



Technical Information

Standardisation¹

The standardisation of *Baseline* took place between September and October 2014. 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 categorised as either an independent or grammar school; then five categories of school intake were applied, based on overall school performance for primary schools using the Key Stage 2 outcomes.

Schools were selected by stratified random sampling procedures within these groupings.

The data from the standardisation study were analysed to provide information on the difficulty level of each question, 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 version of the assessment.

The *Baseline* standardisation was based on a sample of 1,737 children with approximately equal numbers of boys and girls. Most of the children were aged between 4:00 and 5:00.

Students who took part in the *Baseline* standardisation in September/October 2014 were followed up and were given the *Baseline Progress* assessment in June 2015. Some schools dropped out from the follow-up study and the standardisation of the *Baseline Progress* assessment was based on a sample of 1,270² children.

Test reliability

The reliability of an assessment is a measure of the consistency of a student's 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.

¹ *Baseline* is comprised of pre-existing trialled and standardised assessments for phoneme awareness and maths and a published, criterion-referenced assessment for language. Likewise, *Baseline Progress* is comprised of pre-existing, trialled and standardised assessments for letter-sound knowledge and word reading; the maths component is derived from a published test with new questions better suited to cover EYFS goals. In the latter there were more questions than would be needed for the final selection, however as all questions performed well, only one question was dropped post-standardisation.

² The sample sizes are large for an individually administered assessment.



The reliability of the two assessments was estimated using the Cronbach's Alpha formula, which produces values ranging from 0 to 1. The reliability values for the various sections of each assessment are given in the table below and show that the overall assessments and their component parts are very reliable. The reliability for the Mathematics section is lower as it is a shorter assessment.

Test	Reliability
Overall <i>Baseline</i> Test	0.93
Literacy and Language & Communication	0.92
Mathematics	0.77
Overall <i>Baseline Progress</i> Test	0.94
Literacy	0.94
Mathematics	0.76

When interpreting the score of an individual student, the standard error of measurement (SEM) is a more useful statistic than a reliability coefficient. The SEM 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 the overall *Baseline* test is 4.0, and for an average-performing student with an overall Standard Age Score (SAS) of 100, there is a 90% chance that the student's true SAS will be in the range +/- 7.0, i.e. from 93 to 107.

Test	SEM	90% SAS confidence band (+/-)
Overall <i>Baseline</i> Test	4.0	7
Literacy and Language & Communication	4.2	7
Mathematics	7.2	12
Overall <i>Baseline Progress</i> Test	3.7	6
Literacy	3.7	6
Mathematics	7.3	12

Test-retest reliability: *Baseline*

A test-retest study was carried out involving 503 children who completed the *Baseline* assessment twice. A small number of children within a school were randomly selected to be re-assessed with either the same assessor or with a different assessor. Administration was carried out within one to two weeks of the original administration to lessen learning effects but to allow the children a reasonable gap between sessions.



The results of the test-retest study involving the same assessor show a very high degree of correlation (0.95) between the raw scores at two administrations. The raw scores at second administration were on average around 2 raw score points higher. This increase gain is expected due to the learning the children would have received within the two week period of re-test.

The results of the test-retest study involving a different assessor on the second administration also showed a similar very high degree of correlation (0.93) between the raw scores at two administrations.

Test-retest: *Baseline to Baseline Progress*

Around 1,250 children who had completed *Baseline Progress* in June 2015 had previously been administered *Baseline* in autumn 2014. The overall correlation between *Baseline* and *Baseline Progress* standard age scores was 0.64. This relationship enables us to track children’s progress. The level of progression has been divided into five categories based on the proportion of children in these categories nationally as shown in the table below.

Progression category	Percentage of children nationally
Much higher than expected progress	10%
Higher than expected progress	15%
Expected progress	50%
Lower than expected progress	15%
Much lower than expected progress	10%
Total	100%

Schools can download the online report showing how their children have progressed during the school year as shown in the example pages below.

