

Technical note to Writing development report

April 2022



Methods

This note describes how the research team considered the uncertainty of the changes measured through the analytical methods applied in the report 'Writing development: what does a decade of NAPLAN data reveal?' and explains how the team accounted for the limitations noted in the report.

When measuring changes over time, as the analysis in this report did, it is important to consider the uncertainty of the changes. Where such information is publicly available, the report included it. For example, statistical significance of the changes in the overall scores was included in the report (for example, p.25, 32). However, for the main analysis which concerned the trends of the criterion scores, such information was not available publicly and it was not possible for the team to perform statistical testing of the significance in changes in criterion scores either.

To test for statistical significance, information about the sampling error, measurement and equating error for each criterion is required. However, this information was not available to the research team. Further information about the importance and availability of each of these 3 elements required for significance testing is as follows:

- **Sampling error** can be calculated but is small given that trends of large cohorts at the population level were being examined.
- **Measurement error** associated with the **criterion** scores arises from factors that influence score generalisability, including marking inconsistency of an **individual criterion** in a test year. The only measurement error ACARA publishes is that of the overall writing scores, however this is different from the error associated with criterion scores. To clarify, measurement error of an overall score is estimated from repeated measures of criterion scores from the student across the 10 writing criteria and the same information from all other students. Both marking inconsistency **across criteria** (that is, markers being more lenient on some criteria than on others) and **of individual criteria** contribute to this (published) measurement error.
- **Equating error** is the uncertainty associated with any equating shifts estimated from a process used to equate one year's results to other years. The size of the error depends on the equating process used and how well it removes the influence of factors such as marking inconsistency over time on the comparability of results in a trend. ACARA does not conduct equating processes for individual criterion, so there is no equating error information reported for the criterion scores.

As it is not possible to perform the statistical significance testing of changes in criterion scores using publicly available information, and to ensure the reliability of our findings from the descriptive analysis, the research team used a different method (referred to as the Differential Item Functioning (DIF) analysis in the report) to independently verify the trends from the descriptive analysis. The report notes that it's more important to consider practical significance than statistical significance when interpreting DIF for large samples (see the Methodology section). The same consideration applies to the interpretation of the trends of DIF when large samples are involved. The report also notes the implications of the different sources of uncertainty (for example, prompt effect, cohort effect, marker effect) on the interpretation of the study

findings in the Limitations section and where appropriate other places in the report. These discussions were supported by relevant available evidence published to date.

The accuracy of this note was verified by Dr Ray Adams, Chair of the Australian Curriculum, Assessment and Reporting Authority (ACARA) Measurement Advisory Group.



For more information visit
edresearch.edu.au

