TO ACCESS THE TOOL AND START GETTING MORE FROM YOUR TEST DATA VISIT WWW.NFER.AC.UK/TOOL.

Interpreting test outcomes and reports generated by the NFER Tests Analysis Tool

Discover how to gain rich insights from your test data with the online tool

NFER Classroom

Interpreting test outcomes and reports generated by the NFER Tests Analysis Tool

At NFER, we believe that the real value of great assessment lies in the rich insights gained from the assessment data. To help you get the most from the time you and your staff dedicate to classroom assessment, we have put together this guide to help you easily interpret test outcomes and get the most value out of the reports generated by the NFER Tests Analysis Tool.

Teachers have the option to manually convert raw test scores from NFER Tests to more useful outputs such as standardised scores. The NFER Tests Analysis Tool will do this automatically, once pupils' scores have been entered into the platform. The tool also generates visual reports, which may make it easier for you to compare pupils' scores from one test to another and to monitor progress visually rather than viewing standardised scores as a list. Whilst this guide makes use of the reports generated by the online tool, the same principles apply to the interpretation of standardised scores in a list.

The progress information provided by the NFER Tests should inform rather than determine teachers' understanding of how much progress has been made by an individual pupil relative to other pupils in their year group. Looking at how a pupil is progressing, compared with their peers, will help teachers to identify pupils who may be in need of additional support in order to maintain an appropriate level of progress.

It should be remembered that in order to obtain reliable progress information, you should administer the test according to the guidance given in the relevant teacher guide. It is particularly important that you observe the time limits given in the test instructions, and mark questions strictly according to the mark scheme. If not, the data derived from the test scores cannot be used reliably.

How to access the NFER Tests Analysis Tool

The NFER Tests Analysis Tool is available for all purchasers of a teacher guide and can be accessed at the following website: https://hub.nfer.ac.uk. Your login details will have been provided with your first order of NFER Tests, but if you need a reminder please email products@nfer.ac.uk with your school name, LA/DfE number and an official school email address. For guidance in generating any of the reports available in the tool, please refer to the support pages, which can be accessed from the top right-hand corner of the tool, once you are logged in.

To access the NFER Tests Analysis Tool, visit: https://hub.nfer.ac.uk

Introduction to standardised scores

To begin, it is useful to know which standardised scores are indicative of low, average and high pupil performance.

A pupil who achieves a standardised score of 100 is demonstrating an average level of attainment for their age group. Approximately two-thirds of pupils will have scores between 85 and 115 and therefore pupil scores in this range can be described as 'average'. Thus, pupils who score below 85 are working below average, and those who score 115 or more are working above average.

It may be useful to further divide scores that lie between 85 and 115. Scores between 85 and 94 can be described as 'low average', and scores between 106 and 115 can be described as 'high average'. Scores from 95 to 105 remain as 'average'. This is illustrated in the table below:

Standardised score	Description of pupil attainment	
70 to 84	Below average	
85 to 94	Low average	All pupils within this range are working at an average standard
95 to 105	Average	
106 to 115	High average	
116 to 140	Above average	

Almost all pupils' scores fall within the range 70 to 140, so scores outside this range can be regarded as exceptional.

It is worth noting that the scaled score of 100 defined by the Department for Education (DfE) as the national expectation at the end of key stage 2 is not the same as, nor equivalent to, a standardised score of 100 on NFER Tests. A standardised score of 100 on NFER Tests represents the average performance, based on a normal distribution, of the sample of pupils on which the tests were standardised. At the end of key stage 2, the DfE's scaled score of 100 represents the 'expected standard' and is not the average.

Standardised scores enable comparisons to be made between the performance of a specific pupil and that of other pupils who have taken the same test. Additionally, comparing one pupil's standardised scores over time enables monitoring of that pupil's progress.

Comparing standardised scores over time

Comparing standardised scores over time gives an indication of how well a pupil is progressing relative to other pupils in the year group. The information below describes what any difference in standardised scores means.

Key points to remember:

If a pupil's standardised scores remain approximately the same, that pupil is making the same amount of progress as pupils in the large nationally representative standardisation group, i.e. they are progressing as expected.

If a pupil's standardised scores increase, that pupil is making more progress than others in the year group in the standardisation sample.

If a pupil's standardised scores decrease, that pupil is making less progress relative to others in the year group in the standardisation sample.

Confidence bands

Confidence bands are used to show the extent of the margin of error in the standardised scores. In other words, they show how accurately the test measures pupil attainment.

The margin of error is simply a statistical estimate, based on the fact that tests can only sample the particular area of learning which they assess and that therefore the score a pupil achieves may vary within a few points of their 'true score'. The '90 per cent confidence band' indicates how wide the margin of error is likely to be and that we can have 90 per cent certainty that the true score lies within the confidence band.

The confidence bands are helpful when deciding whether an increase or decrease in successive standardised scores is indicative of progress that is better or worse than expected respectively.

Key points to remember:

If the confidence bands surrounding a pupil's standardised scores on two successive tests **do overlap**, this indicates that the pupil has made progress that is typical of pupils at that level of attainment. In other words, standardised scores with overlapping confidence bands are likely to indicate expected progress.

If the confidence bands surrounding a pupil's standardised scores on two successive tests **do not overlap**, then it is likely the pupil is making greater progress than expected (standardised scores have increased) or less progress than expected (standardised scores have decreased).

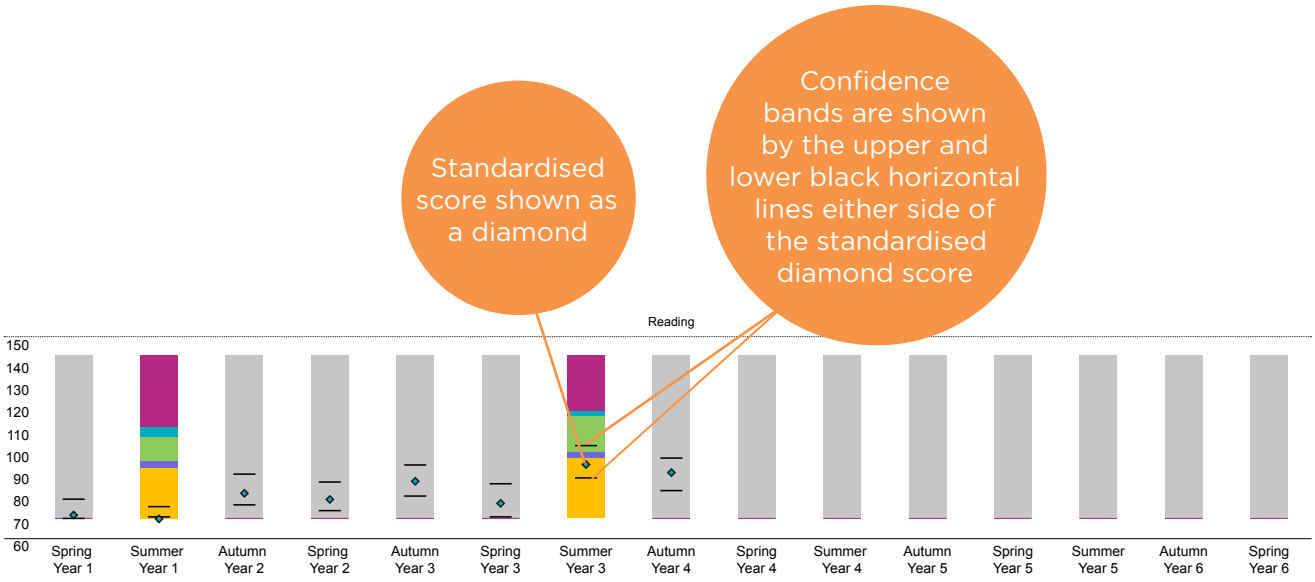
NB: Confidence bands are not the same as a statistical test of the difference between two scores. It is possible for confidence bands to overlap to some degree but for a statistical test to still reveal a significant difference between the two scores. However, they are useful in providing a visual way to interpret pupil scores.

