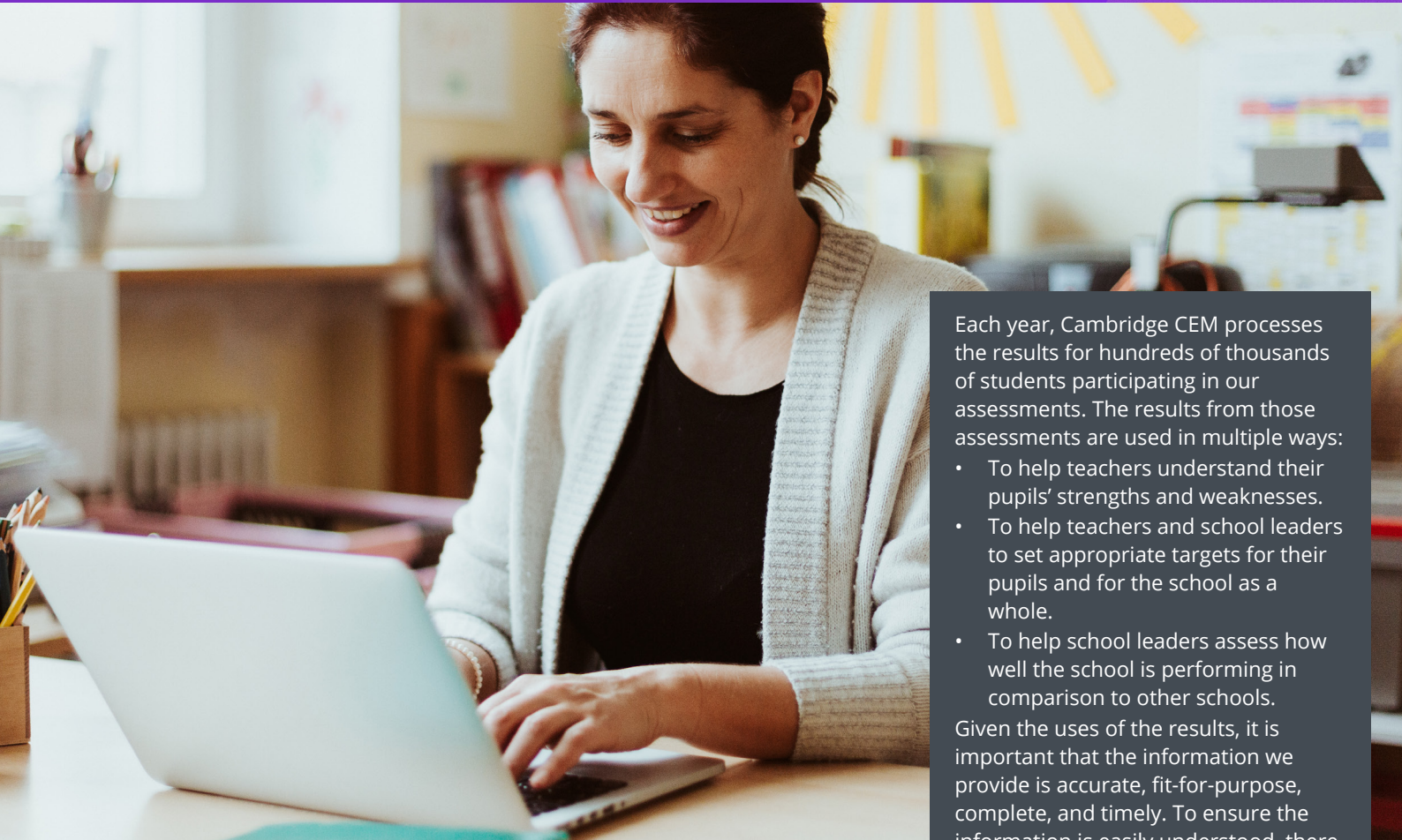


# Data Reliability for Cambridge CEM assessments



Each year, Cambridge CEM processes the results for hundreds of thousands of students participating in our assessments. The results from those assessments are used in multiple ways:

- To help teachers understand their pupils' strengths and weaknesses.
- To help teachers and school leaders to set appropriate targets for their pupils and for the school as a whole.
- To help school leaders assess how well the school is performing in comparison to other schools.

Given the uses of the results, it is important that the information we provide is accurate, fit-for-purpose, complete, and timely. To ensure the information is easily understood, there is a glossary at the end of this handout.

