

Technical Information

New Group Spelling Test® (NGST)

Trials

The questions for *New Group Spelling Test (NGST)* were developed by English subject experts. These questions were trialed in January/February 2017. Two or three paper booklets for each age group were created. 857 questions were trialed and 472 of these were single word questions and 385 were spelling in context questions. There were common questions between booklets within the same age as well as common items between booklets across the age groups. The use of common questions between booklets enables us to calculate the relative difficulty of all questions across all the booklets. Every question was taken by a minimum of 250 students and on average each question was administered to 665 students.

The numbers of students taking part in the paper trials were as follows.

Age (years) level	Number of students
7	800
8	896
9	834
10	887
11	971
12	1,702
13	1,283
14+	1,233
Total	8,606

The data from the paper trials were analysed to provide information on the difficulty level of each question by age, its ability to discriminate between high and low scorers, and the extent to which it proved equally difficult for both genders, once each gender's general level of performance was taken into account. This information was then used to select questions for the final standardisation digital versions.

Standardisation

The individual question statistics derived from the paper trials were used to map all the spelling questions across all the booklets onto a common difficulty scale. Item Response Theory was used to derive the common difficulty scale and also model the adaptive method used to select questions in the digital versions of the tests.

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 standardisations were conducted in the United Kingdom in May/June 2017. A national database of schools was created and schools were grouped

into categories by country (Wales, Scotland and Northern Ireland). In England, schools were further grouped into independent or grammar plus five categories of school intake based on overall school performance at end of primary schooling using Key Stage 2 outcomes, or for secondary schools using the GCSE outcomes.

Schools were selected by stratified random sampling procedures within these groupings. As this was a national sample, many schools taking part in the standardisation had never used English assessments from GL Assessment before.

Students were randomly allocated to take one of the three adaptive test forms and the total numbers of students taking part in the standardisation were as follows.

Age (years) level	Number of students
7	843
8	994
9	1,008
10	1,059
11	878
12	1,063
13	854
14+	744
Total	7,443

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 considered to be very good. The reliability of *NGST* was very high at 0.95.

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 *NGST* is 3.4 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 +/- 6, i.e. from 94 to 106.