The Individual Progress report

The Individual Progress reports in the NFER Tests Analysis Tool show all the recorded test scores for one pupil across each subject. They enable comparison of performance across NFER Tests allowing progress to be viewed over time. The standardised scores are shown on each bar of the chart as a diamond (◆).

Generally the bars are grey but where colour is used for the reading and mathematics summer tests, this also indicates whether the pupil has achieved the age-related expectations. The confidence bands are shown by the upper and lower black horizontal lines either side of the standardised score diamond.

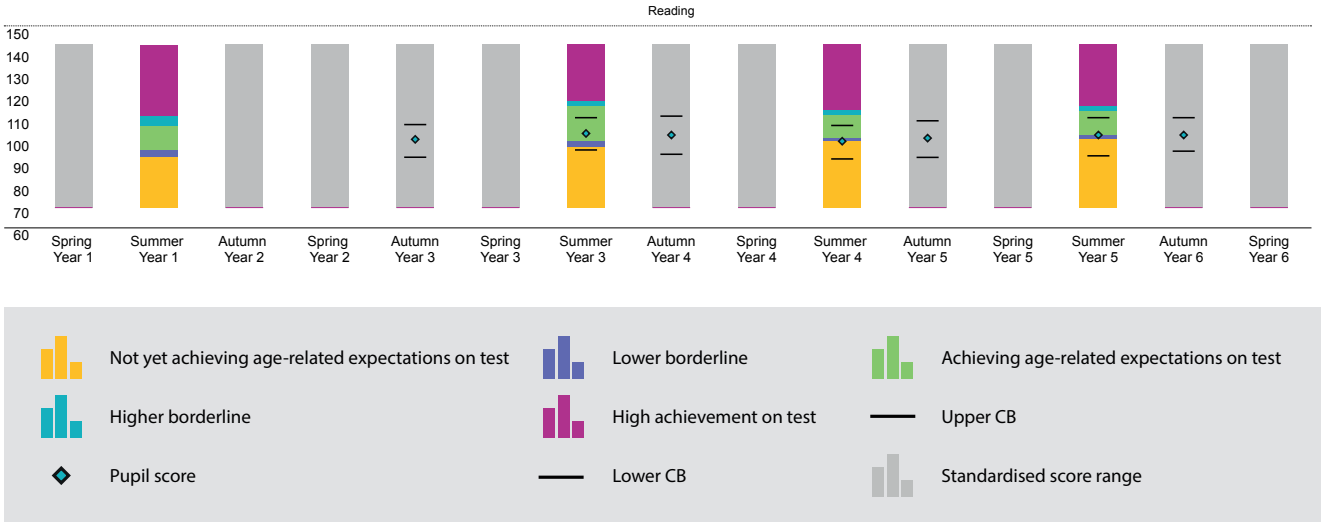
Figure 1. **An example of the Individual Progress report**



Interpreting the Individual Progress reports

To assist you in interpreting the Individual Progress reports, we have included a range of examples below. They show some typical trends and discuss the conclusions that can be drawn from studying how high or low the standardised scores are and whether or not they change over time.

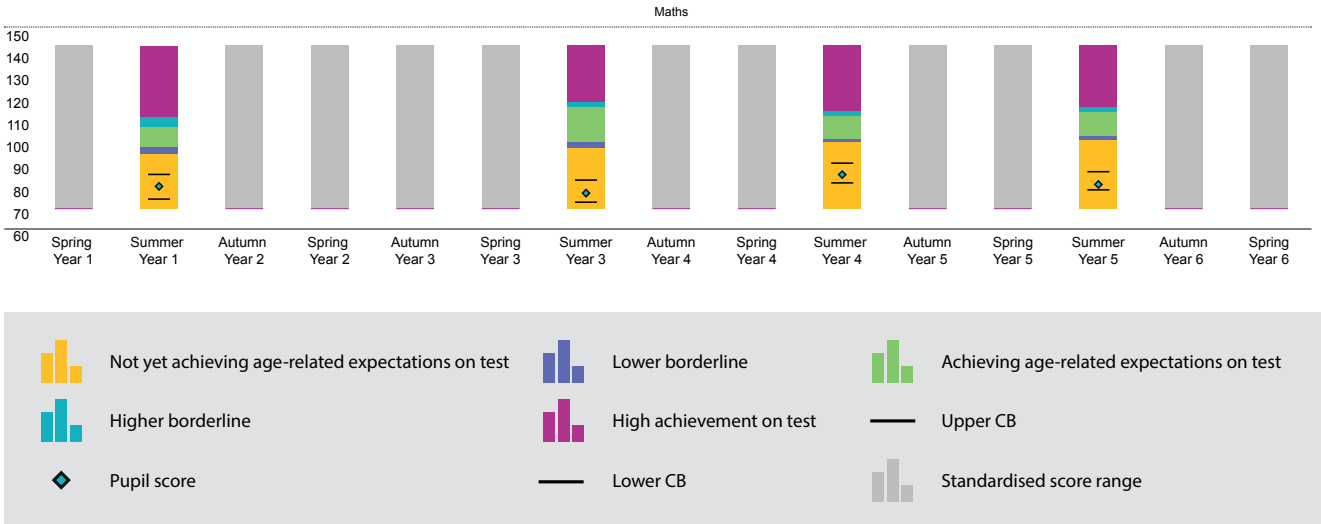
Figure 2. An average scoring pupil making average progress



The pupil in figure 2 has completed the summer and autumn tests throughout key stage 2 and is consistently achieving a standardised score of around 100. This falls into the average ability range between 85 and 115 inclusive.

Over time the standardised scores have neither increased nor decreased to a large extent so the pupil has maintained their position compared to their peers, including in the national sample. Therefore the pupil has made expected progress.