## Quality assurance of Cambridge CEM assessments

When we develop new assessment questions and new assessments, these must meet several quality control criteria:

### Validity

- Does each assessment section measure the construct that it is intended to measure?
- Are the inferences that may be made from the assessment's outcome appropriate?
- Does an assessment cover a sufficient range of the construct and not include construct-irrelevant items?
- Are questions statistically robust, and do they perform well across a range of abilities and at the appropriate difficulty for the assessment year group?

### Reliability

- Do the questions in a section measure the same construct?
- Is the marking of questions consistent and accurate?

### Fairness

- Are the questions unbiased towards different sub-groups within the assessment year group, e.g., gender, culture, and ethnicity?
- Are questions accessible to the assessment year group regardless of their context and background?

### Accessibility

- Is the assessment accessible to all or are there adaptations that need to be incorporated such as extra time provision?

There may be circumstances which affect a student's progress through an assessment to the extent that their results may be unreliable, for example if they feel ill, disinterested, or do not complete the assessment in the time available. Where possible we identify these issues, and those results are automatically flagged.

## Meaningful and accurate scores

To make sure that our standardised scores are meaningful, we have a clearly defined target population that we standardise against. For example, for MidYIS and Yellis we provide nationally standardised scores, with the target population being all students in mainstream Secondary schools.

Additionally, to check that our standardised scores are representative of those student populations we need to obtain accurate estimates of the mean and standard deviation of students' performances in those populations. To achieve this, we use the following techniques:

- We weight our sample of students' performances by school sector (state, grammar and independent), so that the weighted percentages of students in each sector match the national proportions of 89.1%, 3.9% and 7% respectively.
- We control for any bias in the ability profile of our school sample by using regression models that link schools' performances on the Cambridge CEM assessments to their performances at GCSE or A-Level.

## How accurate are the scores for an individual student?

No assessment can measure a student's ability with absolute accuracy and there are well established ways to estimate the measurement 'error' on a student's result. We present these on each student's Individual Student Reports by providing 95% confidence intervals around the student's scores.

*"We look at the discrepancies in the data – if, say, a student has high vocabulary and maths scores, but a lower non-verbal and reasoning score, we might explain this to the teachers and show them how that might affect the students in class."*

**Karen Barker, Deputy Head & SENDCo**

## Quality Assurance on the predicted grades and value-added data

The predictive data we provide are calculated using data from previous students who have taken a Cambridge CEM assessment and then later taken external examinations such as GCSE, Scottish Nationals, IB Diploma, or A-Levels. We must have at least 10 schools and at least 100 students at each school.

To be included in the set of subjects that we provide predictions and value-added data for, a subject's sample must meet our quality control criteria in terms of sample size, number of schools, sampling error and correlation.

To ensure that the national value-added samples are nationally-representative in terms of school sector (i.e., state schools vs independent schools), we apply weighting factors so that the percentage of students from independent schools matches the national figure of 7%.

We typically update the predictions for active students (i.e., students yet to take their examinations) in late spring each year to use the most up-to-date information we have. That typical annual cycle has been disrupted and made more complicated in recent years by the pandemic, but the same principle always applies - we want the predictions used in the value-added to be as accurate and appropriate as possible.

## The validity of value-added calculations for 2021 and 2022

The CEM value-added for both 2020 and 2021 did not take the 'generosity' of grading from Centre-Assessed Grades (CAGs) or from Teacher Assessed Grades (TAGs) into account. Hence, most schools will find that the value-added for those years is higher than they would usually achieve.

In both years, prior to the release of examination results, it was not clear how grading standards would be affected by the unusual grading process. While this information did become available later, it also became clear that grading standards had varied somewhat between schools. Hence, any attempt to 'correct' the value-added by taking account of the grading generosity at a national level, would not have led to fair estimates of schools' value-added. Instead, we decided to leave the value-added as it was and to explain its limitations.

## Glossary

- **Construct:** in the context of assessment, a construct is the theoretical concept such as knowledge, skill or ability that the assessment is aiming to measure.
- **Sample size:** the number of students in a dataset.
- **Regression line:** the result of fitting a statistical model to a dataset. It is the "line of best fit" and it provides a simple summary of the relationship between a predictor variable (e.g., a CEM score) and an outcome variable (e.g., an exam grade).
- **Chances graph:** shows a student's likelihood of achieving each of the possible grades in a particular examination subject.