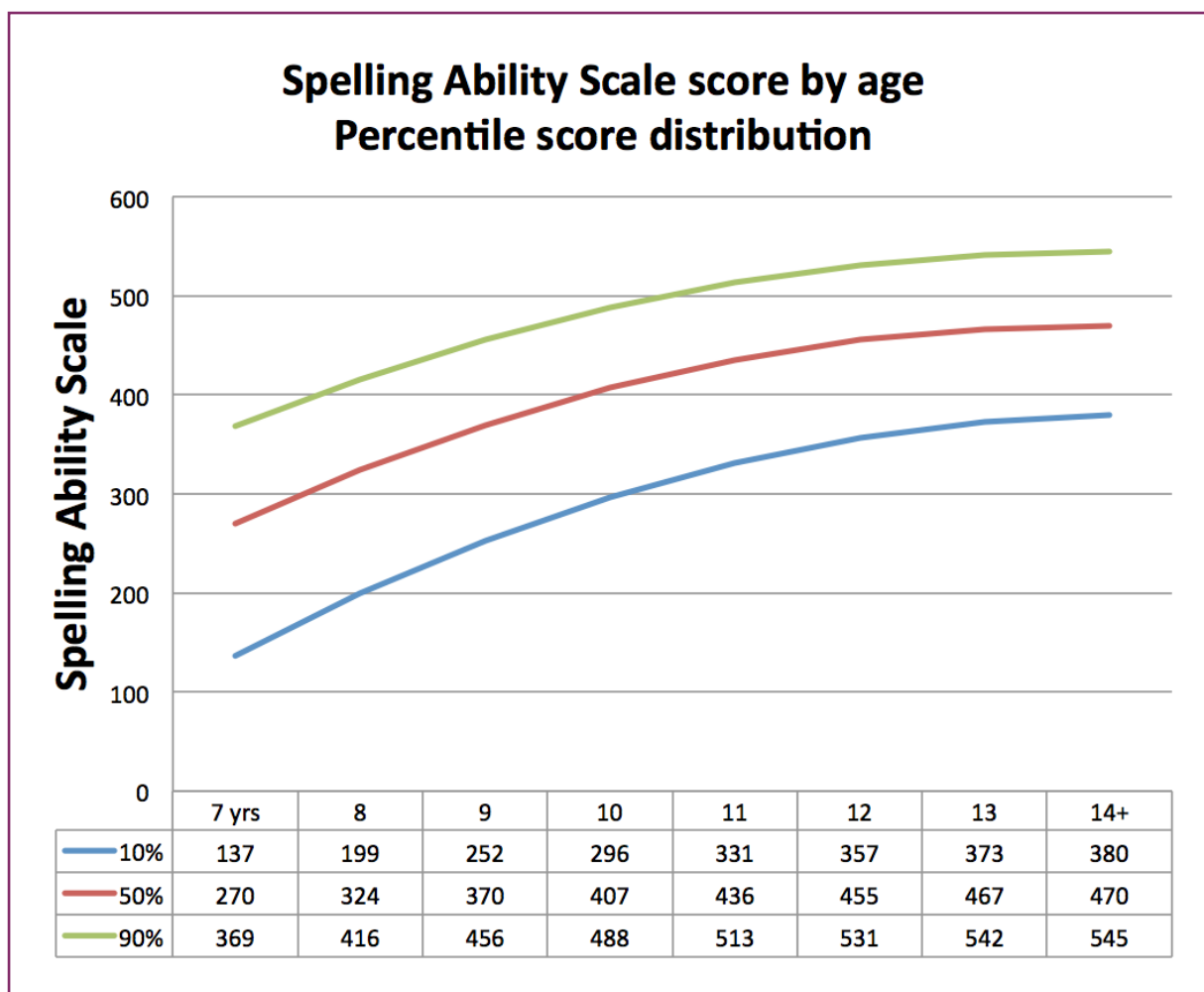
Gender differences

The tests have been age standardised to a national mean of 100 and standard deviation of 15. There were approximately similar numbers of males and females in the standardisation. Overall females perform better than males by an average of 1.8 SAS points. The difference is small but is statistically significant.