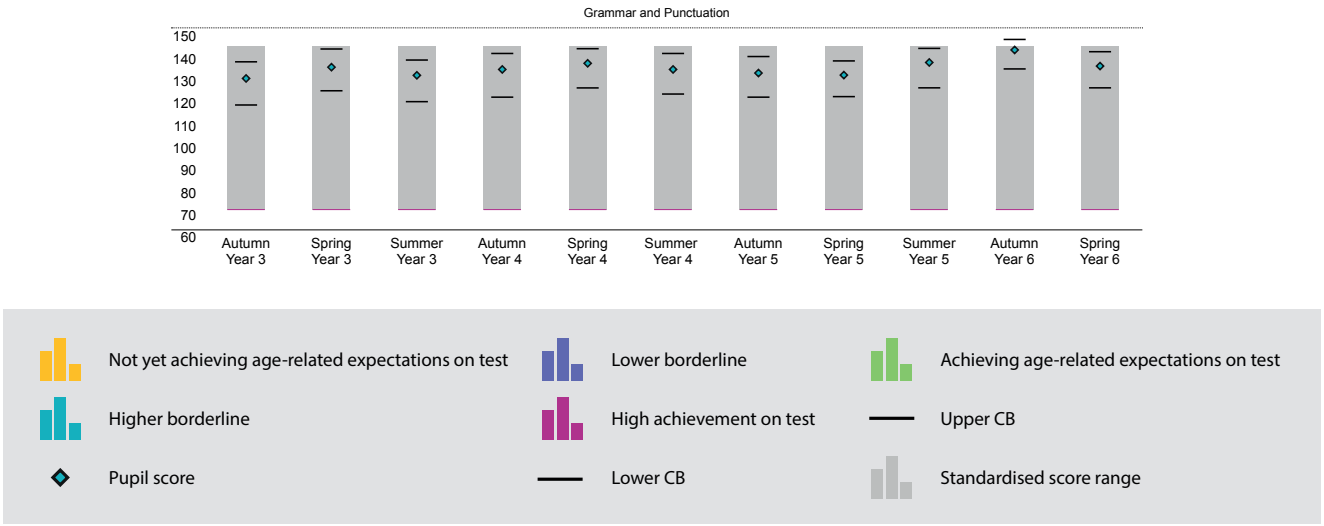
Figure 3. A lower achieving pupil making average progress



This pupil's school only complete the summer tests. The pupil is performing below average as their standardised scores are generally below 85.

Although there is some fluctuation in the standardised scores, they are not consistently increasing or decreasing but hovering around an average of 81. The pupil has therefore maintained their position compared to other pupils of a similar ability and has therefore made expected progress.

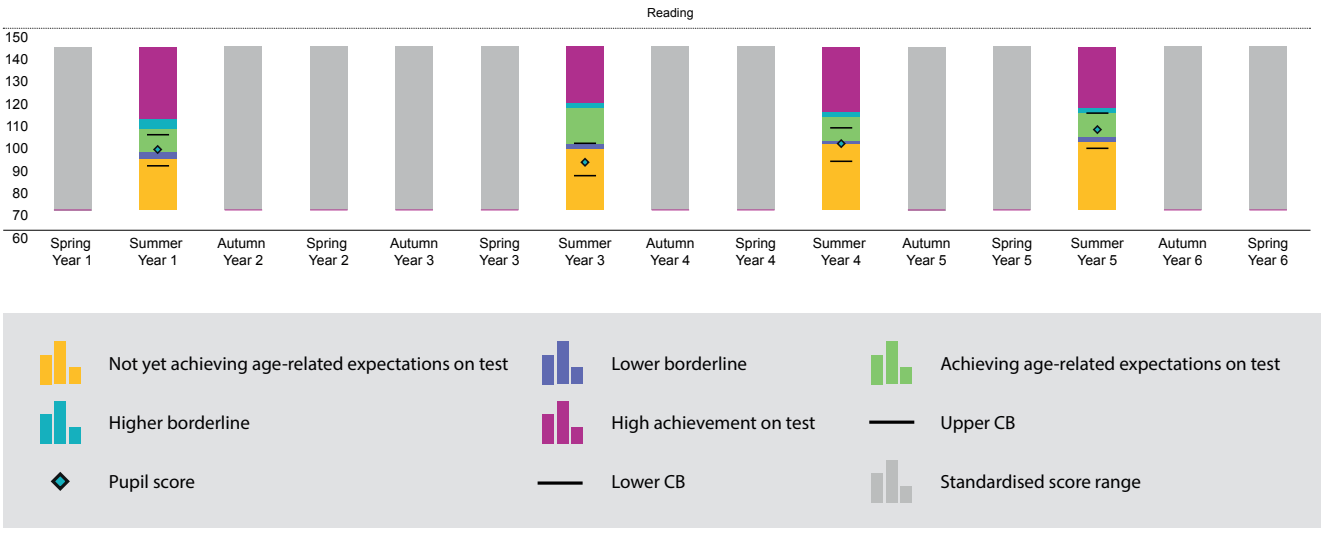
Figure 4. **A higher achieving pupil making average progress**



Grammar and punctuation papers in the NFER Tests range are available from year 3 onwards, therefore only data from year 3 onwards is shown. The pupil is performing above average as their standardised scores are higher than 115.

Although there is some fluctuation in the standardised scores, they are not consistently increasing or decreasing but hovering around an average of 132. The pupil has maintained their position compared to other high achieving pupils and has therefore made expected progress.

Figure 5. **An average scoring pupil making average progress**

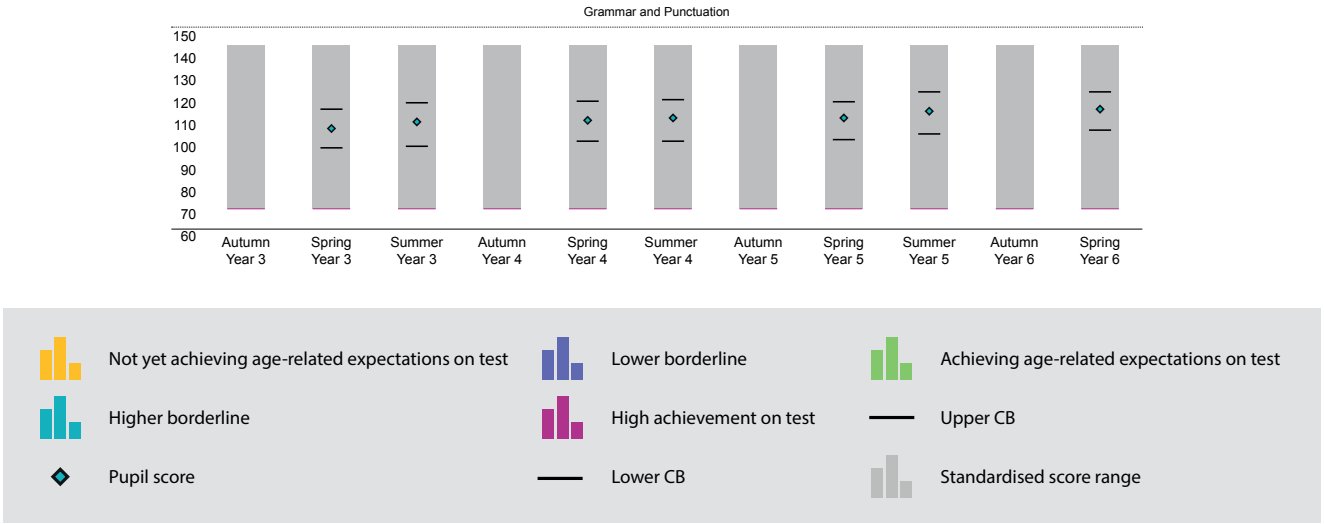


This pupil's school only completes the summer tests. The pupil is of average ability as the standardised scores are in the range between 85 and 115 inclusive.

