depending on their abilities. Weaker students attempt easier questions whereas more able students attempt harder questions. Therefore a weaker or a more able student may get the same number of questions correct but they will have different abilities. The ability scale score takes into account the difficulty of the questions a student has attempted and is the adaptive equivalent to the raw score for a static test. Any use of raw scores in a fixed test applies to the ability scale. The advantage with *NGRT* scale score is you can use it to track the absolute Reading ability of an individual over time. SAS scores are adjusted for age and the average SAS score at any age group nationally is 100.

The chart and table below shows the percentile distribution of the scale score by age. It shows that the average (50th percentile) scale score for students at age 7 years is 228 and increases to 342 at age 12. The top 10% of students (i.e. 90th percentile) at age 7 have a scale score of 308 or higher and at age 12 the top 10% of students have a scale score of 400 or higher.



## Reading Ages

Reading ages are not the same as Reading attainment. Reading ages are derived from the average raw or scale scores at different age points. The red line in the chart shows the average scale scores e.g. the average scale score for 8 year old students is 272. Therefore, any student with a scale score of 272 will have a Reading age of 8 years. For example, a 12 year old student with a scale score of 276 and in the bottom 10th percentile will have an age-equivalent close to 8 years (as shown in the black lines below).

The year on year average growth in Reading ability is much higher at younger ages (e.g. average increase from 228 to 272 between ages 7 and 8 years). For older students the year on year growth is much smaller over ages 13 years (e.g. average increase from 368 to 376 between ages 14 to 15 years). Therefore, a Reading age of 17 years is unlikely to be that much different from a Reading age of 14 years. In most cases it is not sensible to relate scores for students with above average ability to Reading age as Reading ages relate to an average. In this example, the most we can say about a 12 year old student with a scale score of 400 and in the top 10th percentile is that his Reading age is higher than 15 years (refer to the orange line). Therefore, there are issues with using Reading ages especially with above average students and it is best to use standard age scores (SAS) for measuring progress or monitor trends.