Spelling Ability Scale

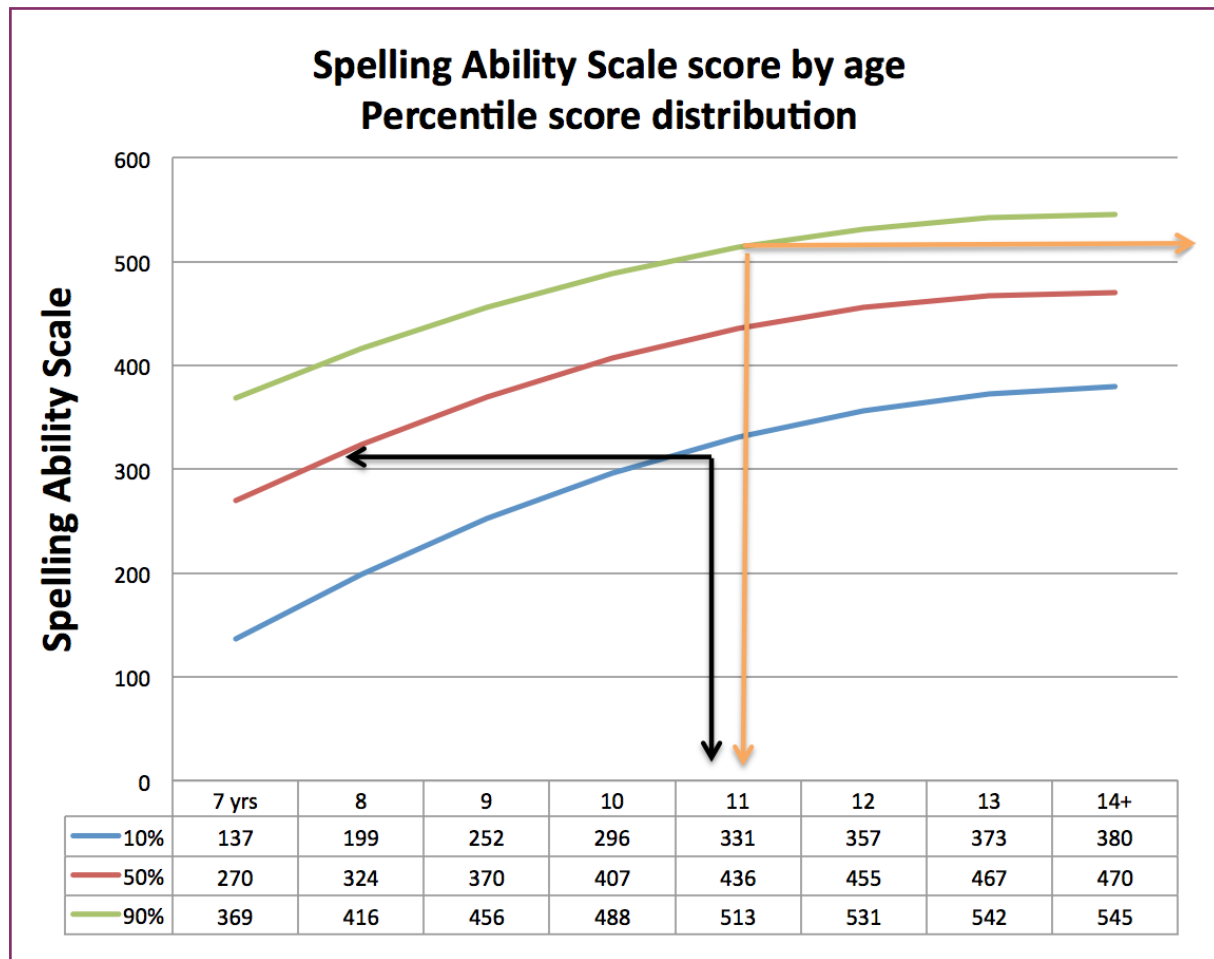
With an adaptive test, students within a class answer different questions depending on their abilities. Weaker students attempt easier questions whereas more able students attempt harder questions. Therefore a weaker or a more able student may get the same number of questions correct but they will have different abilities. The ability scale score takes into account the difficulty of the questions a student has attempted and is the equivalent to the raw score for an adaptive test. Therefore any use of raw scores in a fixed test applies to the ability scale. The advantage with *NGST* scale score is you can use it to track the absolute spelling 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 270 and increases to 436 at age 11. The top 10% of students (i.e. 90th percentile) at age 7 have a scale score of 369 or higher and at age 11 the top 10% of students have a scale score of 513 or higher.



Spelling Ages

Spelling ages are not the same as spelling attainment. Spelling 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 324. Therefore, any student with a scale score of 324 will have a spelling age of 8 years. For example, a 11 year old student with a scale score of 324 and in the bottom 10th percentile will have an age-equivalent of 8 years (as shown in the black lines below).



The year on year average growth in spelling ability is much higher at younger ages (e.g. average increase from 270 to 324 between ages 7 and 8 years). For older students the year on year growth is much smaller and nearly plateaus over ages 13 (e.g. average increase from 467 to 470). Therefore, a spelling age of 17 years is unlikely to be that much different from a spelling age of 14 years. In most cases it is not sensible to relate scores for students with above average ability to spelling age as spelling ages relate to an average. In this example, the most we can say about an 11-year-old student with a scale score of 513 and in the top 10th percentile is that their spelling age is higher than 14 years (refer to the orange line). Therefore, there are issues with using spelling ages especially with above average students and it is best to use standard age scores (SAS) for measuring progress or monitoring trends.