The standardised score range is quite wide (between 91 and 104 inclusive) indicating that their performance on the tests is variable. The scores do not show a steady increase or decrease, so it appears that the pupil has made expected progress. However, we can check this by looking at the confidence bands for each score.

This pupil's true score on the summer year 3 test is likely to lie in the confidence band between 85 and 99 inclusive. On the summer year 4 test the pupil's true score is likely to lie in the confidence band between 91 and 106 inclusive. These score ranges overlap across the values between 91 and 99 inclusive meaning that the pupil's true score could actually be the same on each of the tests. Therefore we cannot say, for certain, that the pupil has improved their performance in the year 4 test, even though the standardised score has increased. Even if we take the lowest standardised score (year 3: true score between 85 and 99 inclusive) and the highest standardised score (year 5: true score between 97 and 112 inclusive), the confidence bands still have an overlap, albeit small, between 97 and 99 inclusive. This pupil is has therefore likely made expected progress.

Figure 6. **An average scoring pupil making average progress**

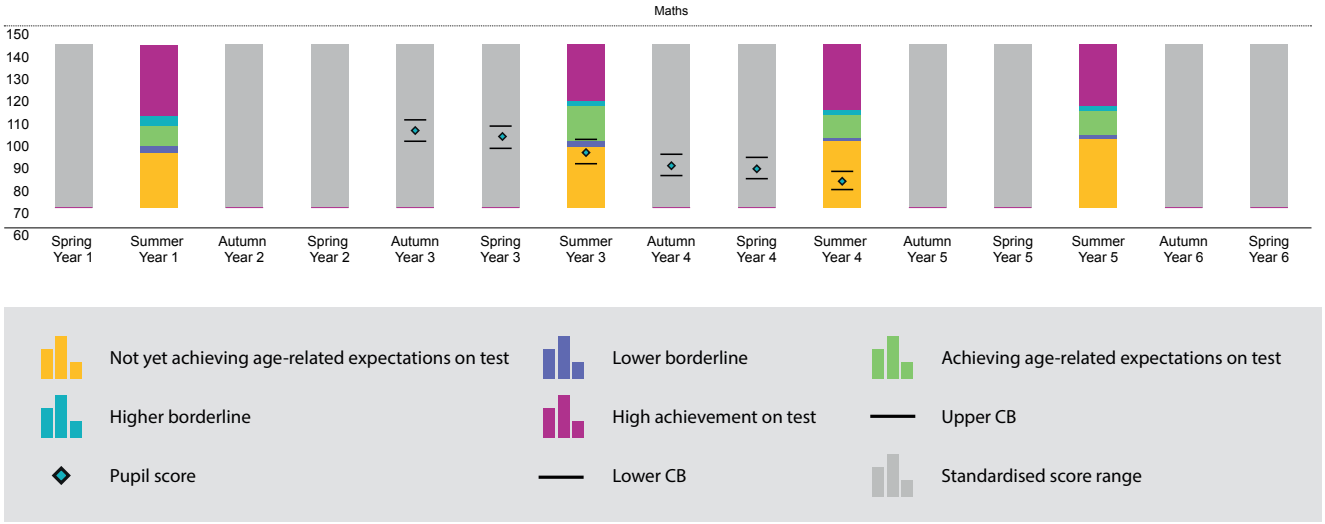


This pupil has completed spring and summer tests in key stage 2 only. With standardised scores from 106 to 114, the pupil is performing at the higher end of the average ability band.

The standardised scores are gradually increasing over time and so could potentially indicate the pupil has made more progress than expected because they appear to be improving their performance in comparison to their peers. However, to be certain, the confidence bands can be checked to see if the true score has likely increased.

For spring year 3, the true score is likely to lie in the confidence band range between 97 and 114 inclusive and for spring year 6 the true score is likely to lie between 105 and 121 inclusive. The ranges overlap between 105 and 114 inclusive so although the scores are increasing we cannot be certain that pupil progress is beyond what might normally be expected.

Figure 7. **An average scoring pupil making less progress than expected**



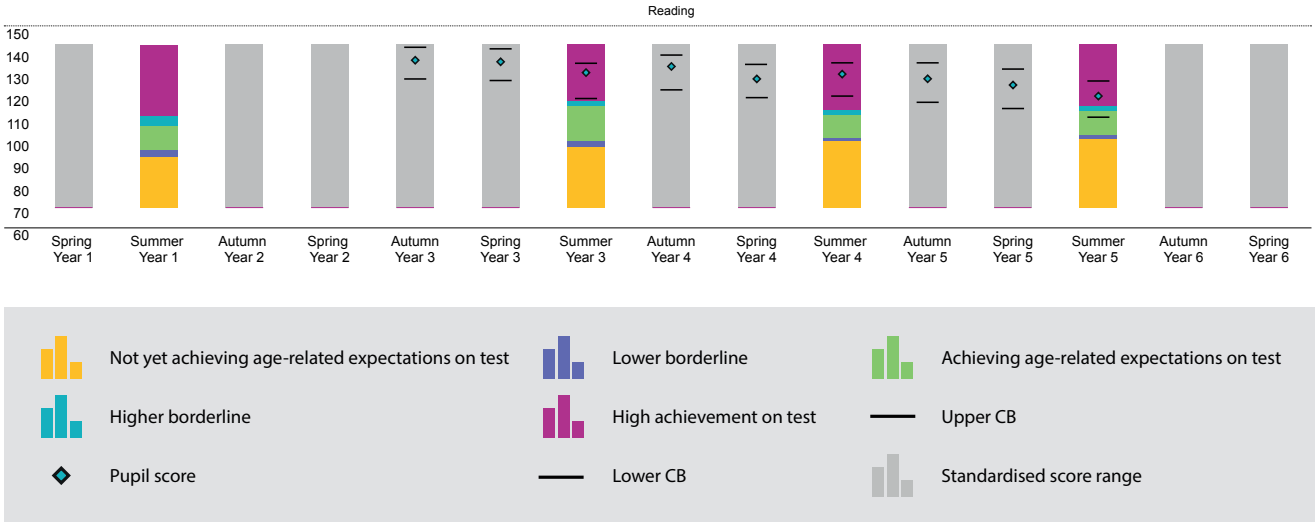
This is a pupil who is just finishing year 4 and who has completed all the tests in key stage 2 so far. The pupil is of average ability, generally having standardised scores in the range between 85 and 115 inclusive.

There is a downward trend in the standardised scores achieved at the start of key stage 2, possibly indicating that the pupil has made less progress than expected. This can be confirmed by looking at the confidence bands.

For each adjacent test there is an overlap in the confidence bands. Even between year 3 summer and year 4 autumn, where there is one of the biggest differences in standardised scores, there is an overlap. The summer year 3 confidence band (and the likely true score) is between 90 and 100 inclusive while the autumn year 4 confidence band (and the likely true score) is between 84 and 94 inclusive. This gives an overlap (between 90 and 94 inclusive) in which the true scores for both tests could lie. Looking at any two adjacent tests, the progress of the pupil looks satisfactory.