School: Sample School	
Group: Sample Group	
Date(s) of testing: 04/05/2016	No. of pupils: 12

Scores for the group (by surname)

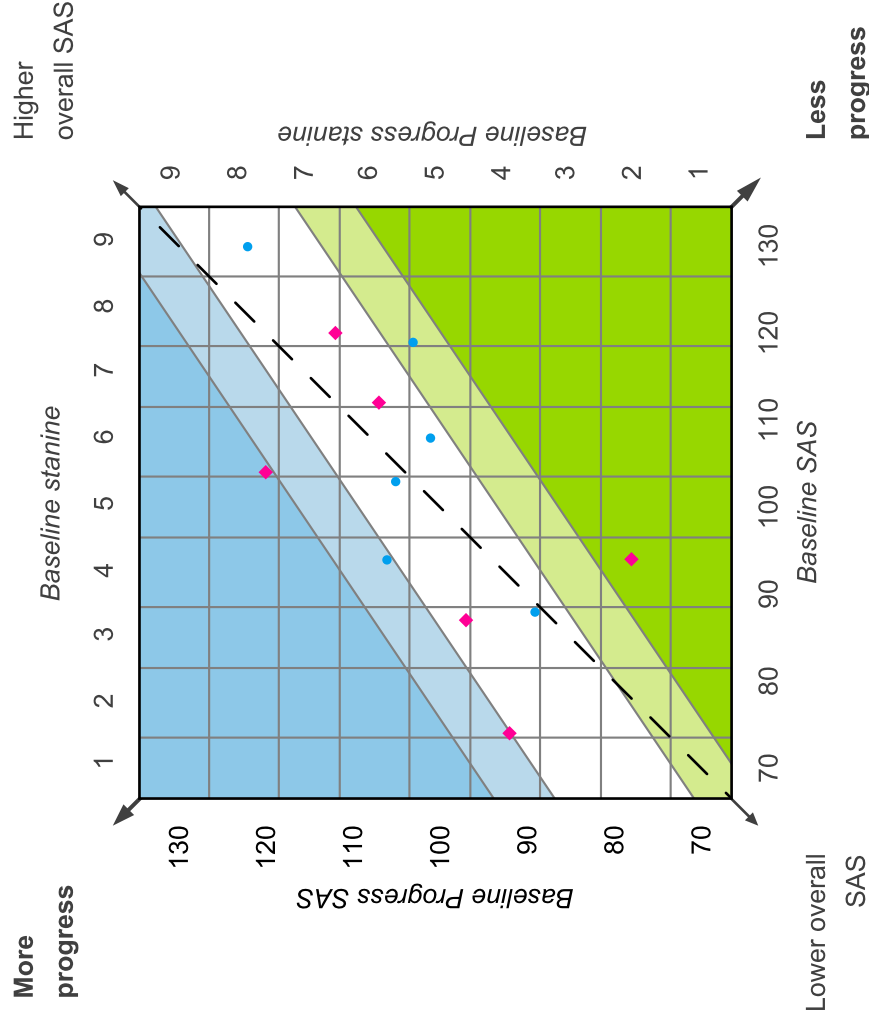
Pupil name	Overall										Literacy			Mathematics			Overall Progress Category
	T-score	Overall SAS	SAS (with 90% confidence bands)			Overall ST	Overall NPR	Overall GR (/12)	SAS	NPR	ST	GR (/12)	SAS	NPR	ST	GR (/12)	
Peter Adams	44	101				5	53	8	110	74	6	3	85	16	3	=9	Expected
Tom Albright	61	106				6	66	5	104	60	6	7	102	55	5	=7	Higher than expected
Natasha Anthony	46	97				5	42	9	105	63	6	=5	84	14	3	11	Expected
Jenny Dunn	57	107				6	68	4	103	58	5	8	113	80	7	=1	Expected
Christopher Gibson	42	89				4	24	11	84	14	3	12	106	66	6	5	Expected
John Jameson	50	103				5	58	7	101	53	5	9	112	78	7	3	Lower than expected
Alice May	62	112				7	78	3	109	72	6	4	111	77	6	4	Expected
Nita Moss	42	92				4	30	10	99	48	5	10	85	16	3	=9	Higher than expected
Rose Nash	68	120				8	91	2	115	84	7	2	113	80	7	=1	Much higher than expected
Simon Shah	50	105				6	63	6	105	63	6	=5	105	63	6	6	Expected
Sophie Turay	35	78				2	7	12	85	16	3	11	69	2	1	12	Much lower than expected
Peter Watt	61	122				8	93	1	129	97	9	1	102	55	5	=7	Expected

