A guide to standardised tests
What are standardised tests?

Standardised tests are tests which require all test takers to answer the same questions in the same way, and that are scored in a consistent manner, which makes it possible to compare the relative performance of pupils or groups of pupils.

Many assessment experts consider standardised tests to be a fair and objective method of assessing pupils, mainly because the standardised format reduces the potential for favouritism, bias, or subjective evaluations.

" Standardised tests measure performance relative to all other pupils taking the same test. "

Standardised tests are used for a number of educational purposes, for example, they may be used to determine what a child knows and can do on entry into school or to identify pupils who need special education support. The following are two types of standardised test offered by GL Assessment:

- **Abilities test**, such as the Cognitive Abilities Test (CAT4) [https://www.gl-assessment.co.uk/products/cognitive-abilities-test-cat4/](https://www.gl-assessment.co.uk/products/cognitive-abilities-test-cat4/), are designed to predict a pupil’s ability to succeed in an academic endeavour by evaluating verbal, non-verbal, quantitative and spatial ability. Abilities tests are “forward-looking” in that they predict how well pupils will do in the future based on their abilities which support academic progress, such as types of reasoning. General cognitive ability is the single strongest predictor of how well a child will do in their GCSEs.

- **Attainment test** such as the Progress Test Series [https://www.gl-assessment.co.uk/products/progress-test-series/](https://www.gl-assessment.co.uk/products/progress-test-series/), are designed to measure the knowledge and skills from key areas of the curriculum pupils have learned in school or to determine the progress they have made over a period of time. The tests may also be used to evaluate the effectiveness of a schools and teachers, which is the case with key stage 2 national tests and GCSEs. Attainment tests are “backward-looking” in that they measure how well pupils have learned what they were expected to learn.

" Abilities tests are designed to indicate a pupil’s propensity to succeed at school, while attainment tests measure the curriculum knowledge and skills acquired. "

Why use standardised tests?

Schools and Multi-academy trusts (MATs) or school groups choose to use standardised tests in order to gain consistency of assessment across the school or schools and to gain important insight into how pupils and schools are performing in relation to other schools nationally. School leaders are able to benchmark a school’s performance nationally against other schools by using standardised tests. Standardised tests are used as part of an assessment regime alongside tests of curriculum knowledge and understanding and in-class teacher assessment as well as the outcomes from national tests, such as those at the end of Key Stage 2.
What are the benefits of standardised tests?

There are a number of key benefits to using standardised tests within schools or groups of schools:

- See where a pupil or group of pupils is strong or requires additional support; can be applied to improve teaching and learning.
- Provide a more reliable comparison of the test outcomes than non-standardised tests.
- Provide quantifiable measures, such as age-standardised scores (SAS) and indicative prediction of key stage 2 national test or GCSE performance.
- Indicate how a pupil or groups of pupils have performed in relation to others nationally.
- Use at regular intervals over time, allowing progress to be tracked in an effective and objective way.
- Used to measure the impact of interventions. For example, NGRT can be used to measure reading age before and after an intervention and thereby provide evidence of the impact of the intervention.

The NGRT can be used to objectively measure reading age before and after an intervention and thereby provide evidence of the impact of the intervention.

What are the limitations of standardised tests?

Standardised tests form part of an effective assessment system, but they cannot measure everything. For this reason standardised tests are best used alongside regular in-class formative feedback about what a pupil knows or can do.

- Any test will reflect a pupil’s performance at a point in time, and this may be affected by factors such as tiredness or illness.
- Some pupils with Special Educational Needs may be unable to access particular tests.
- Some pupils with very high attainment will reach the “ceiling” of a static test so the information from the test is not that helpful to the teacher.
How can standardised tests be made appropriate for a range of pupils?