However, it may be more useful to look at the pattern overall. For autumn year 3 the true score is likely to lie between 99 and 109 inclusive and for summer year 4 it is likely to be between 78 and 87 inclusive. The two confidence band ranges do not overlap, therefore, the true score on the autumn year 3 test is likely to be higher than the true score on the summer year 4 test. This means the standardised scores have decreased sufficiently for us to say that over years 3 and 4, the pupil has probably not made the progress that might be expected. The pattern is apparent even comparing the two autumn term scores and it indicates that the pupil needs help to stop the downwards trend continuing in year 5 and to prevent performance reducing further to below average.

Figure 8. A higher achieving pupil making less than average progress



This is a pupil who is finishing year 5 and whose school completes only the key stage 2 papers. The pupil is performing at an above average standard as their standardised scores are higher than 115. There is some fluctuation in the standardised scores with some increases but mainly a downward trend.

The confidence bands show that the true score of the pupil on the autumn year 3 test is likely to lie between 126 and 139 inclusive, and on the summer year 5 test it is likely to be between 109 and 125 inclusive. There is no overlap in the ranges and therefore the true score at the end of year 5 is likely to be lower than that at the start of year 3. Therefore, it can be said that this pupil is likely to have made less progress than other higher achieving pupils.

Figure 9. A lower achieving pupil making more than average progress

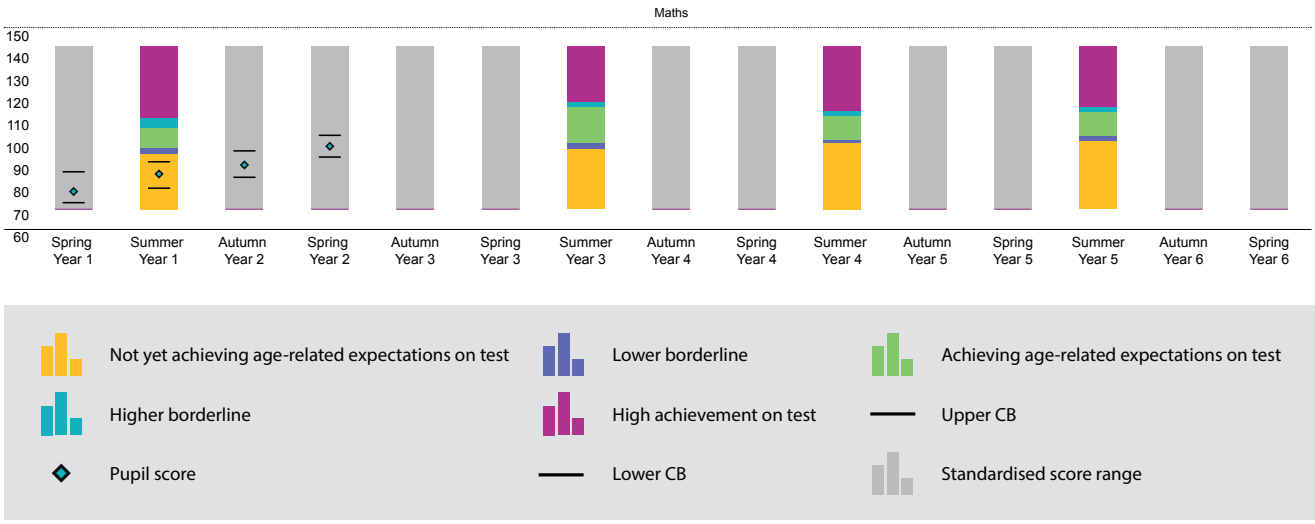
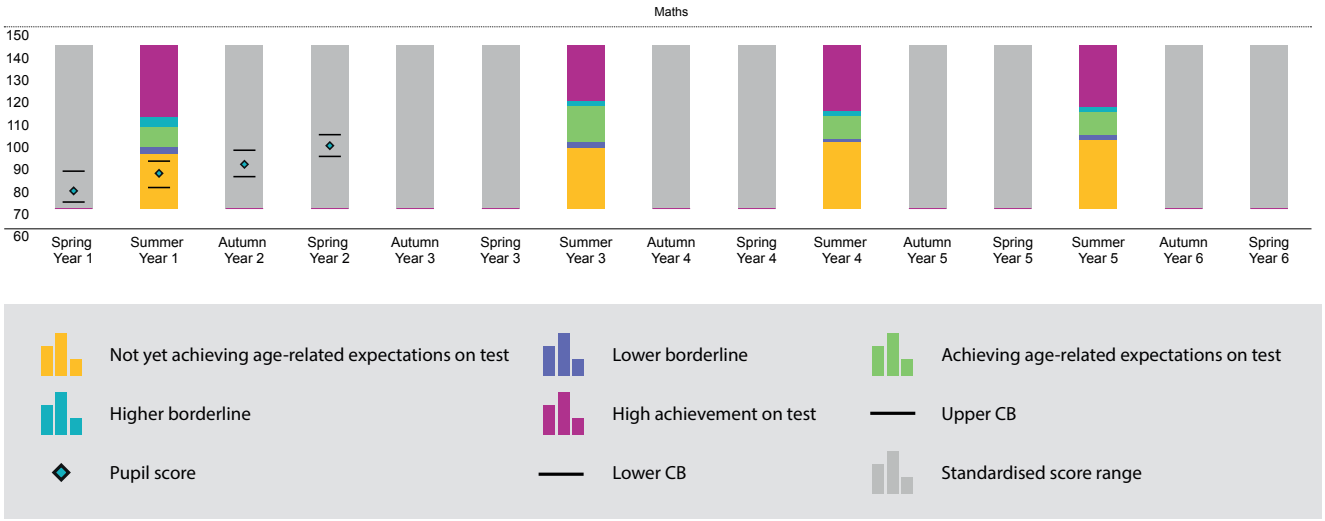


Figure 9. **A lower achieving pupil making more than average progress**



This pupil has only just completed key stage 1 and therefore there is no data for any of the key stage 2 papers. At the start of key stage 1 this pupil is performing below average as shown by a spring year 1 standardised score of 78. However, over the course of the two following years the standardised score increases to 97 in the summer of year 2. This suggests their mathematical ability has improved and can be classified as average.

The confidence bands confirm that more than average progress is likely to have been made as the true score on the spring year 1 test is likely to be between 72 and 86, while on the spring year 2 test it is likely to be between 92 and 102. The true score in year 2 is therefore likely to be above that in year 1, suggesting that the pupil has made greater than average progress.

The Attainment Comparison report

You can also use the NFER Tests Analysis Tool to make a direct comparison between pupils' scores on two different assessments. This can be done at two time points, within a year (e.g. year 5 autumn compared with year 5 summer) or between years (e.g. year 3 summer compared with year 5 summer). Such comparisons provide an overview as to whether a class or cohort is making the expected amount of progress between the two time points. This will enable you to gain an indication of pupil progress and identify pupils that are making less progress compared to their peers, highlighting a need for a more in depth examination of their individual progress. Comparisons can also be made across subjects (e.g. year 5 reading compared with year 5 mathematics). If the plot is used to compare assessments from different subjects then an indication of the subject in which pupils are attaining a higher standard is given. In the examples that follow, we focus on Attainment Comparison reports comparing the same subject at two different time points.