School: Sample School	
Group: Sample Group	No. of pupils: 12
Date(s) of first test: 07/09/2015	
Date(s) of second test: 04/05/2016	

Progress profiles

In this report progress is calculated from each pupil's overall score for *Baseline* and a comparison is made to each pupil's overall score for *Baseline Progress*. The national data is used to determine whether progress is as expected (that is the relationship between the scores is typical) or is above or below what is predicted by the national data. It will be important to consider the overall score level as well as progress. Progress is categorised as expected, higher or lower than expected or much higher or much lower than expected.

Note that only those students who have completed *Baseline* and *Baseline Progress* are able to have performance compared and therefore progress reported in this section.



- Much higher than expected progress
- Higher than expected progress
- Expected progress
- Lower than expected progress
- Much lower than expected progress

- Boys
- Girls



The table below shows the number of pupils in each progress category against the national distribution.

Process category	National	Group	
	%	%	No. of pupils
Much higher than expected progress	10%	8%	1
Higher than expected progress	15%	17%	2
Expected progress	50%	58%	7
Lower than expected progress	15%	8%	1
Much lower than expected progress	10%	8%	1

Differences by sub-groups

The tests have been age standardised to a national mean of 100 and standard deviation of 15. The table below shows the mean SAS scores for various sub-groups of children who had completed both the *Baseline* and *Baseline Progress*. Mean SAS differences of more than 3 points can be considered to be significant.

- Girls scored higher than boys by on average 3.7 SAS points at the initial *Baseline* and the gap narrows slightly to 2.4 SAS points by the end of the school year.
- Children with English as an additional language scored on average 4 SAS points lower at the initial *Baseline* but they made greater progress and their average score by the end of the school year was 2 SAS points higher than those children with English as a first language.
- Children receiving free school meals (pupil premium) score on average 4.6 SAS lower at the initial *Baseline* and the gap narrows slightly to 3.5 SAS by the end of the school year.



		Overall Mean SAS score		Number of students
		<i>Baseline test</i>	<i>Baseline Progress test</i>	
Gender	Girls	100.4	101.1	636
	Boys	96.7	98.7	606
	Mean SAS difference	3.7	2.4	
English as an additional language	No	99.6	99.3	904
	Yes	95.7	101.4	359
	Mean SAS difference	4.0	-2.1	
Children receiving free school meals (pupil premium)	No	99.0	100.6	861
	Yes	94.4	97.1	237
	Mean SAS difference	4.6	3.5	

Test-retest: *Baseline to Progress Test in English 5 and Progress Test in Maths 5*

644 children who had previously been administered *Baseline* in autumn 2014 completed the *Progress Test in English 5* and *Progress Test in Maths 5* in April and May 2015. The correlation between overall *Baseline* outcomes and *Progress Test in English 5* SAS scores was 0.56 and was 0.63 with *Progress Test in Maths 5*.

The level of progression varies depending on the initial *Baseline* score. The table below gives an approximate indication for an average child. For example, if a child with an initial *Baseline* SAS score of 100 subsequently has a SAS score of 112 in *Progress Test in English 5*, the 12 point increase in score places the child in the 'Higher than expected progress' category.

Difference in SAS scores	Progression category
Increase by > 15	Much higher than expected progress
Increase by 10 to 15	Higher than expected progress
Within +/- 9	Expected progress
Decrease by 10 to 15	Lower than expected progress
Decrease by > 15	Much lower than expected progress