Standardised tests designed to include the majority of pupils will not cater well to those with very low or exceptionally high attainment. This is why computer-adaptive tests are designed to adjust the difficulty of questions—based on the responses provided—to match the knowledge and skills of a test taker. If a pupil gives a wrong answer, the computer follows up with an easier question; if the pupil answers correctly, the next question will be more difficult. So, computer-adaptive tests measure more precisely than fixed-form standardised tests.

Diagnostic tests are designed to highlight particular errors and misunderstandings which indicate a key learning need. If a pupil has very weak skills, they may require a test which allows them to work with material which is matched to their skill level, so scores reflect both the age of the pupil and the difficulty of the material used for testing. Diagnostic information is thus enhanced. The York Assessment of Reading for Comprehension (YARC), for example, gives scores for reading rate, error and comprehension.
How are standardised tests developed?

Standardised tests are developed in a very structured way to ensure that they have validity (they test what they claim to test or the results predict future behaviour) and reliability (the test gives the same or similar results repeatedly over time). The rigorous development process can take between two to four years to complete and involves a number of stages and experts.

1. Design a test framework (which for curriculum tests samples the knowledge and skills to be assessed)
2. Write a large number of content questions
3. Trial questions with pupils in schools
4. Reject questions that perform badly in trialling
5. Develop tests for a standardisation trial
6. Conduct the standardisation trial with a statistically significant and nationally representative sample of pupils
7. Develop the norm-referenced measures (such as age-standardised scores and percentiles) to enable comparison of pupil or school performance to performance nationally.

How are standardised test results described?

**Raw score:** The raw score is the total number or points or marks the pupil has scored on the test. Standardised tests convert raw scores of, for example 33 out of 50, to scores on a readily understandable scale, a normal distribution curve.

**Age-standardised score:** An age standardised score converts a pupil’s raw score to a standardised score which takes into account the pupil’s age in years and months and gives an indication of how the pupil is performing relative to a national sample of pupils of the same age. The average score is 100. A higher score is above average and a lower score is below average. The SAS is key to benchmarking and tracking progress and is the fairest way to compare performance of different pupils within a year group or across a year group.

“The Age-standardised score (SAS) is a recognised benchmark to measure against a national sample of pupils of the same age.”

**Stanine:** The stanine places the pupil’s score on a scale of 1 (low) to 9 (high) and offers a broad overview of performance.

**Confidence bands:** The confidence band is an indication of the range within which a pupil’s score lies. The narrower the band the more reliable the score. Reliability values range from 0 to 1, with 0.9 being very high and 90% confidence bands showing strong reliability. This means that on 9 out of 10 occasions, the true value of the score is within the score band. Reliability can be determined by testing and re-testing a group of pupils and seeing how well the scores correlate between the two testing occasions.

**Group Rank (GR):** Group Rank shows how each pupil has performed in comparison to those in the group.
National Percentile Rank (NPR): The National Percentile rank relates to the SAS score and shows the percentage of pupils obtaining a certain score or below. A NPR of 50 is average since 50% of pupils obtained an SAS of 50 or below. A NPR of 5 indicates a pupil’s score is within the lowest 5% of the nationally representative sample and a NPR of 95 means that a pupil’s score is within the highest 5% of the national sample.

Reading age: Also known as age-equivalent score, reading age is the age at which a particular score is obtained by the average pupil. So for example, if the average raw score for a 7-year-old on a reading test is 50, any pupil with a raw score of 50 will have a reading age of 7 years. Reading age can also be used to measure the impact of reading interventions. In most cases, it is not sensible to relate scores for pupils with above average reading ages (above their chronological age) as age-equivalences by definition relate to an average.

Performance indicators: Indicators showing potential pupil attainment at the end of KS2 and GCSE. They are based on the significant and positive correlation (a link supported by statistical data) between a pupil’s scores on tests, and his or her performance in Key Stage 2 national tests and GCSE examinations. The performance indicators provide a reliable indicator of future performance. This analysis is based on results from a large sample of schools and pupils, and the indicators will be updated regularly to reflect changes in national KS2 attainment.