The graphs produced by the Attainment Comparison reports show the standardised scores for the first selected test/time point along the horizontal axis. The scores for the second selected test/time point appear on the vertical axis. Expected progress is indicated by standardised scores on or close to the diagonal line on the graph. Along this line the standardised scores at the two time points will be the same, indicating the performance on the two tests was similar. Pupils whose scores are considerably above the diagonal line may be making more than average progress as their score at the second time point is higher than at the first time point. Pupils whose scores are considerably below the line may be making less than average progress as their score at the second time point is lower than at the first time point. It is the confidence bands that help to determine whether the distance away from the diagonal is sufficient to indicate that more progress or less progress has been made.

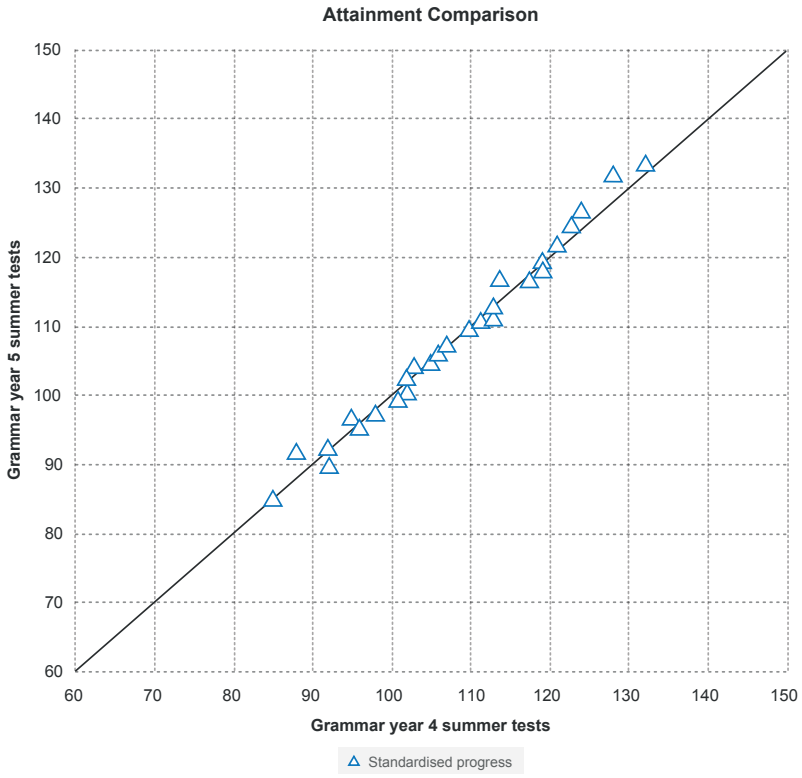
Key points to remember:

If the confidence band for a pupil's standardised score crosses over the diagonal line, then the pupil is likely to have performed similarly in the two assessments, meaning they have made an expected level of progress.

If the confidence band for a pupil's standardised score does not overlap the diagonal line, then they are likely to be making more progress than expected (score plot lies above the diagonal) or less progress than expected (score plot lies below the diagonal).

The following charts show some typical patterns and discuss the conclusions that can be drawn from them.

Figure 10. An example of the Attainment Comparison report showing a class making average progress

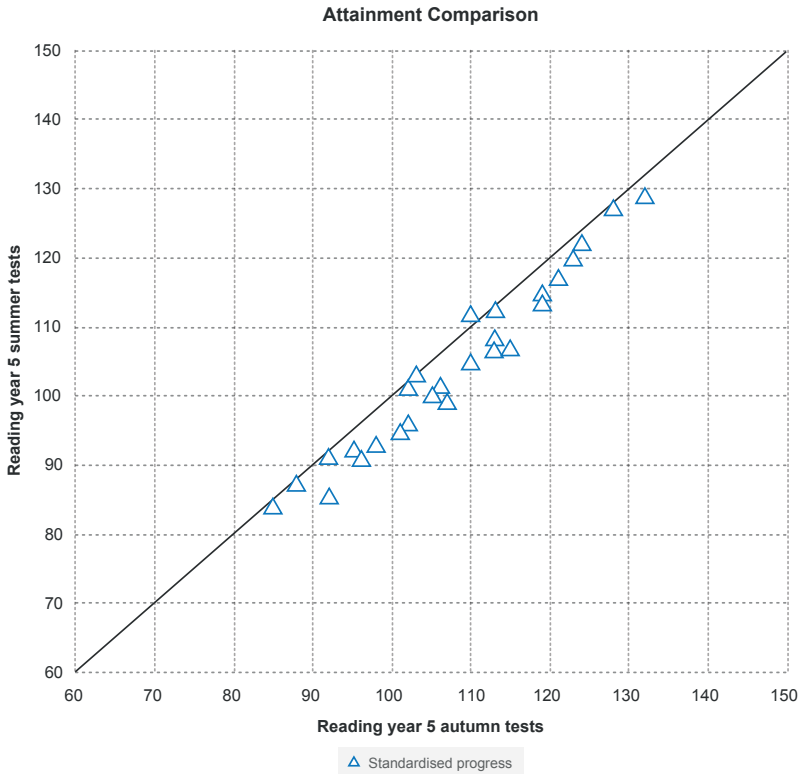


In this report, the standardised score plots are all close to the line indicating that pupils have generally scored similar standardised scores in both tests and have therefore progressed as expected.

Individual pupil plots lying close to but above or below the line are to be anticipated, and a balance either side of the line indicates that the class as a whole is progressing in line with pupils nationally.

NB: Confidence bands are not the same as a statistical test of the difference between two scores. It is possible for confidence bands to overlap to some degree but for a statistical test to still reveal a significant difference between the two scores. However, they are useful in providing a visual way to interpret pupil scores.

Figure 11. A class making less than average progress

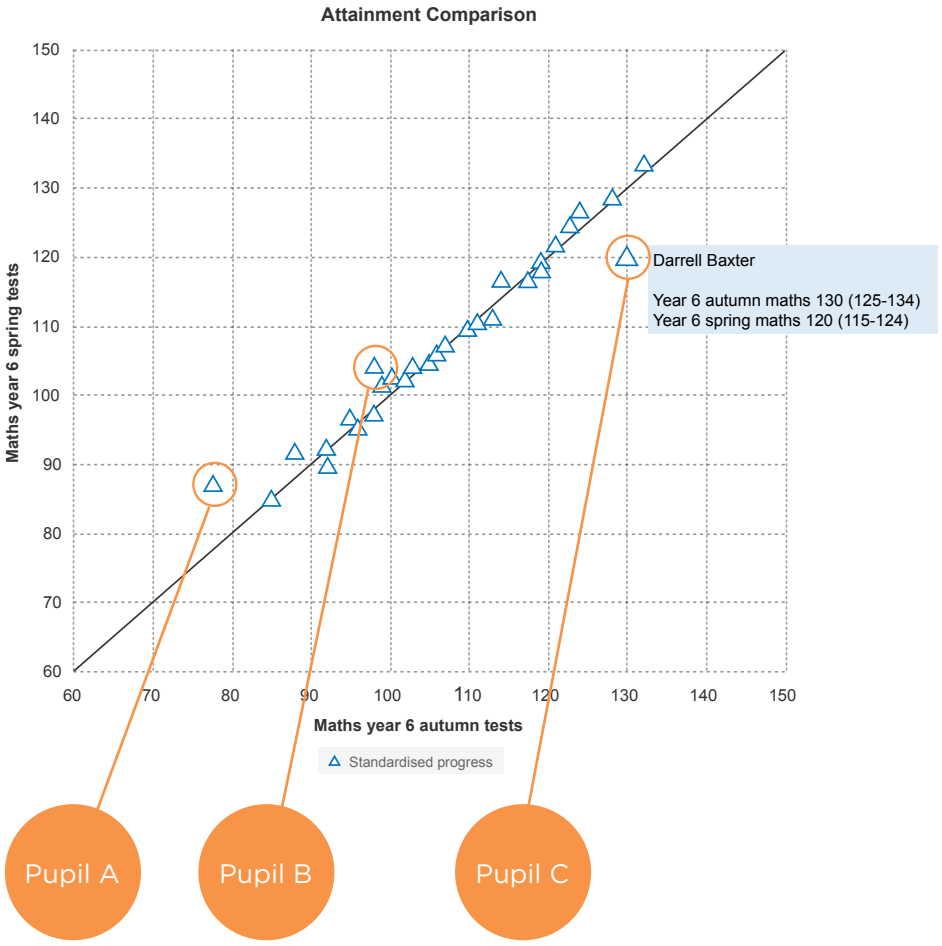


In this report, many of the standardised score plots fall below the line, showing that these pupils are scoring lower on the summer test than the autumn test. It would be sensible for the class teacher to use the confidence bands to check how substantial the difference in performance is. Plots lying close to the line are not a cause for concern but larger differences in standardised scores are more likely to indicate less progress has been made.

If many pupils' standardised score plots fall a considerable distance below the line, it may indicate the class as a whole has made less progress compared with pupils nationally. This might happen with the reading tests for example, if pupils have not had exposure to a wide range of text types.

A graph with more standardised score plots above the line may indicate that the class is making more progress than might be expected when compared with the national sample. Again, this should be confirmed by looking at the confidence bands and considering whether the plots lie a considerable distance above the line. You might expect increases in whole class performance if the school has introduced a new reading intervention. Figure 12 below gives examples of how to use the confidence bands to check how substantial the difference in performance is across two tests.

Figure 12. A class making average progress but with three pupils whose performance seems to have changed over time (Pupils A, B and C)



This report shows a class that is generally making average progress, but there are three pupils whose performance seems to have changed over time.

Pupil A scored as follows:

Test	Standardised score	Confidence band
Maths year 6 autumn	76	72-81
Maths year 6 spring	87	83-91

Pupil A has been achieving below average scores but is likely to have made progress than might be expected compared to pupils in the standardisation sample. We can see this by studying the confidence band for spring (83-91) which lies above the diagonal. There is no overlap in the ranges and therefore the pupil's spring score is higher than might be expected if they had been progressing as standard. This might reflect additional remedial help the pupil had been receiving.

Pupil B scored as follows:

Test	Standardised score	Confidence band
Maths year 6 autumn	98	94-103
Maths year 6 spring	104	100-108

Pupil B has attained average scores. There is an increase in standardised scores between the two periods (from 98 in the autumn to 104 in the spring). The pupil's standardised score plot therefore lies above the diagonal line. However, when we consider the confidence band range we can see that there is overlap across the diagonal. It is possible that the pupil's true score could be the same on both tests so this pupil is not likely to have made more progress than might be expected. The teacher may find it useful to make comparisons between the year 6 autumn test and test results in previous years to see if the improvement in standardised score is part of a longer-term increase.

Pupil C scored as follows:

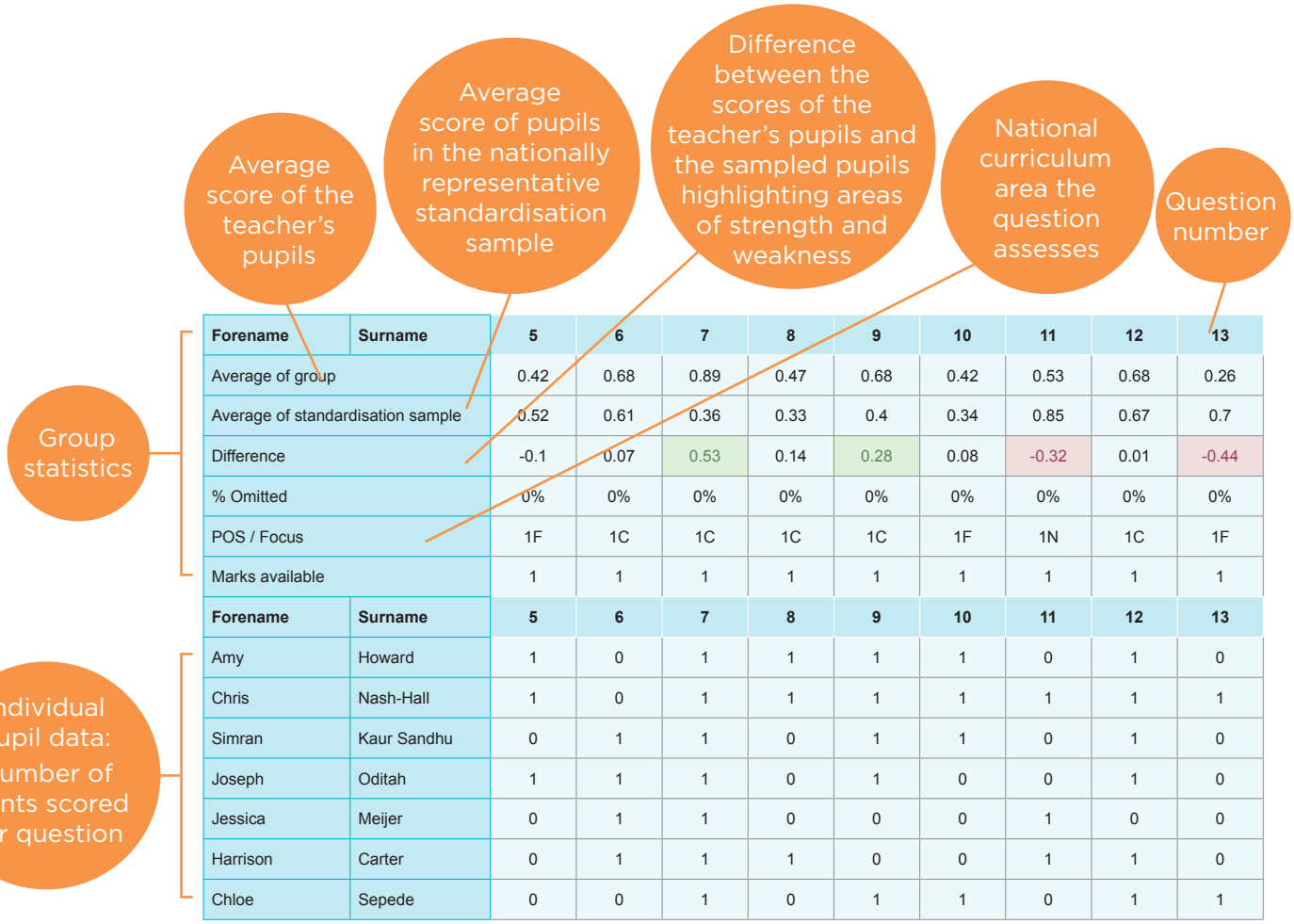
Test	Standardised score	Confidence band
Maths year 6 autumn	130	125-134
Maths year 6 spring	120	115-124

Pupil C has achieved standardised scores that are above average but seems to have made less progress than might be expected compared with pupils in the standardisation sample. We can infer this as the confidence band for spring (115-124) does not overlap with the diagonal line. The pupil's spring standardised score is therefore less than might be anticipated indicating that progress is probably less than expected. The teacher might want to investigate reasons for this e.g. whether or not the pupil has understood a new topic in year 6, such as long multiplication.

The Question Level Analysis report

The NFER Tests Analysis Tool includes a Question Level Analysis report. This can enable teachers to compare how their pupils have performed on individual items against the performance of the nationally representative sample which trialled the tests. It requires the score that each pupil gets on each question to be entered into the tool but the information returned can be very helpful. Where scatter plots have indicated that the class as a whole has made less progress than expected, the question level report may indicate particular areas of the curriculum which the pupils have found challenging. It also highlights the questions on which pupils have done particularly well compared to the standardisation sample. This enables teachers to identify which topics need a greater focus or perhaps a new approach as well as those where high performance can be celebrated and good practice shared.

Figure 13. **An example of the Question Level Analysis report showing question level analysis for questions 5 through to 13 of a year 1 mathematics assessment**



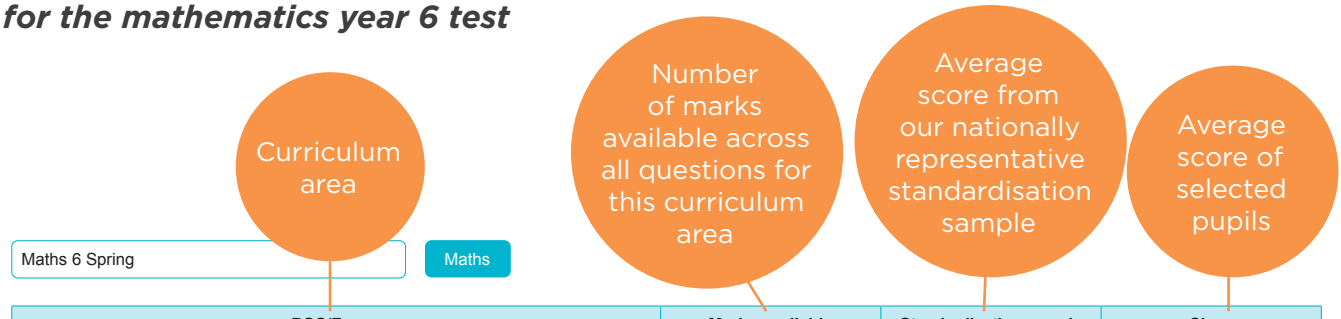
Forename	Surname	5	6	7	8	9	10	11	12	13
Average of group		0.42	0.68	0.89	0.47	0.68	0.42	0.53	0.68	0.26
Average of standardisation sample		0.52	0.61	0.36	0.33	0.4	0.34	0.85	0.67	0.7
Difference		-0.1	0.07	0.53	0.14	0.28	0.08	-0.32	0.01	-0.44
% Omitted		0%	0%	0%	0%	0%	0%	0%	0%	0%
POS / Focus		1F	1C	1C	1C	1C	1F	1N	1C	1F
Marks available		1	1	1	1	1	1	1	1	1
Forename	**Surname**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Amy	Howard	1	0	1	1	1	1	0	1	0
Chris	Nash-Hall	1	0	1	1	1	1	1	1	1
Simran	Kaur Sandhu	0	1	1	0	1	1	0	1	0
Joseph	Oditah	1	1	1	0	1	0	0	1	0
Jessica	Meijer	0	1	1	0	0	0	1	0	0
Harrison	Carter	0	1	1	1	0	0	1	1	0
Chloe	Sepede	0	0	1	0	1	1	0	1	1

All questions in which the teacher's pupils have an average score that is 0.20 or more greater than that of the standardisation sample are highlighted in green. For example, on question 7, which focuses on calculations (POS/Focus = C), the teacher's pupils performed extremely well (achieving an average score of 0.89) compared with the pupils in the standardation sample (average score 0.36).

This green highlighting indicates that the class have performed significantly better than the standardisation sample pupils. All questions in which the teacher's pupils have an average score that is 0.20 or more lower than that of the standardisation sample are highlighted in red. For example, on question 11, which focuses on number and place value (POS/Focus = N), the teacher's pupils did not perform well (average score 0.53) compared with the pupils in the standardisation sample (average score 0.85). Red highlighting indicates that the class have performed worse than the standardisation sample.

The Programme of Study report

Figure 14. An example of the Programme of Study report for the mathematics year 6 test



POS/Focus	Marks available	Standardisation sample	Class
Total	110	61.03	63.86
N Number and Place value	13	9.42	10.13
C Addition, subtraction, multiplication and division (calculations)	35	21.86	21.53
F Fractions, decimals and percentages	26	14.29	15.66
R Ratio	6	2.21	1.98
A Algebra	6	3.3	1.79
M Measurement	11	5.14	7.21
G Geometry - properties of shapes	5	1.84	1.92
P Geometry - position and direction	3	1.46	1.55
S Statistics	5	1.51	2.09

Information: The programme of study, or content domains, for mathematics are shown.
The number of marks available is given alongside standardisation sample and the class average. For tests with more than one paper, both the individual papers and the test total can be viewed.

The NFER Tests Analysis Tool includes a Programme of Study report which groups the items in the tests according to the curriculum area assessed and then summarises performance according to each curriculum area. Teachers can compare the average score in each curriculum area that their pupils achieved with the total number of marks available in each curriculum area and, additionally, with the performance of the nationally representative sample. The two comparisons are important as some curriculum areas may be difficult for all pupils, i.e. those in the national sample as well as those in the teacher's own school. For example, without comparison to the national sample, the class whose results are shown in Figure 14, appear to have scored proportionally fewer marks in both the Ratio and Algebra curriculum areas than in other curriculum areas (they have achieved about a third of the available marks.) We can see that nationally many pupils have found Ratio difficult and the pupils in the standardisation sample have also achieved about a third of the available marks. The class teacher might be more concerned about the pupils' performance in Algebra in which the standardisation sample appears to have achieved a greater proportion of the available marks (about a half). The class teacher might also note that there are some curriculum areas in which his or her pupils are performing well in, for example: Measurement and Statistics.

Further help

We hope that you found this guidance useful. If you require any further assistance, please email products@nfer.ac.uk and our specialist team will be on hand to help.

First published in September 2020

By the National Foundation for Educational Research, The Mere, Upton Park, Slough, Berkshire, SL1 2DQ

www.nfer.ac.uk

www.nfer.ac.uk/tests

@NFERClassroom

To access the NFER Tests Analysis Tool, visit www.nfer.ac.uk/hub.

If you have any further questions, our products team can be contacted by emailing products@nfer.ac.uk.

NFER Classroom

National Foundation for Educational Research

October 2020

P